HANDBOOKS IN ECONOMICS

Barry R. Chiswick
Paul W. Miller

Economics of International Migration

VOLUME 1B

NORTH-HOLLAND

Handbook of THE ECONOMICS OF INTERNATIONAL MIGRATION

Volume 1B The Impact and Regional Studies

Handbook of THE ECONOMICS OF INTERNATIONAL MIGRATION

Volume 1B The Impact and Regional Studies

Edited by

BARRY R. CHISWICK
George Washington University
and
IZA—Institute for the Study of Labor

PAUL W. MILLER
Curtin University

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo
North-Holland is an imprint of Elsevier

N·H

CHAPTER 18

Migration, International Trade, and Capital Formation: Cause or Effect?

Gabriel Felbermayr*, Volker Grossmann†, Wilhelm Kohler‡
*IFO Institute – Leibniz Institute for Economic Research, University of Munich, Poschingerstr. 5, D-81679 Munich, Germany
†University of Fribourg, Bd. de Pérolles 90, CH-1700 Fribourg, Switzerland
‡University of Tübingen, Mohlstrasse 36, D-72074 Tübingen, Germany

Contents

Handbook of the Economics of International Migration, Volume 1B
ISSN 2212-0092, http://dx.doi.org/10.1016/B978-0-444-53768-3.00018-7

1. INTRODUCTION

Migration is the oldest action against poverty (*Galbraith, 1979*)

Global economy . . . a gated wealthy community consisting of the advanced countries, surrounded by impoverished ghettos, with immigration restrictions preventing the ghetto residents from moving to where their productivity and well-being would be higher (*Freeman, 2006*)

In *grand historical perspective*, globalization is, first and foremost, a story of migration. Thirteen millennia of human migration and settlement, from Africa over Eurasia to the Americas, as described by Diamond (1997), still form the basis of world trade. Over the very long-run horizon, then, there is no doubt migration has caused trade, and not the other way round. The key force was that in their "new countries" migrants eventually ended up producing goods that were in short supply in their "old countries", mostly for reasons of nature and climate.[1] Since trade is a precondition for capital movements, capital movements thus appear at the end of the line of causation.

[1] A modern version of this is Asian workers migrating to Sweden, albeit on a temporary basis, picking blueberries that Sweden then exports to Asia; see "Berry pickers, unite!", The Economist, 4 August 2012.

Over *shorter horizons*, looking at the recent waves of modern economic globalization, the relationship between international migration and international trade as well as capital movements is considerably more involved. Moreover, since capital movements are typically related to capital formation, the relationship should also be seen as one between trade, migration, and capital formation. In this chapter, we want to give an overview of what modern economic analysis tells us about this relationship.

Currently, an estimated 3% of the world population, or some 232 million people, live outside their countries of birth. This is commonly regarded as a low figure, although it has increased by more than 50% from less than 2% in 1960.[2] But what is the benchmark against which to judge whether the present stock of migrants is high or low? Perhaps more informative is a comparison of living conditions in different parts of the world, since this tells us about remaining incentives for, and gains from, international migration. In 2012, the 25th percentile of GDP per capita (at international PPP) as a fraction of the 75th percentile was a mere 0.136, up from 0.119 in 1980. For private household expenditure, the fraction was 0.185 in 2012, up from 0.074 in 1980.[3] Figures of a similar magnitude have been presented by Freeman (2006), based on occupation-specific wages taken from the NBER Wages around the World database. What is striking is both the degree of inequality, as well as the slow speed of convergence, if any. Figure 18.1 gives a somewhat more comprehensive picture by plotting country-specific deviations from an unweighted mean for 1980 against 2012. Data points above (below) the line in the positive (negative) orthant indicate divergence. A similar picture is obtained using household expenditure per capita.

If known to poor individuals in poor countries, income gaps of this magnitude must be expected to constitute powerful incentives for migration. But such incentives may be misleading and distorted. The key question arising in this context are (i) whether international income or wage inequality mirrors differences in *country-specific* determinants of workers' wages, and (ii) whether wage incomes in different countries correctly reflect marginal productivities of employed workers. If the answers to both questions are yes, then international migration is an important key to improving living conditions of the world's poor, since movement of labor from low to high productivity countries would clearly increase the *efficiency* of worldwide factor use. But, setting wage distortions aside, inequality may also mirror differences in *worker-specific* wage determinants, which are unlikely to change much through migration alone. As we shall see below, empirical evidence suggests that a large part of international wage gaps must be attributed to *individual*

[2] See also United Nations Secretary General (2013), where it is estimated that in the period from 2000 up to 2013 the number of migrants had increased by 32 million in the North and 25 million in the South.

[3] Our focus here is not global inequality among *individuals*, which would require looking at internal distribution of income within countries. Specifically, the lack of international convergence suggested by the above simple measures is perfectly consistent with a reduction through the same period in global international inequality, as portrayed in Sala-i-Martin (2006). Our point is that enormous income gaps still exist between countries and that this will be a powerful incentive for future migration.

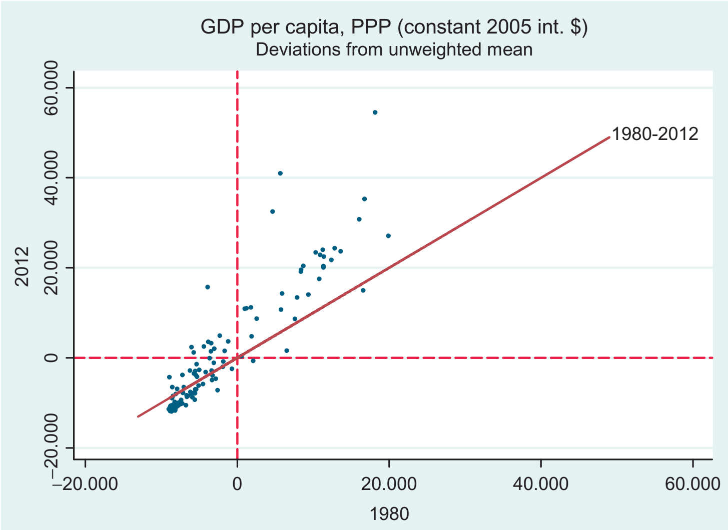

Figure 18.1 International income inequality and convergence.

characteristics of the worker, like education and skills, and not to characteristics of the *country* of work.[4] For individuals, it will often be hard to tell whether it is country-specific determinants or individual characteristics that are responsible for observed wage gaps.

At the present, however, would-be migrants face stiff immigration restrictions in most rich countries. Looking at policies pursued in the developed world with respect to different forms of globalization, one observes a certain amount of asymmetry. *Goods markets* are characterized by several decades of multilateral and regional trade liberalization after World War II. Despite the fact that negotiations towards further multilateral liberalization in the Doha round of the WTO presently seem stuck, the policy rhetoric is imbued with the idea of gains from trade. In a similar vein, ever since the breakdown of the Bretton Woods system in the early 1970s, almost all countries of the world seem committed to *capital mobility*, although there are widely shared concerns about detrimental effects of speculative short-run capital flows. But when it comes to *international migration*, the policy rhetoric as well as the policy practice is characterized by the notion of a country's "natural right" to protect its domestic labor market.[5] As we shall see in Section 2, this asymmetry sets the second wave of economic globalization in the late twentieth century apart from its nineteenth century counterpart.

Economic globalization after World War II has featured a strong increase in all forms of internationalization, not just migration. Indeed, trade and capital movements appear as more important drivers of globalization than does migration. Figure 18.2

[4] See the study by Hendricks (2002).
[5] This striking policy asymmetry is also emphasized by Freeman (2006).

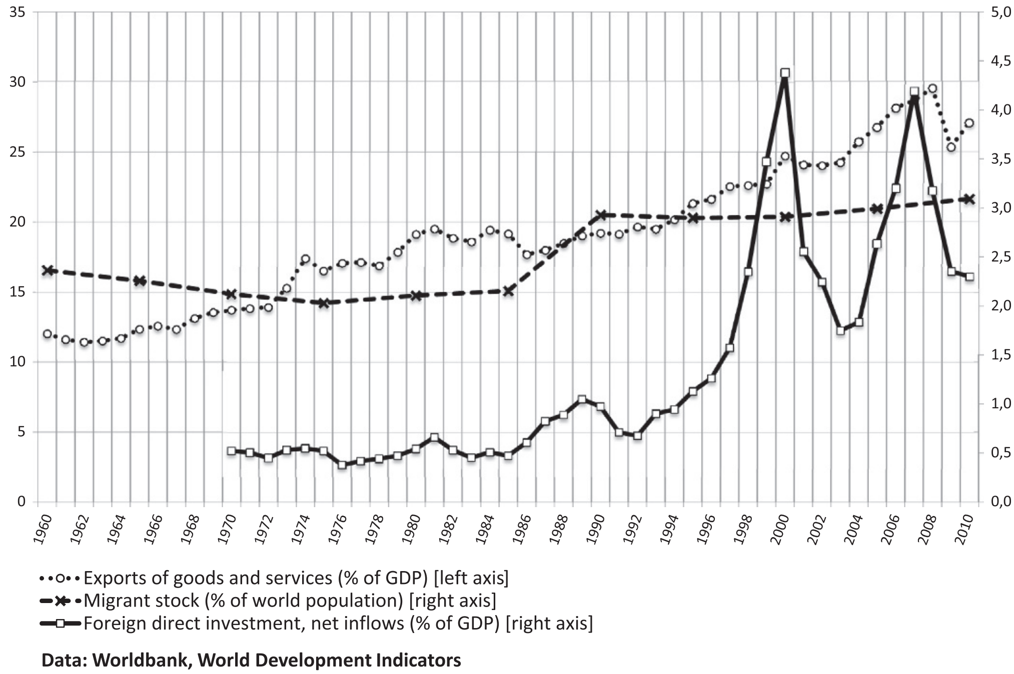

Figure 18.2 Evolution of world trade, capital flows, and migration.

presents a quick overview of the evolution of worldwide international migration, measured through gross migrant stocks, alongside international trade and capital flows over the past five decades. We see that, on a global scale, the major surge in migration took place in the 1980s, but this conceals much variation across countries. Table 18.1 therefore presents annual net migration rates over the recent past, as well as the present stock of migrants for major industrialized countries. We see considerable idiosyncratic variation. Perhaps most striking is the surge of immigration into Spain during the first decade of this century. The increase in world trade, measured as gross trade in % or GDP, appears somewhat more evenly spread out over the decades considered, although the figure clearly demonstrates the impact that financial disruptions, such as the Asian crisis of 1997 and the more recent crisis of 2007/08, have on global trade. The figure also demonstrates that foreign direct investment (FDI), measured as net inflows in % of GDP, has not taken off before the late 1980s. Unsurprisingly, FDI flows appear as a highly volatile form of globalization, as evidenced by the sharp downturn after 2000, due to the "dotcom bubble" of 2000, which was followed by a fast recovery and a further downturn caused by the financial crisis of 2007/09. Overall internationalization of capital markets may be measured by total foreign investment stocks (assets) in % of GDP. For major industrialized countries, this figure has risen from 18% in 1960 to 112% in 2000; see Obstfeld and Taylor (2004). This was importantly driven by a

Table 18.1 Immigration flows and stocks of immigrants for selected countries

Country	Net migration rates (5-year averages)							Stock of migrants*
	1980	1985	1990	1995	2000	2005	2010	2010
Australia	1.56%	3.11%	3.91%	2.05%	2.43%	3.24%	5.04 %	24%
Canada	1.63%	1.27%	3.20%	2.19%	2.38%	3.37%	3.22 %	21%
France	0.46%	0.51%	0.48%	0.21%	0.31%	1.21%	0.77%	10%
Germany	0.42%	−0.14%	2.04%	4.05%	1.02%	0.93%	0.67%	12%
Italy	0.29%	0.47%	−0.02%	0.27%	0.40%	3.16%	3.30%	7%
Spain	0.21%	−0.11%	−0.17%	0.81%	1.98%	6.52%	4.88%	15%
United Kingdom	0.07%	−0.17%	0.17%	0.35%	0.73%	1.61%	1.64%	11%
United States	1.55%	1.39%	1.52%	1.67%	3.02%	2.10%	1.60%	12%

*In percent of total population
Source: World Bank, World Development Indicators.

reduction in capital controls after the collapse of the Bretton Woods system in the early 1970s and the revolution of information technology over the past two decades.

However, from Figure 18.1 we are tempted to conclude that these secular trends of enhanced economic globalization have done relatively little to reduce international inequality. What, then, are the dimensions in which *countries* (as opposed to individuals) may differ and which may explain individual well-being and inequality in incomes across countries? This is probably one of the most intensively researched questions in economics. On a fundamental level, we may identify five *country characteristics* that may determine the economic perspectives of a country's inhabitants, in absolute terms and relative to other countries. Ordering by the degree of exogeneity, we may list: (i) Its climate, (ii) its size and geographic proximity to other big countries, (iii) its institutions, (iv) its level of technological knowledge, and (v) its factor endowment. In this chapter, we are mainly concerned with characteristics (iv) and (v), that are subject to medium-run change and policies, and our primary focus lies on migration.

Against the backdrop of international income inequality and the associated inefficiency of world factor use, what is the appropriate policy stance vis-à-vis trade, migration and capital flows? What are the effects of trade liberalization and loosening migration restrictions respectively on internal income distribution and aggregate welfare within the countries involved? Will enhanced trade among poor countries or between poor and rich countries have a tendency to reduce international income gaps, thus also reducing migration incentives? Will international migration, through its effect on countries' relative factor endowments, reduce the scope for international trade, in addition to leveling out international wage gaps? Is there a reverse causality in that an enhanced network of migrants facilitates easier and more gainful trade? What are the roles that migration may play in the process of accumulation of physical or human capital?

A large part of the literature tries to address at least some of these questions focusing on countries' factor endowments. As stressed very forcefully by Ohlin (1933), *endowment-based* models of trade imply that trade and migration as well as capital movements are all working towards a reduction of inefficiencies in the worldwide use of different types of factors, thereby also leading to convergence of factor prices and towards a reduction of existing international income gaps. The political challenge, according to this view, is to foster those forms of globalization (trade, migration, and capital flows) that represent the least costly way towards a more efficient use of world factor endowments.

However, a long history of empirical literature has demonstrated that (Heckscher–Ohlin-type) endowment-based models of trade have a hard time in explaining observed patterns of trade. Does this mean that using the theoretical framework provided by such models in an attempt to answer the above questions is necessarily wrongheaded? We argue it is not. Indeed, it seems difficult to think of a world where the uneven distribution of factor endowments across countries does not matter at all for international inequality and trade. And since labor migration as well as capital movements and capital accumulation by definition alter these endowments, the theoretical perspective of endowment-based trade models should reveal important insights into the *interrelationship* between the two types of factor movements and trade.

One of the reasons for the empirical failure of endowment-based trade models has to do with the role and treatment of time. Empirical tests of such models typically look at cross-country patterns at some given point in time. It is all too obvious that this does not allow one to trace out how changes in endowments over time, whether through international movements or accumulation, affect trade, and vice versa. Whether or not we find the effects that play out over time also in cross-sectional perspective depends on the speed of adjustment, particularly in movements of factors across sectors and between countries. Given the assumptions of endowment-based models, i.e., mobility of at least some factors across sectors and complete immobility of factors between countries, one should not be too surprised that their predictions find only limited support in cross-section data.[6]

The aim of this chapter, however, is not to test theory or predicting detailed empirical patterns, but to develop an understanding of the role that the interaction of migration, trade, and capital movement as well as capital formation plays for the evolution of income inequality and convergence between countries, and to identify patterns of causality between these essential forms of economic globalization. Given this aim, a reasonable first approach is to remain agnostic about differential adjustment speeds and simply explore the welfare, inequality and convergence effects of any one of these globalization phenomena in isolation, and then to explore what these effects imply for the underlying incentives for the other forms of globalization. This is what we do in Section 3 of this chapter, relying on a simple, yet fairly general neoclassical model abstracting from the time and cost of adjustment. Obviously, this does not lead to a clear prediction on how trade, migration and capital movements evolve through time, and on the exact line of causation. But, as we hope will become evident below, it nevertheless produces important insights.

Any treatment of the interrelationship between migration, trade and capital movements and accumulation would be highly incomplete if restricted to the view of endowment-based trade models. These models mostly assume an identical technology in all countries. Perhaps more importantly, in their mainstream versions they also assume that this technology is convex, thus ruling out economies of large-scale production. It must be expected that non-endowment-based determinants of trade and/or the presence of economies of large-scale production profoundly affect the interrelationship between our three forms of globalization. In Section 3 of this chapter, we will therefore also extend our analysis in order to examine the interaction between trade and migration in the so-called models of the new economic geography where increasing returns to scale play a prominent role. These models also introduce costliness of trade, an element to which we shall return in more detail in Section 6 of the chapter.

[6] See Leamer (2012) for an excellent treatment of this issue.

Section 3 takes a static view and is, therefore, restricted to capital *movements*. In Section 4 we go further in exploring the relationship between migration and physical capital *accumulation* in a dynamic model that focuses on costly adjustment. In doing so, we shall further explore the implications of agglomeration effects for migration in the context of physical capital formation. In Section 5, the dynamic analysis shifts to human capital formation and the role of high-skilled migration, focusing on product innovation, i.e., the introduction of new products, as well as quality improvements of existing products. A further important issue dealt with in Section 5 is the potential of brain drain through emigration of high-skilled labor.

Section 6 explores the interrelationship between international migration and trade, focusing on the costliness of trade. While trade costs are to some extent taken up already in Section 3, they rise to prominence in Section 6, with a view on geographic and cultural distance between countries. This focus is a characteristic element of recent trade literature, which has led to a renaissance and refinement of the so-called gravity approach to the explanation of trade patterns. In addition to explaining the detailed *pattern* of world trade among many countries, this approach has proven a very handy tool for attempts to quantify the importance of such things as the use of a common currency or membership in the WTO for the *volume* of trade between two countries. In Section 6, we use it to investigate the role that migrant networks play for trade.

Before turning to theoretical analysis in Sections 3–6, Section 2 provides a brief tour through the history of international migration and pertinent issues. And before turning to Section 2, we now take a very rough look at the data in search for indications of causality. Figure 18.3 plots different countries' net migration rates over five-year-periods from 1960 to 2010 against five-year-averages of annual growth rates of the same countries' merchandise trade (exports plus imports). The top panel of figure 18.3 (henceforth 18.3a) plots lagged net migration rates against trade growth, while the bottom panel (henceforth 18.3b) plots net migration rates against lagged trade growth. Lags are meant to allow for lagged responses. Both figures plot the entire pooled data. Given what we have seen above, we expect positive growth rates for trade for most data points. For negative causal effects, we would expect strong emigration or immigration in Figure 18.3a to be associated with low subsequent increases in trade. Conversely, in Figure 18.3b we would expect strong increases in trade to be followed by low emigration or immigration rates. For positive causal effects, in Figure 18.3a we would expect strong (lagged) net emigration or immigration to be associated with, or cause, high growth rates of trade. By the same token, if trade causes migration, then we would expect strong growth rates in (lagged) trade to be associated with high emigration or immigration rates in Figure 18.3b. It is relatively obvious that Figure 18.3 does not support causality in either direction.

Figure 18.4a and b repeat this little empirical exercise for migration and capital flows. In Figure 18.4a, negative causal effects would imply that high emigration rates would be associated with low capital exports, i.e., low values of the current account, and vice versa for high immigration rates. In Figure 18.4b, it would mean high capital exports to be

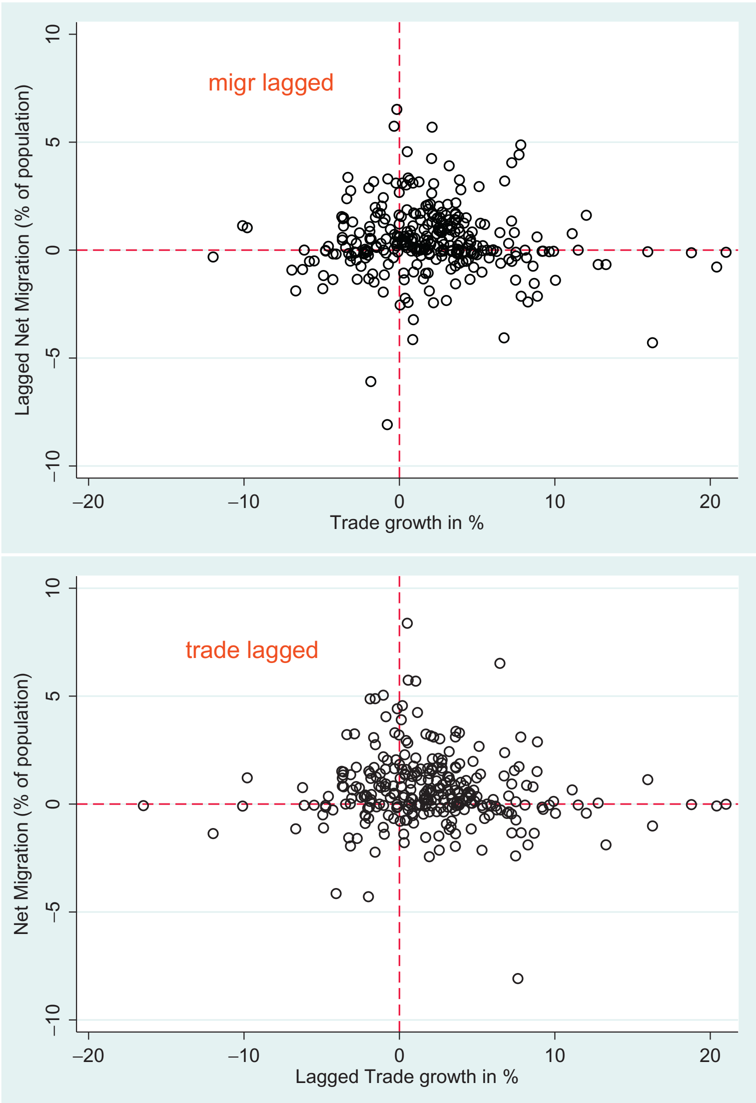

Figure 18.3 Net migration and trade growth across countries.

associated with low immigration rates, and vice versa. In contrast, positive causal effects would be indicated by high (lagged) capital exports being associated with high emigration rates, and vice versa. Again, these figures do not indicate any kind of causality.

2. A BRIEF TOUR THROUGH HISTORY AND ISSUES

Before jumping into a focused theoretical and empirical analysis, we offer a brief tour through recent history and issues, aiming at a theory-guided comparison between the

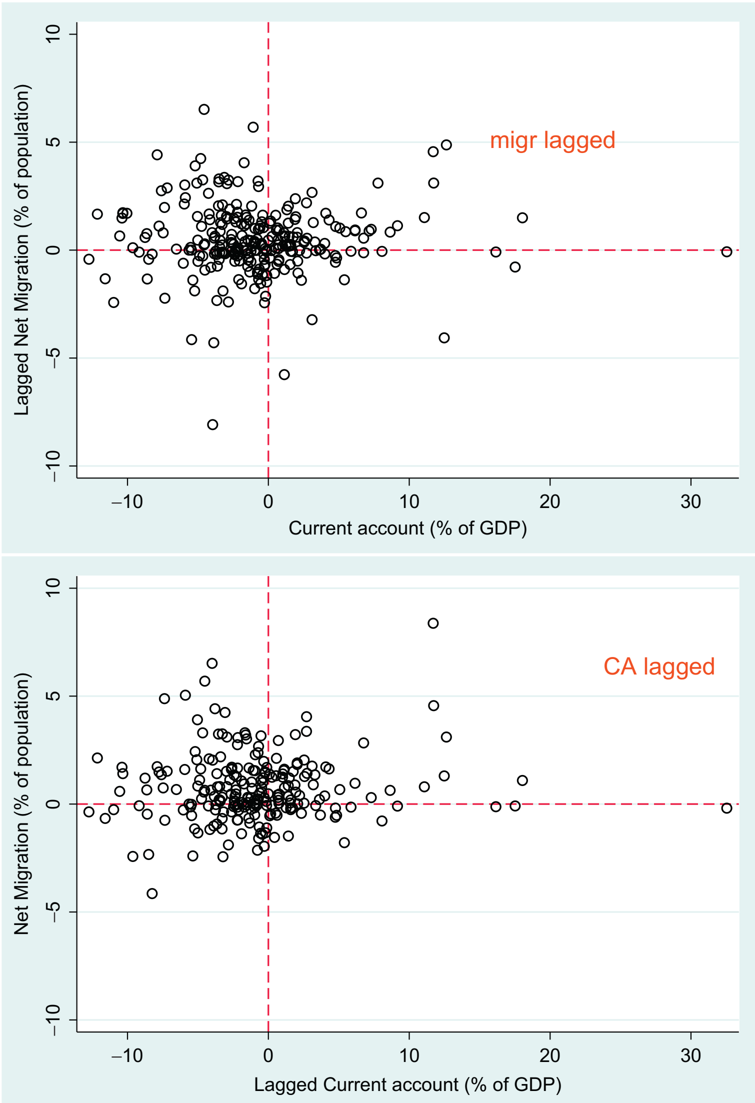

Figure 18.4 Net migration and capital flows across countries.

migration that took place during the first wave of economic globalization in the second half of the nineteenth century until the Great War and the migration that took place 100 years later.

2.1 Mass migration of the nineteenth century

Although human history is a history of migration, it was not until modern times that massive flows of migration have occurred in short periods of time, spanning no more than the length of human life, and in search of a better life. The period from 1850 to 1914 has witnessed more than 55 million people migrating, mostly from Europe to the Americas, in response to the combination of huge real wage gaps between the two sides of the "Atlantic economy" and a dramatic fall in the cost of ocean travel. Compared to earlier episodes of migration, this era, which has become known as the "era of mass migration", was characterized by three important novel features. The first was that migration took place between nation states. The second was that it has caused strong international convergence of workers' earnings perspectives between the sending and receiving countries. And the third was that it has changed the degree of inequality within these countries, favoring (harming) workers who were close substitutes to migrants in sending (receiving) countries.[7]

The model of trade that most economists would invoke in order to explain the aforementioned income convergence and inequality trends observed during 1870–1910, both between and within different countries of the "Atlantic economy", is the well-known factor proportions theory of trade, due to Eli Heckscher and Bertil Ohlin. According to this theory, commodity trade among countries with different factor endowments is indirect factor trade. A country's exports embody its abundant factors, while its imports embody its scarce factors — the so-called Heckscher–Ohlin–Vanek proposition. Thus, indirect factor trade (through the factor content of goods traded) tends to level out differences in factor scarcity across countries, leading to factor price convergence. Moreover, according to the famous Stolper–Samuelson theorem, it leads to a change in income distribution within countries, harming a country's scarce factors and favoring its abundant factors.[8] But what is true for indirect factor trade should also be true for direct movements of labor and capital. Hence trade and migration or capital movements should be substitutes for each other.

This view of trade and factor movements has strong policy implications. In particular, it implies that restricting trade in order to avoid unwelcome income distribution effects might be frustrated unless migration is restricted as well. It is interesting to note that the policies observed during the era of mass migration between 1870 and 1910 do not square

[7] For extensive documentation of these aspects of the first era of mass migration, see Williamson (1997), Hatton and Williamson (1998, 2005), Aghion and Williamson (1998), and O'Rourke and Williamson (1999). The numbers are huge. To give just a few examples, the estimated labor force reduction of the two biggest sending countries over the period 1870–1910, was 45% for Ireland and 39% for Italy, with an estimated positive impact on the real wage for emigrant-competing workers equal to 32% and 28%, respectively. The labor force increase of the two biggest receiving countries over the same period was 86% for Argentina and 42% for Australia, with an estimated negative effect on the real wage rate for immigrant-competing workers equal to 21and 15%, respectively. These latter countries have experienced a sharp increase also in internal inequality, with the wage-to-rental ratio falling to a quarter (one-fifth) of its initial level in Australia (Argentina). In contrast, European sending countries have experienced a sharp increase in this ratio during the same period; see O'Rourke and Williamson (1999, Chapter 9).

[8] See Feenstra (2004a) for a convenient survey of these propositions also for higher dimensions.

with this implication. Most of the labor receiving countries were pursuing protectionist trade policies in the 1850s, at a time when they were still vastly open to immigration. On a general level, this combination of policies appears at odds with a Heckscher–Ohlin narrative of nineteenth century globalization.[9]

This narrative also fails to convincingly explain the observed patterns of trade and factor movements. As indicated above, the factor proportions approach implies that trade and factor movements are substitutes. It thus implies that factor flows would reduce the incentives for trade, and vice versa. In contrast, the nineteenth century has witnessed significant surges of trade and migration at the same time. Moreover, the era of mass migration was also an era of large capital movements. Interestingly, against the backdrop of the factor proportions approach, for a large part labor and capital was moving in the same direction, i.e., to the "Atlantic economy" and the Scandinavian periphery, where labor appeared as a scarce factor (therefore drawing migration), rather than labor abundant places like Asia or poorer parts of Europe.[10] There are at least two possible explanations for this. First, substitutability holds for given barriers to trade and factor movements. In contrast, as already indicated above, the nineteenth century has witnessed a huge reduction of the cost of transport, particularly sea transport, which acted as a key barrier to both trade and migration.[11] The transatlantic telegraph cable installed in 1858 had a similar effect on capital movements. With a simultaneous fall in barriers to all forms of globalization, a simultaneous surge of trade as well as factor movements cannot, of course, be taken as violating Heckscher–Ohlin-type substitutability.

The second explanation has to do with technology. The factor proportions approach assumes away technological differences across countries, which is questionable empirically. To see what relaxing this assumption means in the present context, consider a highly stylized neoclassical model of the world economy with two types of goods and two countries. Suppose that one country has a Hicks-neutral technological advantage over its trading partner in labor-intensive goods, and assume that this is the only asymmetry between countries. In such a model, any trading equilibrium that satisfies the law of one price on goods markets will feature a higher wage rate in the technologically superior

[9] Hatton and Williamson even speak of a policy paradox; see Hatton and Williamson (2005, Chapter 8). One of their explanations of this paradox is that migration flows in the mid-nineteenth century were still regarded as quite welcome due to a self-selection effect (towards high-skilled labor) that was regarded as favorable for the receiving country. Any selection effect present in the labor inflow is certainly easier to observe than the labor content of trade for different types of labor. However, a skill-biased selection effect was no longer observed towards the end of the century, whence immigration restrictions started to surge as well; see O'Rourke and Williamson (1999, Chapter 10).

[10] See O'Rourke and Williamson (1999, Chapter 12). This is an early instance of what is presently known as the "Lucas paradox"; see Lucas (1990).

[11] See again O'Rourke and Williamson (1999, Chapter 3).

country than in the lagging country.[12] Allowing for factor mobility, an inflow of labor into the high-wage country will add a Heckscher–Ohlin rationale for further trade, since immigration will turn the technologically superior country into a relatively labor abundant economy as well. This is the essence of a result demonstrated by Markusen (1983), which has sparked off a series of theoretical papers addressing conditions under which trade and factor flows are complements, rather than substitutes. We do not want to pursue this question any further, but simply want to point out that allowing for differences in technology serves to reconcile the factor proportions approach with the pattern of trade and factor movements that we observe for the nineteenth century.[13]

A technology-driven scenario also allows us to invoke familiar factor proportions logic to explain the observed convergence of factor prices. Remaining within the aforementioned stylized economy, the inflow of foreign labor will eventually wipe out the high-wage advantage afforded by superior technology in the labor-intensive sector. Interestingly, in such a scenario, it takes more labor movements to achieve complete international convergence than would be the case with endowment-based trade alone. As pointed out by Markusen (1983), complete convergence in the sense of factor price equalization will occur only once the inflow of labor has driven the superior country to complete specialization. In this sense, the technology-driven scenario features slower convergence of factor prices.

One may question whether technological advantage is a plausible paradigm to apply for nineteenth century migration to the Atlantic economy and for trade between the new and the old world. A more plausible line of argument would invoke a different driving force, namely abundance of land in the new world. But intuitively, as regards wages and capital rentals, abundance in a third factor (land) should act pretty much like technological superiority in a two-factor environment (capital and labor). Formally, with three factors and two goods, factor prices are no longer determined by goods prices alone (through the pricing equations, or zero profit conditions), but are also determined by endowments (through the full employment conditions). As a result, even with the law of one price in full force through free trade, factor prices will be different across countries if their endowments with land are different. But the same is true if their technologies are different and endowments are the same, as in the model invoked above. O'Rourke and Williamson (1999) outlined a model with three factors and two goods where land is specific to food production and capital is specific to manufactures with labor, and where the outcome is trade as well as labor and (potentially) capital movements to the land abundant economy.

[12] It is very likely that the two countries will have different factor prices also in the autarky equilibrium. But depending on the relative strengths of the income and substitution effects the autarky difference in factor prices can go either way. In the knife-edge case of Cobb–Douglas preferences the autarky equilibrium would have a lower relative price of the labor-intensive good, without any difference in factor prices across countries. In this case, we could say that the entire difference in the wage rate that arises in the free trade equilibrium is *caused by trade*.

[13] For a more detailed treatment of possible interpretations of complementarity against the backdrop of the era of mass migration, see O'Rourke and Williamson (1999, Chapter 13).

2.2 Characteristics of modern migration

All forms of globalization were hit by a strong backlash in the first half of the twentieth century, to be followed by a gradual recovery of globalization, first through tariff liberalization after World War II. The recovery features a distinct sequence, starting with a revival of trade under the GATT in the 1950s, to be followed by a surge of migration starting in the 1960s and, finally, by the abandonment of capital controls subsequent to the breakdown of the Bretton Woods system in the 1970s. We are now into the fourth decade of what may be called the second wave of economic globalization, again covering trade, migration, and capital flows. This second wave of globalization differs from the first wave 100 years ago in many important respects.

2.2.1 Migration patterns

As regards the regional pattern of migration flows, Hatton and Williamson (2005) spoke of "seismic shifts" between the first and the second wave of globalization. Perhaps most significantly, some of the Western European countries changed from sending to receiving countries, with a lot of migration also taking place within Europe from the south to the north. Conversely, Latin American countries changed from receiving to sending countries, mostly sending emigrants to North America. And finally, Africa and Asia have appeared as important source countries of migration, while the Persian Gulf has appeared as a new destination country. These shifts mainly reflect differential evolution of income per capita in different parts of the world, which were exogenous to migration but have had important consequences for migration incentives. In addition, in some cases major political events have had the important consequences of opening borders to emigration, the most important case in point being the fall of the Iron Curtain in 1989/90, which has led to significant east–west migration within Europe.

In the era of mass migration yearly (net) immigration rates close to and above 1% of the population were quite common. In the past decades, such net immigration of this magnitude was the exception, not the rule. Looking at 13 countries classified by the OECD as receiving countries over the period 1956–2009, we observe no more than 20 instances (out of 702) where net annual immigration rates were above 1%.[14] But the receiving countries of the second wave have typically had lower growth rates of the labor force than the receiving countries in the era of mass migration, at least towards the end of the period considered. For instance, Hatton and Williamson

[14] The one country that has experienced an extended period of annual immigration rates above 1% was Spain, which, oddly, is listed as an emigration country by the OECD. See the dataset underlying OECD (2011), URL: http://dx.doi.org/10.1787/888932446759. Setting the threshold at 0.5%, a value almost always surpassed by immigration countries of the nineteenth century, the number of such cases in the late twentieth century increases to 162, which is still less than a quarter. See also Felbermayr and Kohler (2006b).

(2005) calculated that the contribution of immigration to the labor force growth in the US in the 2000s was comparable to the peak during the first mass migration.

2.2.2 Selection effects

In the era of both mass migration in the nineteenth century and globalization in the late twentieth century the source country composition of migration flows has changed over time. Any such change is likely to also affect the skill composition of the aggregate immigration flow of receiving countries, because each sending country sends its own distinct *selection* of migrants.[15] Different sending countries have different skill distributions of their respective populations, and the decisions to migrate or stay typically led to a selection of migrants that is no random draw from the population. Hatton and Williamson (2006) showed that in both waves of globalization this selection effect has worked towards a deterioration over time of the "labor market quality" of US immigrants, relative to the domestic workforce. This trend was aggravated by a trend towards a higher skill level of the domestic labor force in the receiving country. It is difficult to compare nineteenth and twentieth century migration with respect to the relative "labor market quality of migrants", but the negative trend was more pronounced in the twentieth century than in the nineteenth century.[16]

Migrant selection effects have been extensively addressed in modern literature. Perhaps the most well-known explanation of selection effects is the Roy–Borjas model, where earnings of workers in both the sending and the receiving country are stochastic and where the selection of migrants is driven by the variance and covariance of the earnings distributions in the two countries. For instance, a positive selection effect, i.e., a higher expected earnings in the group of migrants than in the entire population of the sending country, occurs if the covariance of earnings in the two countries is sufficiently high and the variance of earnings is higher in the receiving country than in the sending country.[17] A skill-biased selection effect then follows, if we assume that the distribution function for earnings follows from a skill distribution function among workers plus some earnings function translating skills into earnings. The intuition for the aforementioned condition is clear. The fatter upper tail of the earnings distribution in the receiving country draws highly talented people into migration, while a fatter lower tail deters less skilled people from leaving. But the condition may or may not be satisfied.

[15] For instance, Friedberg and Hunt (1995) calculated that among US immigrants arriving 1980–1990 the fraction with a lower than high-school level of education was 76.1% for immigrants from Mexico and 48.4% for immigrants from other Latin American countries, compared to 19.3% for immigrants from Europe and 26.4% for Asian immigrants.

[16] For details, see Hatton and Williamson (2006). They argued that the long-run trend of a falling "labor market quality" has contributed significantly to anti-immigrant sentiment already in the nineteenth century.

[17] See the classic paper by Borjas (1987) and Borjas (1999).

The fundamental message delivered by the Roy–Borjas model is that a positive skill-based selection effect is by no means a foregone conclusion, even if migrants are maximizing expected incomes. However, a recent paper by Kreickemeier and Wrona (2011) demonstrated a further channel for a positive selection effect that may arise if individual skills are unobservable. If migration is costly, then the highly skilled may have an incentive to reveal their skills through migration, provided that the status of being a migrant is observed. Assuming an "O ring technology", they showed that the equilibrium involves two-way migration with a positive skill-based selection effect. This effect still derives from an earnings-driven migration decision. Different selection effects will arise if migration is driven by concerns other than expected income.[18] For instance, Fan and Stark (2011) argued emigration may be a way to reduce the amount of stigmatization that derives from working in a certain sector of the economy, because doing so in a foreign economy may cause less stigmatization. If individuals differ with respect to their sensitivity to stigmatization, then those with a high sensitivity will select themselves into emigration, while those with a low sensitivity will stay. Whether or not this is coupled with a skill-based selection effect then depends on the correlation, if any, between the skill level and the sensitivity to stigmatization.

Why are selection effects important? For two reasons. First, policymakers are preoccupied with selection. Sending countries are worried about a skill bias in emigration, because the associated loss in human capital may impair development and growth. Receiving countries typically aim for a skill bias in their immigration flows because of specific "labor market needs" or because of a more general presumption that an inflow of talent is beneficial. And secondly, as we shall see in more detail in Section 3, the composition of the migration flow determines its effect on income distribution as well as welfare. A general conclusion from the theoretical literature is that heterogeneity among individuals will almost always generate a migration flow that involves a selection effect. But this need not be a skill-biased selection, as feared by sending and hoped for by receiving countries. It is therefore not too surprising that the empirical literature on selection effects has so far produced mixed results, both for the era of mass migration in the nineteenth century and the late twentieth century.[19]

2.2.3 Restrictive immigration policies

Modern migration is characterized by fears of unwelcome labor market effects in receiving countries. Such fears have developed towards the end of the nineteenth century as well, particularly in the US, but it was not until the 1920s that immigration restrictions

[18] See Stark (1993) for a more elaborate analysis of the emigration decision that goes beyond expected earnings.

[19] See Hatton and Williamson (2005, 2006) for the nineteenth century, and Borjas (1987), Chiswick (1999, 2000), Chiquiar and Hanson (2005), Grogger and Hanson (2011), and Moraga (2011) for evidence on the late twentieth century.

were imposed by many countries.[20] After World War II, immigration countries have repeatedly changed these restrictions, but the policy of quantitative restrictions on immigration was never abandoned. Despite economic analysis suggesting favorable welfare effects for natives as well as econometric evidence questioning the widely held belief that immigration causes wage pressure for natives, a generally restrictive policy stance by destination countries still marks the global migration landscape of today.[21]

Restrictions are mostly quantitative in nature and they are often highly selective, tailored to domestic "labor market needs", however vaguely defined, and mostly aiming at particular skills of migrants that are deemed in short domestic supply. Also, receiving countries of today typically have relatively large welfare states, and restrictions are often aimed at avoiding additional welfare-state cost for public budgets of the receiving country. By and large, quantitative restrictions have been binding. Hence, unlike the mass migration of the nineteenth century, international migration flows of the late twentieth century have been determined to a large extent by receiving countries' restrictive immigration policies. With little exaggeration, one can state that migration today is primarily seen and discussed as *im*migration, whereas in the nineteenth century it was mostly seen as *em*igration.

The restrictiveness of present day immigration policies is perhaps best realized by looking at the amount of illegal immigration and the cost incurred to jump immigration barriers. The two most important destinations for modern migration are the US and the EU. An estimated 4% of the US domestic population (and more than a third of its immigrants) are illegal residents, about 76% of them Hispanics.[22] In Europe, numbers are harder to come by, but the annual inflow of illegal immigrants at the beginning of the past decade was estimated to be of the same magnitude in the US and the EU, at roughly 0.15% of the population.[23] With a total inflow of less than 1%, this is substantial. The cost of jumping restriction is sizable too. Between 1993 and 2007, around 7000 people have died trying to get into the EU.[24] Naturally, the "quota rent", defined as the money cost incurred by migrants in order to migrate illegally, as compared to legal migration, is difficult to measure, but the evidence available is alarming. According to a New York Times report, immigration authorities in Ecuador, Mexico, and the US have estimated this rent to be around $20 bn per year at the beginning of the past decade.[25] For the EU, the figure is estimated at €4 bn.[26]

[20] See Hatton and Williamson (2006).
[21] We turn to the welfare effects on natives and the econometric evidence regarding wage effects in somewhat more detail in the next section.
[22] See Passel and Cohn (2009) and Hanson (2009b). For an economic analysis, see Hanson (2006).
[23] See "The best of reasons", The Economist, 31 October 2002.
[24] The number of people who have lost their lives trying to cross the border from Mexico into the US between 1997 and 2007 is larger than the number of people who have lost their lives trying to get through the Berlin Wall during its entire existence between 1961 and 1989; see Legrain (2009, pp.29 and 34).
[25] See "By a Back Door to the U.S.: A Migrant's Grim Sea Voyage", New York Times, 14 June 2004.
[26] See "Decapitating the snakeheads", The Economist, 5 October 2005.

2.2.4 Migration and development

A further characteristic that sets the twentieth century globalization apart from the era of mass migration is that migration is now an important element in global development policy. Migration from poor to rich parts of the world is perceived as a powerful vehicle to alleviate world poverty. However, as opposed to many other development policies, migration not only reduces inequality, but at the same time is likely to improve efficiency of world factor use. The reason is that existing gaps in factor earnings between rich and poor countries of the world indicate differences in marginal productivity and, thus, potential efficiency gains from international migration.

Numbers suggest vastly more gains from further migration than from trade liberalization. Based on the Wages Around the World dataset, Freeman (2006) compared wage gaps within occupations, taking the bottom and top 20 percentage points of the worldwide distribution for the period 1998–2002. Converting to PPP, he obtained bottom-to-top ratios ranging between 0.139 and 0.286. Compared with the wages at the beginning of mass migration in 1870, these are very large gaps. For instance, based on the real wages reported by Taylor and Williamson (1997) the average for European sending countries was no less than half the average in 1870, and it rose to 53% by 1910.

Large income gaps suggest large potential gains that workers may derive from migration. The World Bank has used its LINKAGE Model in order to estimate the gains from an "enhanced migration" scenario, which increases the share of migrant workers (from poor countries) in high-income countries from 6% to 8% (from 7.8% to 10.5% for low-skilled and from 2.2% to 5% for high-skilled workers). Migrants are estimated to enjoy an income gain of about 600%.[27] This seems like an enormous gain, but judged from the above-mentioned wage gaps they would seem plausible.

However, these income gaps are misleading, if they reflect differences in human capital embodied in the worker, which will not change through migration per se. Other principal explanatory factors for wage gaps are TFP and (non-human) capital per worker, and it is these factors that the worker will immediately benefit from once moving. Hendricks (2002) decomposed observed 1990 wage gaps to the US for a large sample of countries. For low-income countries, defined as countries with an observed gap larger than 60%, the average of observed country-specific income gaps is 82.3%. Differences in countries' physical capital endowments are able to explain a very small part of this gap, namely 17.6 percentage points.[28] Adding observed worker skills increases the explained

[27] See World Bank (2006).

[28] The small contribution of differences in capital stocks to the explanation of wage gaps can be seen as an explanation of the "Lucas paradox"; see Lucas (1990). The basic "Lucas calculation" attributes the entire wage gap to a gap in the capital stocks per worker, and it typically comes up with implausibly low capital stocks per capita in low-wage countries and a correspondingly high marginal return to capital. The other two explanatory factors mentioned above, then, are responsible for why the difference in the marginal return to capital is much lower.

gap to 46.9 percentage points, which leaves a residual, unexplained gap equal to 35.4 percentage points. Attributing this residual to TFP, the gain in marginal productivity through migration, which derives from TFP gains and increased capital per worker, reduces to $1/(1-(0.354+0.176))=2.13$ (from $1/(1-0.823)=5.65$).[29] These number are admittedly somewhat outdated, but they serve very well to make a fundamental point. Estimates of potential welfare gains from international migration, such as the above mentioned World Bank estimate of a 600% income gain accruing to migrants, are overly optimistic in that they take existing wage gaps as indicating the gains in marginal productivity of migrants to be earned by migration.

2.3 Modern migration, trade, and income distribution

How do the factor price and convergence trends observed during the second wave of globalization compare to the first? Research on the nineteenth century trends in income distribution has focused on labor income relative to non-labor income, whereas the literature on migration and trade during the past couple of decades has focused on wages of skilled relative to unskilled labor, the so-called skill premium. Specifically, it has often been pointed out that the past decades have seen an increase in factor price inequality to the disadvantage of low-skilled labor. As pointed out above, the era of mass migration has seen something like this happening as well, but only in receiving countries of migration, whereas in sending countries the opposite trend was observed. In contrast, in the late twentieth century the trend has been observed for almost all countries at the same time.[30] This alone should caution against any expectation that the distributional effects of trade and migration in the second wave of globalization can be explained as the outcome of Heckscher–Ohlin-type mechanisms. For trade, this mechanism operates through the above-mentioned factor content of trade, for migration it works through a direct change in the domestic labor supply. In either case, a country receiving a certain factor, say unskilled labor, whether directly through labor flows or indirectly through the labor content of trade, should experience a wage effect that is the opposite of that in a country sending this factor. In the empirical literature, trade and migration have mostly been analyzed separately, and with different methodological approaches.

2.3.1 *Trade*

There is a voluminous empirical literature aiming to quantify the explanatory power of trade for the relative (and sometimes even absolute) decline in wages for low-skilled

[29] The numbers are from Table 1 in Hendricks (2002). For the five lowest income countries, the reduction is from $1/0.058=17.24$ to $1/(1-(0.406+0.196))=2.51$.

[30] The trend has been documented many times, and we do not go into details here. For more details, we refer to Feenstra (2004b, Chapter 4), which also includes a literature survey. For a more recent survey, see Harrison et al. (2011).

workers that has been observed over the past three decades. Given the fundamental change in the nature of trade during the twentieth century and given that the wage trends have been observed equally in almost all countries, it shouldn't be too surprising that the evidence indicating a significant explanatory role of trade along the factor content logic is weak. Indeed, the consensus reached in the empirical literature towards the end of the 1990s was that the explanatory potential of trade is rather small, and that the wage trend for the larger part is a story of technological change.[31]

However, estimating the wage effects of (an increase in) trade is fraught with methodological problems.[32] One of the problems is aggregation. The level of disaggregation is restricted by the need to observe production data as needed to calculate factor contents. Observed factor contents are thus likely to mask vertical specialization within certain industries, based on skill-intensity differences between different parts of value added. If this is true, then trade may have a larger effect on relative factor demands than would appear from measurable factor contents, since seemingly skill-intensive exports from less developed countries may be an artefact of ignoring vertical specialization.[33] Moreover, from a single country's perspective a shift in the structure of vertical specialization, or offshoring, is equivalent to a change in technology, which *may* lead to higher demand for skilled labor. After all, trade is an inherent part of the technology that a country may use for turning its own resources into goods available for *consumption*. And a structural change in trade towards vertical specialization may conceivably have an effect similar to a skill-biased technical change that occurs in several countries at the same time, hence it may explain the *worldwide* nature of wage trends.[34] Summarizing more recent literature that duly takes into account this change in the nature of international specialization, the role of Heckscher–Ohlin-type trade as an explanatory factor for wage trends *may be* larger than the early consensus of the 1990s has suggested. But broad and robust statistical support of this hypothesis is still wanting, as it requires more refined data than are presently available.[35]

[31] For the trade literature, see for instance Richardson (1995), Krugman (1995), Borjas et al. (1997), and Cline (1997). Important papers demonstrating the pervasiveness of skill-biased technical change are Berman et al. (1994, 1998).

[32] One of the problems is that according to conventional models of trade, factor prices are linked to goods prices, and not necessarily to quantities traded, as emphasized by Leamer (1997). Moreover, neither goods prices nor trade volumes are exogenous; they are jointly endogenous to changes in trade barriers. The key issue, then, is whether calculating (changes in) the factor content of trade will deliver any information on the associated factor price movements. For an in-depth discussion of these problems, see Deardorff (2000), Krugman (2000), Leamer (2000), Panagariya (2000) and, more recently, Krugman (2008).

[33] This point has recently been made by Krugman (2008). As a case in point, Krugman invoked the computing industry.

[34] This point was made early on by Feenstra and Hanson (1996, 1997, 1999).

[35] This point has recently been made by Krugman (2008).

The modern literature proposes several new determinants of trade, which also imply novel channels for trade to affect the skill premium.[36] For instance, Epifani and Gancia (2008) pointed out that economies of scale as typically assumed in monopolistic competition models may explain the observed *worldwide* increase in the skilled premium as a result of enhanced trade, if the more skill-intensive sectors feature a higher degree of such scale effects. An increase in the skill premium may also arise as a result of selection effects, as emphasized by the recent literature emphasizing firm heterogeneity. In a recent paper, Burstein and Vogel (2012) argued that more productive firms often use a more skill-intensive technology. If this is true, then trade liberalization will entail factor reallocation towards more productive firms, if it weeds out the least productive firms.[37] In a recent paper Burstein and Vogel (2012) presented an in-depth analysis of this channel, including a calibration exercise in order to quantify the effect. They also proposed a generalization of the factor content calculation that takes into account inter-firm reallocation effects. Their results suggest a much larger role of trade in explaining the worldwide trend rise in the skill premium. In their scenario, trade liberalization explains up to 80% of this trend. Again, this may be expected to happen *worldwide*, not just at one end of the division of labor.

2.3.2 Migration

The relationship between immigration and wages would seem of more immediate importance for policy than the question of trade and wages, for two reasons. First, immigration is a directly observable change in labor supply. It is therefore more likely to become the *alleged* culprit of unwelcome wage trends. And secondly, immigration policy is much less restricted by international agreements than trade policy. If immigration turns out to be the *proven* culprit for unwelcome distributional trends, then restrictive immigration is the likely response. The intuition for this is open to common sense. If demand for a certain type of labor is downward sloping in the respective wage rate, then—other things equal—an increase of supply of this type of labor through immigration should put downward pressure on this wage rate.

However, this logic is less valid than it may seem at first sight. Trade theorists typically hasten to point out that in many sector economies the labor demand function need not be downward sloping for its entire range, but may have at segments. Assuming a standard neoclassical production function for each sector, this is the case whenever the receiving economy is (and remains within) a given cone of diversification—the well-known "factor price insensitivity" result, which is a logical corollary of the Rybczynski

[36] For a survey on the wage inequality effects of trade that includes novel theoretical approaches featuring non-Heckscher–Ohlin trade, see Harrison et al. (2011).

[37] This type of selection effect is a hallmark of modern trade literature, sparked off by Melitz (2003), that allows for firm heterogeneity.

theorem.[38] However, if there are many sectors differing in factor intensities, these at segments are very small. In the limit, with a continuum of industries, cones of diversification have zero measure, whence any labor demand function becomes continuously downward sloping. While this may seem reassuring in suggesting a well behaved aggregate labor demand function, it has the uneasy implication that different economies would be specialized in disjoint sets of industries with non-overlapping ranges of factor intensities, which seems counterfactual.[39]

The empirical literature has pursued different approaches to quantify the wage effect of migration, mostly looking at immigration rather than emigration. The so-called "area approach" applies regression analysis to exploit cross-sectional variation in the share of immigrants in different regions (e.g., cities, counties) of the receiving country. The aim is to estimate reduced form coefficients telling us how the equilibrium wage rates in regional labor markets respond to a change in this regional share of immigrants, controlling for Mincerian wage determinants and allowing for regional fixed effects. Early applications of this approach in the 1980s and 1990s have revealed very low coefficient estimates, with very low economic significance, i.e., accounting for only a small fraction of observed wage movements.[40]

Critics have pointed out that the "area approach" suffers from attenuation bias due to mobility of factors, both capital and labor, across regions. This has prompted researchers to look at economy-wide (as opposed to regional) changes in labor supply brought about by immigration, which obviously eschews the problem of cross-regional factor mobility. This approach, sometimes called the "nation approach", aims at estimating the elasticities of substitution between different types of labor, based on a standard production function. It was developed and first applied by Borjas (2003), and has since found applications for several other countries. Borjas (2003) distinguished between workers of different age and work experience, obtaining wage effects from US immigration that were much larger than those obtained with the "area approach". The baseline estimation results imply that the US immigration between 1980 and 1990, totaling about 10% of the population, has depressed wages paid to the "average" US worker by about 3.5%.

Other researchers have found somewhat smaller effects, both for the US and for other economies. A key question with this approach is whether or not immigrant workers are

[38] This implication of the Heckscher–Ohlin theory is emphasized in Leamer and Levinsohn (1995).

[39] This is described in detail in Dornbusch et al. (1980).

[40] For a survey of early results leading to this consensus, see Friedberg and Hunt (1995), Lalonde and Topel (1997), and Topel (1997). In Borjas et al. (1997), this approach is combined with a factor content calculation, leading to the conclusion of very moderate labor market effect of both trade and immigration. Studies that are notable and often mentioned because they exploit natural experiments are Card (1990) and Friedberg (2001). A more recent study following the "area approach" is Dustmann et al. (2005). For a more recent survey, see Hanson (2009a). All of these studies support the aforementioned consensus of very low immigration on wages earned by natives.

assumed perfect, or are allowed to be imperfect substitutes for native workers with the same labor market characteristic. Naturally, the wage effects will be larger if we assume perfect substitution.[41] A further question is whether we allow for an endogenous reaction of other inputs, particularly of capital. If we do, then the wage effects will naturally be lower than if we don't.[42]

It is somewhat difficult to summarize this literature on the wage effects of migration. The "nation approach" delivers somewhat larger effects than the "area approach", which suffers from attenuation bias, although the magnitudes of the wage effects found are still surprisingly low, given the size of the immigration-induced labor supply shock.[43] But the estimated magnitudes vary across countries. Indeed, an important conclusion to be drawn from this literature is that the wage effects of migration importantly depend on the type of migrants that a country receives. This lends additional relevance to the above-mentioned selection effects in international migration.

2.4 Modern migration and international convergence

The literature concludes that the "mass migration" during the nineteenth century has significantly contributed to international convergence of wage incomes. Taylor and Williamson (1997) have estimated the convergence effect by applying estimated elasticities of the labor demand functions to the measured cumulative migration flows for the various sending and receiving countries of the "Atlantic economy" between 1870 and 1910. This simply calculates the labor market effect of emigration and immigration respectively as an upward or downward movement on a well-behaved aggregate labor demand functions caused by a migration-induced labor supply shock, as for instance portrayed in Borjas (1999). The estimated real wage convergence is substantial. The reduction of the real wage in New World countries that this estimation attributes to immigration is 12.4%. The corresponding increase in Old World sending countries is 9.6%.

If we measure international wage inequality by the square coefficient of variation, how much of the actual reduction in wage inequality can be explained according to this simple calculation of migration-induced wage effects in sending and receiving countries? Taylor and Williamson (1997) pointed out that these calculations would imply a

[41] Borjas (2003) as well as Aydemir and Borjas (2007) assumed perfect substitution, while Ottaviano and Peri (2012) allowed the data to tell about the degree of substitutability between immigrants and natives. For a critical discussion, see Borjas et al. (2012).

[42] Borjas (2003) assumed a constant capital stock, while Felbermayr et al. (2010a) and Ottaviano and Peri (2012) allowed for endogenous reaction of the capital stock. Felbermayr et al. (2010) also allowed for unemployment, while all studies for the US assume full employment.

[43] A recent paper by Aydemir and Borjas (2011) demonstrated that the "national approach" suffers from attenuation bias due to sampling error. Re-estimating the "national approach" regressions on larger samples leads to larger (negative) wage effects of immigration.

migration-induced convergence that exceeds the convergence actually observed. Across all countries considered, measured inequality in 1910 was down to 72% of what it was in 1870. Undoing the migration effect through the aforementioned thought experiment takes inequality up to 7% *above* what is actually observed in 1870. Thus, focusing entirely on migration and assuming a simple downward-sloping aggregate labor demand function "overexplains" income convergence.

2.4.1 Capital

How can we interpret this somewhat paradoxical finding? A crucial point here is that such a partial equilibrium application of aggregate labor demand elasticity ignores two things that are at the core of this chapter: The first is that the immigration countries concerned were not just receiving labor, but capital as well. And the second is that they were trading economies. Ignoring both factors is bound to exaggerate migration-induced convergence. If capital flows into a country alongside labor, then the wage depression effect is mitigated if labor and capital are complementary factors in production, meaning that an increase in the employment of one increases the marginal productivity of the other. Applying this logic to all countries, the convergence effect attributable to labor movement alone is reduced, even in the calculation by Taylor and Williamson (1997).[44]

However, there is an issue of aggregation lurking here. The convergence picture almost automatically becomes much less clear-cut if we consider several types of labor. In a standard neoclassical technology with capital and labor as the only inputs, complementarity between these two inputs (and thus convergence) must prevail. However, even with neoclassical technology, if there are several types of labor and multiple labor movements, then the pattern of wage effects from a given pattern of factor supply shocks is no longer strictly tied down. We shall return to this in the subsequent section. As a result, even absent trade, convergence need not hold across *all* types of labor, or between any pair of factors. Specifically, going back to the calculations by Taylor and Williamson (1997), if the New World part of the "Atlantic economy" has experienced a positive supply shock through inward movements of all types of labor as well as capital, then all we can say from general equilibrium theory is that—loosely speaking—the factor price reactions and the resulting factor price changes must be negatively correlated, but this is perfectly consistent with international divergence in the price of any one factor that moves from one country to the other.

[44] Taylor and Williamson (1997) provided robustness checks of their convergence results, indicating that taking account of capital inflows that "chase" migrating labor the above-mentioned over-explanation of convergence disappears. See also the discussion in O'Rourke and Williamson (1999) and Hatton and Williamson (2005).

2.4.2 Trade

Perhaps more worryingly, however, this way of estimating the convergence effect of factor movements also ignores that the receiving countries were trading economies. Allowing for trade, inflows (outflows) of both labor and capital may be devoid of any factor price effects, if factor price insensitivity obtains (see above). This means that all countries absorb the factor supply shocks through inter-sectoral reallocation so as to maintain each factor's marginal value productivity. With constant goods prices, this implies Rybczynski-type reallocation among tradable goods industries. However, constant goods prices seem highly unlikely, at least if the initial migration is a response to international wage gaps. Hence, we must take a step back and ask why such wage gaps exist in the first place. In principle, such wage gaps may exist for three reasons: international differences in technology, relative labor endowments, and human capital embodied in workers.

Ruling out technology and human capital differences, initial wage gaps imply that countries are in different cones of diversification. This may be the outcome of endowment points lying outside the so-called factor price equalization region, or the outcome of trade barriers.[45] For constant goods prices, labor supply shocks then mean Rybczynski reallocations in sending and receiving countries that appear as something like mirror images of each other. But with disjoint cones of diversification, these reallocations would nonetheless cause disequilibria in world goods markets. For labor receiving countries, the reallocation is likely to cause excess demand of *their* more labor-intensive sectors, for sending countries it is likely to cause excess demand of *their* more capital-intensive sectors. Goods market equilibrium will thus require goods price changes that undermine factor price insensitivity in both countries. By the standard Stolper–Samuelson logic, labor receiving countries will see goods price adjustments that favor capital and harm labor, and the converse will hold for labor sending countries. Clearly, the outcome is international convergence.[46]

The argument can be extended to a case where we have several sending and several receiving countries of migration, as in the "Atlantic economy" of the nineteenth century. Notice that in this scenario migration (or more generally factor flows) may, but need not, enhance trade in the sense of the Markusen result mentioned above, since the reallocation is among disjoint sets of industries where the two countries are specialized. We

[45] On the factor price equalization region, see Dixit and Norman (1980); on trade barriers and cones of diversification, see Deardorff (1979).

[46] A somewhat similar scenario of comparative statics is described in Dornbusch et al. (1980). However, that scenario assumes an exogenous increase in one country's factor endowment for a constant endowment of the other, which is obviously different from factor movements. Our argument above is somewhat of a short-cut in that it ignores a likely shift in the margin that separates industries of specialization in the two countries.

can say that migration will have trade effects, but we cannot unambiguously state that the volume of trade will increase.

For the "Atlantic economy", the scenario is complicated by capital "chasing" labor, i.e., by simultaneous inflow of labor and capital. This cannot be explained by different relative endowments of trade-barrier-induced specialization in different diversification cones. We need to add technological differences. Specifically, the labor receiving countries must be attractive destinations for capital flows as well, because they have superior technology. But with TFP superiority, complete factor mobility would eventually depopulate the inferior economy. To avoid this, we need some type of counter-force, say some form of congestion. For the nineteenth century, O'Rourke and Williamson (1999) argued for the so-called "frontier hypothesis", which is equivalent to technological superiority which eventually peters out. Whatever the details, adding such capital flows to the above scenario of migration-cum-trade does not necessarily reinforce the conclusion of international *wage* convergence although convergence of incomes more generally seems a natural outcome.

What are we to conclude from all of this regarding the above-mentioned paradox of migration "overexplaining" nineteenth century convergence? The general message, not just for this historical episode of strong convergence, is that it seems futile trying to attribute convergence to either factor flows or trade, or to flows of some specific factor, say labor as opposed to capital. Depending on existing international barriers on markets for goods, labor and capital, a certain combination of trade and movements of one or both factors will be the *simultaneous* adjustment to some given initial disequilibrium in the sense of a violation of the "law of one price". The particular combination of trade and factor movements through time may reflect the sequence of historical changes in different types of barriers, but since all of them jointly represent a general equilibrium adjustment to the same disequilibrium, the change to some new equilibrium, say measured in terms of wage convergence, must similarly be considered as the *joint* outcome of both trade and factor movements. Attributing parts of observed convergence to either trade, capital movements or migration seems arbitrary.

2.4.3 Convergence through modern migration?

Mass migration in the nineteenth century, although arguably dominating the picture, must thus be seen as an integral part of an adjustment additionally involving both trade and capital movements, the exogenous shock being a vast reduction in both the costs of, and political barriers to, the movements of goods, labor, and capital. A first rough picture is obtained by comparing average real wages across sending and receiving countries of the "Atlantic economies" 29 in 1870 and 1910, as presented in Taylor and Williamson (1997). As already mentioned above, by 1910 the coefficient of dispersion measured as the ratio of the variance to the squared mean has fallen to 72% of what it had been in 1870. Hatton and Williamson (2005) described a more detailed pattern of convergence by looking at wages in several (sending) European countries, relative to a country-specific weighted average of wage rates in the corresponding destination countries of their respective emigrants. The data do not suggest convergence in all cases but in some cases the convergence was substantial, particularly for Nordic sending countries. In 1870, the unweighted average of this wage gap was 49%, rising to 53% by 1910.[47]

How does this compare to the second wave of globalization in the late twentieth century? Some convergence as a result of migration is implicit in the results obtained by some of the studies using the Borjas-type "nation approach" in order to look at emigration countries, in addition to the traditional focus on immigration. Thus, Aydemir and Borjas (2007) and Mishra (2007) applied this approach to Mexican data, obtaining estimates comparable in magnitude to those obtained by Borjas (2003) and Aydemir and Borjas (2007) for the US and Canada. Taken together these results imply income convergence. However, this is partial evidence, and the overall picture of estimation results for this approach does not support a wider generalization.

If international convergence is more difficult to describe empirically and perhaps a less plausible consequence of modern globalization on theoretical grounds, what certainly separates the present state of the world economy from that of the nineteenth century, both before and after mass migration, is the existing income gaps between potential receiving and sending countries of migration. As we have just seen, the wage rates in source countries of nineteenth century migration on average were about half the wages in destination countries. This level of international inequality pales against all evidence that we have for the outcome of twentieth century globalization. A very rough measure of the extent of international income gaps in the second wave of globalization is obtained by looking at the international distribution of real GDP per capita or national household expenditure per capita, each measured in purchasing power parities. Taking data from the World Bank *World Development Indicators*, and comparing the respective cut-off points for the bottom and the upper quartiles of the world distribution, we obtain numbers that are comparable to those reported for occupation-specific wages by Freeman (2006), taken from the NBER *Wages around the World* database (see above). We may refer to the figures already given in the introduction, where we have summarized international convergence by comparing the ratio of the 25th to the 75th percentile respectively of the world distribution of GDP per capita as well as private household expenditure over time. These ratios have barely increased from 1980 to 2012. For GDP per capita ratio in 2012 is a mere 0.136, up from 0.119 in 2012. For household expenditure, convergence is somewhat stronger, with a value of 0.185 in 2012, compared to 0.074 in 1980.

[47] This masks much more dramatic convergence for some of the emigration countries. For instance, Norway has seen arise from 25% to 50%, and Sweden from 36.7% to almost 60%. For details, see Hatton and Williamson (2005, Table 4.2).

3. A FACTOR PROPORTIONS VIEW ON MIGRATION AND TRADE

For a large part, international migration is a response to international wage gaps. Wage gaps, in turn, are primarily determined by three forces: technological knowledge and practice prevailing in different countries, national factor endowments (including human capital endowments), and trade between these countries. At the same time, international migration directly affects national endowments, as do international movements of capital. In addition, factor movements impact on trade as well as on the accumulation and availability of technological knowledge. Hence, while being determined by factor prices, factor movements are also influencing factor prices in both the receiving and the sending country, which raises the question of international convergence that we have touched upon in the previous section. It is perhaps fair to say that, generally, policymakers tend to view migration as a possible way towards international convergence. However, what makes both migration and trade the subject of controversial political debate is their influence on domestic wage inequality or, more generally, inequality of incomes. Typically, if two or more countries observe enhanced integration of markets, within each country there will be winners and losers. The key question for policy then is whether there are efficiency gains large enough for the economy as a whole to compensate losers, thus achieving a Pareto improvement (welfare gain).

In this section we briefly summarize the key messages that the so-called factor proportions approach holds about these questions. This approach assumes that international movements of goods, labor and capital are responsive to prices, and it focuses on factor endowment as a key determinant of goods and factor prices. Arbitraging on international differences in prices, trade and factor movements generates an international equilibrium where price differences are reduced to the costs of moving goods, capital or labor respectively across borders, plus the price equivalent of policy-induced barriers to such movements.

We need to acknowledge an important asymmetry between international migration and capital movements. International migration always means that existing stocks of labor are being relocated across countries. Thus, migration may be gainfully analyzed independently of population growth. In contrast, capital very rarely moves in the sense of a relocating existing physical capital. Instead, it moves in the sense of *new* capital being invested abroad. Hence, capital movements should preferably be analyzed in the context of capital accumulation. Arguably, the interrelationship between migration and capital accumulation as such is more interesting to analyze than the relationship between migration and international capital movements. The same holds true for accumulation of technological knowledge. We shall, therefore, turn to accumulation issues relying on dynamic models in separate Sections 4 and 5 below. In this section, we want to highlight the interrelationship between international migration and trade, which may usefully be done relying on a static model.

The model that we use in this section is mostly neoclassical in nature. It assumes full employment throughout. For a large part, it also assumes a constant returns to scale technology and markets with perfect competition, but we will relax the assumption of constant returns to scale and allow for moderate market imperfections towards the end of the section. Importantly, and in contrast to much of the literature, we shall look at marginal migration flows, with stocks of cross-border migrants in all countries inherited from the past. Migration will always be between trading economies. We assume that migration flows (changes in stocks of migrants) are driven by changes in barriers to migration, which may be migration costs or immigration restrictions. These changes are assumed exogenous and not dependent on the amount of trade or movements of other factors. Conversely, trade is assumed to be free of any barriers. In particular, changes in migrant stocks do not affect trade barriers. This channel will be taken up separately in Section 6 of the chapter.

3.1 A normative view on migration

A central tenet of neoclassical theory is that factor movements driven by international differences in factor returns increase worldwide efficiency of factor use and should, therefore, deliver welfare gains, provided factor returns reflect marginal value productivities. However, these gains typically accrue very unevenly to different people and different countries. First, there will be internal redistribution effects, as already mentioned above. But these are, in principle, common to both factor movements and trade. What sets gains from factor movements, particularly migration gains, apart from the classic gains from trade, is that, even absent any market distortion, it is typically not true that *both* receiving and sending countries may expect to achieve welfare gains that may be turned into Pareto improvements through suitable compensation schemes. This contrasts with the gains from trade result, which states that with perfect markets all countries can simultaneously gain from trade. The subsequent analysis will substantiate this point.

3.1.1 A simple yet general model

The fact that international migration usually involves movement of people complicates welfare calculations. Specifically, we need to make an assumption as to whether migrants' welfare should be considered as part of the receiving country's or the sending country's welfare, or none of both. In principle all three approaches are possible. The standard approach, however, is to treat migrants as part of the sending country's welfare. Although standard practice in most of the literature, this is a delicate assumption, since it is in direct contrast to the notion of integrating migrants into the host country society. But we do not intend to discuss questions related to integration or assimilation of migrants in the host country.

Given the prevalence of highly selective immigration restrictions that distinguish between several types of labor, and given the above-mentioned selection effects deriving from specific characteristics of different sending countries as well as from emigration decisions, it seems important to allow for more than one type of labor in the analysis. In what follows, we therefore allow for an arbitrary number of factors and goods, but we generally assume that the number of traded goods is larger than the number of primary factors. For tractability we assume two countries. There is free trade between these countries, which enforces the "law of one price" on goods markets. Both countries already have stocks of migrants, which reflect past migration leading to a trading equilibrium with wage differences in line with existing costs of, or policy barriers to, migration. An exogenous reduction in these barriers leads to further migration.

For simplicity, we first assume that both countries share the same technology, which is described by a GDP function, defined as

$$G(p, v) := \max_{q} \{p{\cdot}q, \text{ s.t. } (q, v) \in T(q, v)\}. \tag{18.1}$$

In this definition, $p \cdot q$ indicates a scalar product of vectors p and q, which denote goods prices and outputs respectively. The vector v denotes the quantities of factors *supplied* domestically in this economy. With factor movements, this is different from a country's factor *ownership*. $T(q, v)$ denotes the set of feasible output and input vectors, given the two countries' technological knowledge. Constant returns to scale imply a convex technology set. Moreover, we assume convex preferences, characterized by an indirect utility function $H(p, Y)$, where Y denotes aggregate income. Preferences are allowed to be different between the two countries.

The envelope theorem implies that in a frictionless competitive equilibrium the country's vector-valued supply function emerges as $q(p, v) = G_p(p, v)$, and its factor returns are $w(p, v) = G_v(p, v)$. In a similar vein, the economy's vector-valued demand functions may be written as $-(1/H_Y)H_p(p, Y)$, usually referred to as Roy's identity. From the fundamental properties of G and H, it follows that $G(p, v)$ is convex in goods prices p and concave in factor supplies v, while $H(p, Y)$ is quasiconvex.

We now consider a two-country world with countries A and B, assuming that both countries trade with each other and have cross-border *stocks* of different types of migrants. However, we assume that these cross-border stocks are *one-way* in nature, meaning that within a given type of labor, a country will not simultaneously have emigrants and immigrants. Moreover, we initially assume that there are no capital movements. We assume free and costless trade, so that both countries have the same prices for tradable goods, and for the time being we abstract from non-traded goods. Given identical technological knowledge for both countries, the underlying assumption is that free trade alone does not equalize factor returns across both countries, the reason being relative endowments that lie outside the factor price equalization region.[48] The implication then is that the two

[48] The factor price equalization region has been introduced by Dixit and Norman (1980) and further extended by Helpman and Krugman (1985).

countries are specialized in production, meaning that there is a limited number of goods, smaller than the number of factors, that are jointly produced by both countries.

We use V^A to denote the stock of factors owned by natives of country A, and the vector m_{AB} to denote the stock of immigrants from country A working in country B. Conversely, m_{BA} is the stock of immigrants from country B working in country A. We have $v_A = V_A - m_{AB} + m_{BA}$. Note that the vector V_A includes all factors, so that m_{BA} and m_{AB} are vectors of equal dimension that contain zeros for non-labor factors. By definition m_{AB} and m_{BA} only have non-negative elements, $m_{AB} \geq 0$ and $m_{BA} \geq 0$. Moreover, the one-way nature of migrant stocks implies that $m_{AB} \cdot m_{BA} = 0$. Obviously, we have $v^B = V^B - m_{BA} + m_{AB}$, and the GNPs of the two countries then follow as $Y_A = G(p, v_A) - w_A \cdot m_{BA} + w_B \cdot m_{AB}$ and $Y_B = G(p, v_B) - w_B \cdot m_{AB} + wA \cdot m_{BA}$ respectively. We assume V_A and V_B to be given, and we look at variations in the migrant stocks, i.e., migration flows dm_{AB} and dm_{BA}.

Both migrant stocks and migration flows are determined by an underlying no arbitrage condition on the two countries' wage rates. For simplicity, we do not want to incorporate any of the more complex migration decisions, such as the selection effects considered in Borjas (1987) or relative deprivation effects considered in Stark and Taylor (1991). Instead, we assume that migration decisions are based on direct wage comparisons. Suppose, then, that the cost of cross-border movement for labor of type l is proportional to its wage, denoted by w^l. Formally, a worker of type l from country B would not consider moving to country A if $w_A^l \leq \rho^l w_B^l$, where $\rho^l > 1$. Assuming that the migration cost is symmetric, l-type workers of country A would similarly not consider moving to B if $w_A^l \geq w_B^l / \rho^l$. Hence, if both conditions are satisfied, then no migration flows occur. As depicted in Figure 18.5, the two conditions together span a "*cone of no migration flows*" in wage space for labor of type l, with unique patterns of migration flows dm_{AB} and dm_{BA} outside this cone. As emphasized in the previous section, migration is of course hampered not just by migration cost, but also—and perhaps more importantly—by quantitative restrictions imposed by receiving countries. Thus the ρ values represent the cost-equivalents of such restrictions, in addition to the costs of migration.

What can we say about the relationship between the ρ values and the stocks of migrants m_{BA} and m_{AB} respectively? Stocks reflect past migration flows, hence without knowledge of the history of wage rates and migration cost, it is impossible to establish a connection between present wages and existing migrant stocks that is dictated by present migration cost. For instance, with wages as in point 1, we have an inflow of l-type labor, in point 3 we have an outflow of labor from A, and with point 2 we have none of both. In the subsequent analysis, we shall explore the comparative statics of migration flows dm_{AB} and dm_{BA}. As with migrant stocks, we now assume that migration flows are always one-way too. Within this model, such migration flows must be thought of as the outcome of lower migration cost. For instance, assuming point 2 in the figure as the initial

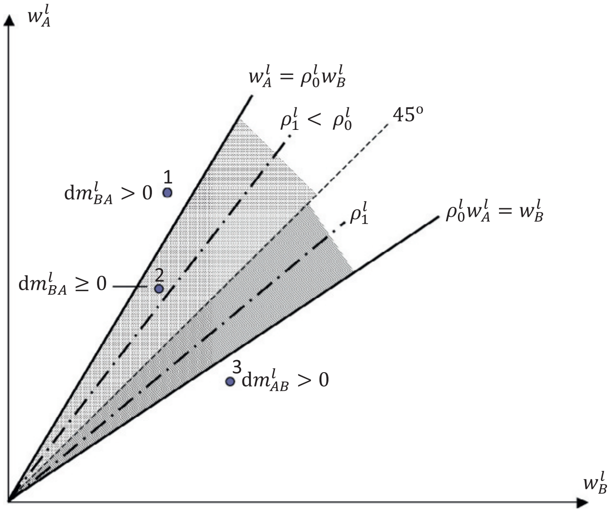

Figure 18.5 Cone of no-migration flows.

equilibrium, and assuming a stock $m_{BA}^l > 0$ to start with, a reduction from ρ_0^l to $\rho_1^l < \rho_0^l$ leads to further outflows of l-type labor from A to B. This might seem like a natural scenario, but the initial equilibrium at point 2 may well feature a stock $m_{AB}^l > 0$. As we have emphasized above, without knowing the history of wages and migration cost, we cannot say anything about the stocks. If point 2 in the figure has $m_{AB}^l > 0$, then a reduction from ρ_0^l to ρ_1^l still leads to $\mathrm{d}m_{BA} > 0$, i.e., an inflow of this type of labor into country A. This could be first-time immigration of country B's natives to country A, or it could be *return migration* of country A's natives from country B to A. The former would, however, lead to a two-way migrant stock, which we want to rule out. Hence, in such cases we must think of $\mathrm{d}m_{BA} > 0$ as return migration.

In this multiple factor setup, wage rates are simultaneously determined by *all* factor supplies. A change in w_A^l and w_B^l is brought about not just by migration flows in l-type labor, but by migration of *other* types of labor as well. Thus, returning once more to Figure 18.5, even without any change in the migration cost wages in the two countries may be changing, such that point 2 in the figure, where no migration incentive for type-l labor exists, moves to point 1, where such an incentive arises. We do not explicitly state any hypothesis about migration flows other than what we indicate in Figure 18.5. This means that we cannot say much about stability of the adjustment paths, or about international convergence. We shall return to this below.

3.1.2 Three welfare channels of migration

How does aggregate welfare of natives of country A and B respectively change upon a change in the migrant stocks? In the following, we answer this question through linear approximations around the initial equilibrium, whereby we simplify by setting $H_Y(p, Y) = 1$ for both countries. In the following, we shall use $m_A := -m_{AB} + m_{BA}$ to denote the *net* stock of country A's *im*migrants, and accordingly for $m_B \equiv -m_A$. Remember that we have assumed stocks to be strictly one-way in nature. The income of country A's net stock of immigrants is $w_B \cdot m_{AB} - w_A \cdot m_{BA}$. This is simply the net value of *stock migration* between countries A and B, evaluated at ongoing factor prices.

Turning to welfare changes, we now use $U_A = H^A(p, Y_A)$ to denote welfare in country A, and analogously for country B. We may then write $\mathrm{d}U_A = H^A_p(p, Y_A) \cdot \mathrm{d}p + H^A_Y \mathrm{d}Y_A$, with $\mathrm{d}Y_A = G_p(p, v_A) \cdot \mathrm{d}p + G_v(p, v_A) \cdot \mathrm{d}m_A - d(m_{BA} \cdot w_A) + \mathrm{d}(m_{AB} \cdot w_B)$, and accordingly for country B. After some straightforward manipulation, we arrive at the following two equations describing welfare effects of *flow* migrations $\mathrm{d}m_{AB}$ and dm_{BA}:[49]

$$\mathrm{d}U_A = G_v(p, v_A) \cdot \mathrm{d}m_A - \mathrm{d}(m_{BA} \cdot w_A) + \mathrm{d}(m_{AB} \cdot w_B) + M_A \cdot \mathrm{d}p, \tag{18.2a}$$

$$\mathrm{d}U_B = -G_v(p, v_B) \cdot \mathrm{d}m_B + \mathrm{d}(m_{BA} \cdot w_A) - \mathrm{d}(m_{AB} \cdot w_B) + M_B \cdot \mathrm{d}p. \tag{18.2b}$$

In the first of these equations M_A denotes the vector of net commodity imports by natives of country A. Formally, $M_A := H^A_p(p, Y_A) - G_p(p, v_A)$, where $H^A_p(p, Y_A)$ is the vector of commodity demands by natives of country A (Roy's identity) and $G_p(p, v_A)$ is the vector of country A's outputs (Hotelling's lemma). A corresponding interpretation holds for M_B. Notice that worldwide goods market equilibrium implies $M_A = -M_B$. The vector of goods price changes, $\mathrm{d}p$, must be thought of as determined by migration-induced changes in goods supply and demand on world markets. Goods price changes are often assumed away in models of migration, but given our assumption of unequalized factor prices with attendant specialization (see above), such changes in the terms of trade are an inevitable consequence of international migration. However, we need not explicitly solve for these feedback effects from goods markets, in order to derive some interesting results with the aid of this model.

For frictionless factor markets, we have $G_v(p, v_A) = w_A$ and $G_v(p, v_B) = w_B$, so that after suitable manipulation the above equations simplify to

$$\mathrm{d}U_A = (w_B - w_A) \cdot \mathrm{d}m_{AB} - m_{BA} \cdot \mathrm{d}w_A + m_{AB} \cdot \mathrm{d}w_B + M_A \cdot \mathrm{d}p, \tag{18.3a}$$

$$\mathrm{d}U_B = (w_A - w_B) \cdot \mathrm{d}m_{BA} - m_{AB} \cdot \mathrm{d}w_B + m_{BA} \cdot \mathrm{d}w_A + M_A \cdot \mathrm{d}p. \tag{18.3b}$$

These equations reveal three principal channels for welfare effects of migration. The first terms tell us that either country may derive a benefit from a flow migration where labor outflows are correlated, across different types of labor, with the initial differences in factor rewards. This correlation must be such that the labor movements follow incentives given by wage gaps. Remember that we have assumed undistorted wages in both countries. Importantly, on this account both countries may simultaneously gain from migration.

[49] This analysis is along the lines of Felbermayr and Kohler (2007).

Adding up the first terms in (18.3a) and (18.3b) we obtain a *world efficiency gain* equal to $(w_B - w_A)\cdot dm_{AB} + (w_A - w_B)\cdot\ dm_{BA}$. This is a *first-order* welfare effect of international migration which is comparable to the gains from trade, and positive for both countries, provided that migration strictly follows wage gaps.

The second and third terms on the right-hand side of (18.3a) and (18.3b) indicate the change in incomes that the two countries earn on their existing immigrant *stocks*, due to migration-induced changes in the wage rates, dw_A and dw_B. This effect essentially is a "*terms of trade effect*" that operates on existing *migrant stocks*. The effect is *second order* in nature, arising only if the pre-existing migrant stocks are non-zero, $m_A \neq 0$.[50] Moreover, like all terms of trade effects, it cannot be positive for both countries at the same time, which is directly evident from (18.3a) and (18.3b).

The fourth terms on the right-hand side capture the welfare effects that derive from the goods price changes brought about by migration-induced shifts in worldwide excess demands for tradable goods. This is a conventional *terms of trade effect* operating on *trade flows*. Technically, the effect is of first order since it arises also if existing migrant stocks are zero. It disappears only for countries that do not trade with each other in the initial equilibrium.

There are two ways to obtain a sharper result or further insights. One is to explicitly solve for equilibrating goods price changes, dp, based on the supply and demand effects of migration in the sending and the receiving countries. This approach is pursued in Dixit and Norman (1980). One may question the empirical significance of this effect for practical migration scenarios, but it is an integral part of any migration scenario. Plausibly, with factor prices in the initial equilibrium being different in the two countries, as seems necessary in any meaningful model of migration, the supply effects in the sending and the receiving country respectively will be no mirror images of each other, so we must expect some goods price adjustment in any migration scenario. Generally, a positive terms of trade effect in (18.3a) is likely to arise if the migration flow $dm_A = -\ dm_B$ leads to a worldwide increase in supply of goods where *natives* of country A are net exporters.[51]

A somewhat less demanding way to obtain further insights is pursued in Felbermayr and Kohler (2007), where the commodity terms of trade effect $M_A \cdot dp$ is related to the factor price effects through the *factor content* of the trade vector M_A. It should be noted that

[50] In the entire paper, when logical operators are applied to vectors, they are meant to apply to at least one element in a vector.

[51] Davis and Weinstein (2002) took a perspective on US immigration where the opposite is the case. They assumed that US trade reflects Ricardian comparative advantage as modeled in Dornbusch et al. (1977), with the US completely specialized in a certain range of goods due to superior technology. Any inflow of migration then causes excess supply in goods where US natives are net exporters, thus causing a deterioration of their terms of trade. The opposite would hold true for an outflow of US labor. They calculated a close to 1% negative welfare effect from US immigration of the 1980s and 1990s. See Felbermayr and Kohler (2007) for an analysis that puts this result into a general perspective.

M_A is not the trade vector of *country* A, but the trade vector of its *natives*. This consolidates all relative price effects to a single effect, driven by changes in factor prices $\mathrm{d}w_A$ and $\mathrm{d}w_B$, which operates not just on net migrant stocks, but also on the indirect factor trade through commodity trade of natives. Indeed, this may be extended to include "trade" of non-traded goods, i.e., the net exchange between natives and resident migrants of non-traded goods; see again Felbermayr and Kohler (2007).[52] It is difficult to evaluate the importance of efficiency effects of migration, which do not involve direct international conflict, relative to the two types of terms of trade effects, which do entail conflict potential. But the general message from the above equations is that conflict is a distinct possibility and that the concern for global efficiency should lead to an argument in favor of something like an international migration organization.[53]

3.1.3 The immigration surplus

Can we say more about the conditions under which a certain pattern of flow migration $\mathrm{d}m_A$ does or does not lead to a welfare gain in country A or B? Concavity as well as linear homogeneity of the GDP function $G(p, v)$ in v help us determine such conditions. The wage effects from $\mathrm{d}m_A$ are given by

$$\mathrm{d}w_A = G_{vv}(p, v_A)\cdot\mathrm{d}m_A \text{ and } \mathrm{d}w_B = -G_{vv}(p, v_B)\cdot\mathrm{d}m_A. \tag{18.4}$$

Concavity means that G_{vv} is negative semi-definite, and hence that $\mathrm{d}m_A\cdot\mathrm{d}w_A = \mathrm{d}m_A\cdot G_{vv}(p, v_A)\cdot\mathrm{d}m_A \leq 0$ and $\mathrm{d}m_B\cdot\mathrm{d}w_B = \mathrm{d}m_B\cdot G_{vv}(p, v_B)\cdot\mathrm{d}m_B \leq 0$. Since $\mathrm{d}m_B = -\,\mathrm{d}m_A$, it follows that

$$\mathrm{d}m_A\cdot(\mathrm{d}w_A - \mathrm{d}w_B) \leq 0. \tag{18.5}$$

To fix ideas, let us focus on the third terms on the right-hand side of (18.3a) and (18.3b). They capture what is usually called the "immigration surplus" in the literature.[54] We have already emphasized above that, in essence, this is a terms of trade effect on the two countries' existing migrant stocks. Like all terms of trade effects, it cannot be positive for both countries at the same time. Indeed, inspection of (18.3a) and (18.3b) immediately tells us that one country's immigration surplus is a mirror image of the other country's immigration loss. It is probably fair to say that this has not been sufficiently acknowledged in the literature. As we have seen above, it is only on account of the first-order efficiency effects in the first terms of (18.3a) and (18.3b) that both countries

[52] This notion serves well to substantiate a popular argument in favor of immigration. According to this idea natives gain from an inflow of foreign labor working in certain non-traded goods sectors that loom much larger in natives' consumption basket than in migrants' consumption baskets. One may for instance think of certain low-skill-intensive services (say cleaning).

[53] For a more detailed analysis of this issue, see Hatton (2007).

[54] Borjas (1999) presented a treatment of this surplus with two types of labor, skilled and unskilled. The second-order property of the immigration surplus was first noted in Berry and Soligo (1969).

may simultaneously gain from migration flows. Without loss of generality, let us focus on country A. Our point is easier to make and more obvious if we rewrite $m_A \cdot (\mathrm{d}w_A - \mathrm{d}w_B) = -m_{BA} \cdot \mathrm{d}w_A + m_{AB} \cdot \mathrm{d}_{wB}$, remembering that migrant stocks are strictly one-way, which means that $m_{BA} \cdot m_{AB} = 0$. Thus, if country A enjoys an immigration surplus, it is the sum of a factor price-induced net loss of income accruing to its existing stock of immigrants from country B, plus a net income gain accruing to its own stock of emigrants working in country B.

From (18.4), a non-negative "immigration surplus" for country A will arise, if and only if

$$-m_{BA} \cdot G_{vv}(p, v_A) \cdot \mathrm{d}m_{BA} + m_{AB} \cdot G_{vv}(p, v_B) \cdot \mathrm{d}m_{AB} \geq 0. \tag{18.6}$$

Applying logic presented in Felbermayr and Kohler (2007) we may state a simple sufficient condition for this, which is that the migration flow $\mathrm{d}m_A$ satisfies $\mathrm{d}m_{BA} = \zeta m_{BA}$ and $\mathrm{d}m_{AB} = -\xi m_{AB}$, where ζ and ξ are positive scalars. This is easily seen by acknowledging concavity of the GDP function in v, which in turn implies that G_{vv} is negative semidefinite. More specifically, under the aforementioned condition we have $-m_{BA} \cdot G_{vv}(p, v_A) \cdot \mathrm{d}m_{BA} = -\zeta\,[m_{BA} \cdot G_{vv}(p, v_A) \cdot m_{BA}] \geq 0$ and $m_{AB} \cdot G_{vv}(p, v_B) \cdot \mathrm{d}m_{AB} = -\xi[m_{AB} \cdot G_{vv}(p, v_B) \cdot m_{AB}] \geq 0$, which leads to (18.6).

It follows from the above that a country may derive a strictly positive surplus from proportionally *reducing* its stock of *emi*grants (return migration) and proportionally *increasing* its stock of *im*migrants, provided that the GDP functions of both countries are strictly concave at v_A and v_B respectively. Obviously, it will be impossible for both countries to achieve this at the same time, just as it will be impossible for both countries to gain from any migration scenario on account of a change in the goods terms of trade. Moreover, note that this will happen as the outcome of market forces only if incentives for *present* migration flows run counter to those of *past* migration flows, which are reflected in present stocks. This may appear somewhat odd, but it is not at all ruled out. Perhaps more realistically, such migration flows may be the outcome not of market forces alone, but of immigration restrictions. Most importantly, however, if a country successfully implements such a policy focusing entirely on the immigration surplus, it risks losing on account of the first-order efficiency effect. For country A, for instance, this effect reads as $-\xi\,(w_B - w_A) \cdot m_{AB}$. Indeed, if migrant stocks are positively correlated with remaining wage gaps, $(w_B - w_A) \cdot m_{AB} > 0$, then this first-order efficiency effect of the policy is clearly negative.

The inequality in (18.6) becomes strong if the GDP function is strictly concave both at v_A and v_B. Conversely, the immigration surplus vanishes altogether if the matrix $G_{vv}(\cdot)$ has zero diagonal elements for certain types of labor. Changes in labor supply, $\mathrm{d}m_{AB}$ and $\mathrm{d}m_{BA}$, may then be absorbed without any changes in equilibrium wage rates. Intuitively, this case may arise through a reallocation of factors towards industries that intensively use the types of labor made more abundant through immigration, and conversely for labor that has become scarcer through emigration. This possibility is well known from trade theory, where it is sometimes referred to as "factor price insensitivity", meaning that international factor movements are devoid of any factor price effects. Formally, this type of insensitivity obtains if two conditions are met. Firstly, that factor prices are uniquely tied down by the zero profit conditions of a competitive equilibrium, and secondly, that migration does not affect goods prices. The first condition is met if the number of factors is lower than the number of goods with different factor intensities, and if migration does not push the country out of its cone of diversification. The restriction on the number of goods is likely to be fulfilled, unless factors are specific to goods, as they may be in the very short run. The second condition is met if the country is open to trade and if it is small. Reallocation will alter output patterns in favor of products intensive in emigrating labor. The logic of the factor proportions approach then implies that migration increases (or causes) net exports of goods intensive in immigrating labor, and conversely for goods intensive in emigrating labor. If trade has been driven by factor abundance and scarcity to start with, and if migration, in turn, is in response of international differences in factor prices that reflect relative factor abundance (assuming non-equalization of factor prices through trade alone), then the outcome is the familiar result that migration and trade are substitutes. For instance, skilled labor will move into a country with scarce supply of skilled labor, provided that despite imports of skill-intensive goods the wage for skilled labor is still higher in this country than abroad. And an increased supply of skilled labor increases outputs of skill-intensive goods, thus reducing imports of such goods. Opposite reasoning applies for unskilled labor and exports of goods intensive in unskilled labor. For a small country all of this will happen without any goods price change. Migration is absorbed through "Rybczynski-type" internal factor reallocation. For a large country, however, goods prices will change to restore equilibrium on world goods markets. But the factor proportions logic implies that the factor price effects of immigration will always be less pronounced in a trading economy than in a closed economy, with the extreme case of insensitivity if the economy is small. Note, however, that these results are partly turned upside down if trade is driven by forces unrelated to factor endowments. A case in point is technology-driven trade, to which we shall briefly turn below.

An interesting variant of the above logic is presented in Davis (1998), where there are two factors (labor and capital) and one of the two economies (EU) has a minimum wage rate and the other (US) has a flexible labor market. The two economies are connected to each other by trade in two goods. There is no migration between the US and the EU, but there is immigration into the US from a third country, say Mexico. With the US being a large economy, any absorption of immigration would imply goods market disequilibrium, which in turn causes an increase in the price of the relatively capital intensive good and a corresponding reduction in the wage rate relative to the capital rental. With free trade between the US and the EU, this same reduction in the relative wage rate would need to take place in both countries. But if this is impossible because of a minimum wage in the EU, then the relative price of the two goods needs to be constant too, implying a constant wage rate in the US as well. The US then undergoes a "Rybczynski-type" reallocation of both labor and capital, and the associated additional US supply of the labor-intensive good is compensated by a corresponding reduction of supply from the EU, which is associated with a "reallocation" of European labor into unemployment.

A further variant of this same logic applies for a small country that is open to movements of physical capital at a constant capital rental.[55] To see this, let us for the moment stick to the simple model with capital and labor as the only factors of production. This need not be the same as the world capital rental. Barriers to capital movements may introduce a wedge between the domestic and the world capital rental. What matters is that the country faces an infinitely elastic supply of capital for this given rental rate. Then, any inflow or outflow of labor would be accompanied by an inflow or outflow also of capital, so as to keep constant the wedge between the domestic and the world market return to capital. But with a standard neoclassical technology, this means that the marginal productivity of labor and thus the wage rate remains constant as well.[56] This is true also for a completely specialized economy. Instead of Rybczynski-type internal factor reallocation, we now have the accompanying capital movement. The more general conclusion in the present context is that international mobility of capital tends to attenuate the wage effects from international labor migration.[57] However, the relationship between migration and capital movements is altered once we relax the assumption of complete international mobility of capital. Other things being equal, immigration of labor will lower the domestic wage rate and thus increase the domestic capital rental. If factor flows are responsive to factor price differences that are outside the cone of non-movement introduced in Figure 18.5, then movement of either factor has the same qualitative effect on both factor prices, hence movement of capital and labor may be said to substitute each other.

3.1.4 Distortions and policy

Immigration countries often impose selective quantitative restrictions on labor inflows, based on some view of positive or negative externalities that employment of certain types of workers exert on the economy as a whole. Let us thus assume that we have $w_A = \mathbf{\Omega}_A \cdot G_v(p,\ v_A)$, where the diagonal matrix $\mathbf{\Omega}_A$ denotes the divergencies between factor returns and marginal GDP effects of the various types of labor in country A. A corresponding expression may be introduced for country B. Formally, the matrices $\mathbf{\Omega}$ capture any deviation from the envelope property of the GDP function, stating that

[55] In the introduction to this section, we have argued that movements of existing physical capital are quite unlikely. The present remark intends to clarify the fundamental logic of the factor proportions approach.

[56] The role of capital mobility for the immigration surplus is extensively discussed in Borjas (1999).

[57] This is an example of the Le Chatelier–Samuelson principle, which states that relaxing constraints (in this case the constraint of a given domestic capital supply) will typically dampen the price effects and enhance the quantity effects in the comparative statics of immigration.

the factor supply gradient of the GDP function is equal to the equilibrium factor prices. The institutional interpretation of such deviations is not restricted to factor market distortions. For instance, suppose there is a positive externality emanating from the *output* of good i. Then, any increase in production of good i has an effect on GDP that is above its ongoing price. Another way of stating this is to say that in a competitive equilibrium G_p exceeds the equilibrium supply of good i. However, what matters here is the effect of a variation in some factor supply, say labor of type l, on GDP. In a distortion-free equilibrium this effect is equal to the wage rate w^l, which is in turn equal to $G_v^l(p, v_A)$. For given goods prices, any variation of factor supply leads to a reallocation of all factors, with an effect on goods supply determined by $G_{pv}(p, v_A) \cdot \mathrm{d}v_A$. Absent distortions, the effects of variations in inputs on the overall value of outputs are zero at the margin, the usual envelope property. With the positive externality in sector i, this is not the case. Specifically, the effect of an increase in labor of type l on the value of output will be larger (smaller) than w^l, if the attendant reallocation favors (works against) output of good i. In the two-by-two case, this depends entirely on whether good i is intensive in factor l. The case where an increase in supply of some factor l works against the output of a good with positive externalities has been discussed extensively in the context of immiserizing growth. We shall not pursue this further in this survey.

For the present purpose, a more interesting case of distortions is a spillover effect of employment of labor in any one firm to the marginal productivity of the same type of labor employment in others. Positive spillovers may arise, for instance, from human capital embodied in certain types of labor, as suggested by Lucas (1988) and applied in a traditional way to the immigration surplus in Wong (1995). Such spillovers would be reflected in values below unity in the diagonal matrices $\boldsymbol{\Omega}$. In principle, there may also be negative external effects, implying that wages are above the true marginal value productivities. This then leads to values above unity in the $\boldsymbol{\Omega}$ matrices. Importantly, these are distortions that are not directly related to the degree of competition on labor markets.

A further interpretation is the presence of labor market institutions leading to wages over marginal value productivities, implying values above unity in the diagonal matrices $\boldsymbol{\Omega}$. Such institutional "failures" could, for instance, arise through collective wage bargaining, efficiency wages, or fair wages. This line of interpretation would, of course, require that we depart from the full employment assumption. For reasons of space, however, we do not pursue such an extension in this survey.

Keeping with the assumption of full employment, the above equations for welfare changes through migration now change to:

$$\begin{aligned} \mathrm{d}U_A = & \left(w_B - \boldsymbol{\Omega}_A^{-1} \cdot w_A\right) \cdot \mathrm{d}m_{AB} + \left(\boldsymbol{\Omega}_A^{-1} - \mathrm{I}\right) \cdot w_A \cdot \mathrm{d}m_{BA} \\ & - m_{BA} \cdot \mathrm{d}w_A + m_{AB} \cdot \mathrm{d}w_B + M_A \cdot \mathrm{d}p, \end{aligned} \tag{18.7a}$$

$$\begin{aligned} \mathrm{d}U_B = & \left(w_A - \boldsymbol{\Omega}_B^{-1} \cdot w_B\right) \cdot \mathrm{d}m_{BA} + \left(\boldsymbol{\Omega}_B^{-1} - \mathrm{I}\right) \cdot w_B \cdot \mathrm{d}m_{AB} \\ & - m_{AB} \cdot \mathrm{d}w_B + m_{BA} \cdot \mathrm{d}w_A - M_A \cdot \mathrm{d}p. \end{aligned} \tag{18.7b}$$

In these expressions, I is the identity matrix, hence $(\Omega_A^{-1} - \mathrm{I})$ gives the matrix of proportional divergencies between country A's marginal productivities of the various types of labor and the corresponding wage rates. It seems that only the first terms, the first-order efficiency effects in the first lines of each equation, are affected through distortions. Formally, as far as these equations go, this is true. But the mechanisms behind the relationship between $\mathrm{d}m_A$ and the attendant wage responses, and thus the terms $\mathrm{d}w_A$ and $\mathrm{d}w_B$, will be affected too.[58] Consider the first-order effects in turn. The first term in (18.7a) gives the net aggregate income effect accruing to country A's natives from the migration flow $\mathrm{d}m_{AB}$. Suppose, for concreteness, that $\mathrm{d}m_{AB} > 0$ and $\mathrm{d}m_{BA} > 0$. Economy A thus experiences emigration of some types of labor and immigration of other types of labor. The private benefit for emigrants is the wage w_B that they receive in country B. Country A, however, loses from emigration in line with the marginal social productivity of labor according to $\Omega_A^{-1} \cdot w_A$. In addition, the social benefit to country A (to country A's GNP) of the increase in the stock of immigrants from country B, over and above what immigrants receive in terms of w_A, is $(\Omega_A^{-1} - I) \cdot w_A$. The first two terms in (18.7b) are interpreted accordingly.

The new first-order efficiency terms appearing in (18.7) are best understood by looking at a simple case. Suppose migration of two types of labor, k and l, responds to $w_A^l > w_B^l$ and $w_A^k < w_B^k$ such that $\mathrm{d}m_{BA}^l > 0$ while $\mathrm{d}m_{BA}^k < 0$ (implying $\mathrm{d}m_{AB}^k > 0$). Without distortions, this would clearly enhance world efficiency: $(w_A - w_B) \cdot \mathrm{d}m_A$; labor unambiguously flows from low to high marginal value productivity. Now assume that labor of type k involves a lot of human capital with a positive spillover effect in the production of country A and a correspondingly high value of $\omega_A^k > 1$, where ω_A^k is the element kk of Ω_A^{-1}. Assume, moreover, that the same is true for labor of type l in country B, with a value of $\omega_B^l > 1$. To simplify, let us assume there are no pre-existing migrant stocks, $m_{AB} = m_{BA} = 0$, and let us rule out goods trade, $M = 0$. Then the two countries are affected by this type of migration as follows:

$$\mathrm{d}U_A = \left(w_B^k - \omega_A^k w_A^k\right)\mathrm{d}m_{AB}^k + \left(\omega_A^l - 1\right) w_A^l \mathrm{d}m_{BA}^l, \tag{18.8a}$$

$$\mathrm{d}U_B = \left(w_A^l - \omega_B^l w_B^l\right)\mathrm{d}m_{BA}^l + \left(\omega_B^k - 1\right) w_B^k \mathrm{d}m_{AB}^k. \tag{18.8b}$$

Suppose that $\omega_A^l = \omega_B^l$ and accordingly for labor of type k. Given the wage gaps and given the migration flows, country A benefits from a high ω^l and a low ω^k, while the opposite is true for country B. It is then clear that there is a potential for international conflict if migration takes place under distortions of this type. However, when adding the two country-specific effects, we obtain a world efficiency gain equal to $[\omega_B^k - w_B^k - \omega_A^k w_A^k]$ $\mathrm{d}m_{AB}^k + [\omega_A^l w_A^l - w_B^l w_B]\mathrm{d}m_{BA}^l$. More generally, this effect may be written as

[58] Specifically, the entire reasoning using concavity of the GDP function in order to determine the wage effects of factor supply changes no longer holds.

$$\mathrm{d}U_A + \mathrm{d}U_B = \left(w_B \cdot \Omega_B^{-1} - w_A \cdot \Omega_A^{-1}\right) \cdot \mathrm{d}m_{AB} + \left(w_A \cdot \Omega_A^{-1} - w_B \cdot \Omega_B^{-1}\right) \cdot \mathrm{d}m_{BA}. \tag{18.9}$$

We thus obtain a familiar result. The presence of distortions may cause a welfare loss in cases where migration would otherwise deliver an efficiency. Obviously, given migration flows $\mathrm{d}m_{AB}$ and $\mathrm{d}m_{BA}$, distortions may also enhance the efficiency gain if they happen to follow a specific pattern across factors and across countries. Generally, if distortions in the receiving countries are larger algebraically than in the sending country, and if this distortion asymmetry is particularly large where migration flows are large, then the first-order efficiency gain is larger than without distortions.

The first lines in equations (18.7) in very general terms capture the attempts that we often observe in practical migration policy to influence the pattern of in- and outflows according to the national advantage. For instance, immigration countries often aim at inflows of certain high-skilled workers, based on the notion, however vague, that such workers are the source of positive spillover effects, in addition to being less likely to draw on welfare state budgets. However, if such spillover effects arise in the same way in both the receiving and the sending country, then the sending country will have the opposite incentive of influencing the pattern of emigration. This concern has been expressed in a very long strand of literature on the so-called brain drain (see Section 5). In addition to selective migration policies, the pattern of migration flows regarding the importance of such spillover effects will also be determined by the selection effects in the emigration decisions that we have briefly addressed in the previous section.

3.2 Technology

So far, we have explored the factor proportions logic for migration between countries sharing the same technology. We now explore some of the implications of this logic for migration if trade is driven by differences in technology, instead of endowments. We look at a rather simple scenario with exogenous Hicks-neutral technological advantage in a two-country world where goods differ by factor intensities. A more focused treatment of the role of migration with endogenous technological change will follow in Section 5 below.

An important result, due to Markusen (1983), is that in such a scenario trade will cause factor movements, and in some sense factor price convergence is slowed down. In the simple case with two goods produced with labor and capital, a Hicks-neutral technological advantage of one country over the other in the labor-intensive good installs technology-based comparative advantage if countries have the same factor endowments and if movements are ruled out. The labor-intensive good will be exported by the country with the superior technology. Specialization in the labor-intensive good requires that both goods are produced with a higher ratio of capital-to-labor inputs, which in turn implies a higher wage-to-rental ratio in the superior country. Allowing for migration in this world, the superior country will draw immigration, which adds a factor-proportions-based advantage to its technology-based advantage for the labor-intensive good. Hence, in this scenario trade appears complementary to migration. Of course, we cannot read causality from this, since the scenario arbitrarily assumes free trade with closed factor markets to start with. Reversing the sequence of opening up, we would first have migration that would seem to cause trade. An interesting aspect of the "trade-first-scenario" with technological advantage is that wage convergence appears protracted, relative to abundance-based trade, since complete wage equalization will come about only once the leading country is completely specialized in the labor-intensive good.

Moving to a more general setup comparable to the above model with M factors and $N > M$ goods, assume that α^i denotes a factor of Hicks-neutral technological superiority of country A, relative to country B, in producing good i. We first look at an equilibrium with free trade in goods, but without any factor movements, meaning in particular that there are no migrant stocks m_{AB} and m_{BA}. In such an equilibrium we have

$$p^i \alpha^i = c^i(w_A) \text{ and } p^i = c^i(w_B) \tag{18.10}$$

for any good i that is produced in positive amounts in both countries. Due to free trade, all goods will have the same price in both countries. Equations (18.10) are standard zero profit equilibrium conditions, where $c^i(w_B)$ is used to denote the minimum unit-cost function of country B, which then implies that the minimum unit-cost function for A is $c^i(w_A)/\alpha^i$. Let us assume that there are at least M goods produced jointly in both countries. We use $\overline{p}$ to denote a vector of prices for these goods. Without loss of generality, we may scale units such that $\overline{p}^i = 1$, whence a free trade equilibrium is characterized by

$$\overline{\alpha} = \overline{c}(w_B) - \overline{c}(w_B) \tag{18.11}$$

where $\overline{\alpha}$ is a vector notation for $\overline{\alpha}^i := 1\alpha^i - 1$, corresponding to $\overline{p}$. Similarly, $\overline{c}(w)$ denotes the vector representation of the minimum unit-cost functions for goods corresponding to $\overline{p}$.

We now write $\overline{c}(w_A) - \overline{c}(w_B)$ as a linear approximation

$$c(w_A) - c(w_B) \approx c_w(w_B).(w_A - w_B), \tag{18.12}$$

where element il of $c_w(w_B)$ is the derivative of $c^i(w_B)$ with respect to w^l. According to Shephard's lemma, this is equal to the cost-minimizing input demand for factor l per unit of good i in country B. We assume that the matrix $c_w(w_B)$ is regular. Loosely speaking, this means that goods differ in their relative per-unit inputs of the various factors. This approximation leads to $\overline{\alpha} \approx c_w(w_B) \cdot (w_A - w_B)$. Observing that $\overline{\alpha} \cdot \overline{\alpha} > 0$, using (18.11) as well as (18.12), we may write[59]

[59] Ethier (1982) invoked the mean value theorem to derive an exact version of this statement.

$$\overline{\alpha}\cdot c_w(w_B)\cdot(w_A - w_B) > 0 \tag{18.13}$$

This has a straightforward meaning: In an equilibrium with free trade in goods and absent factor movements, country A will on average have higher wages than country B for those types of labor that are intensively used in the production of goods where it has a strong technological advantage over country B. This repeats in general terms what we have just seen in the Markusen model above. It is an alternative explanation for the wage gaps $w_A - w_B$ that we have used in the entire analysis above. Instead of factor endowment differences it invokes technological differences between the two countries that vary in degree across goods, which in turn vary in factor intensities.

It can be shown that such an equilibrium features net trade vectors, which are in a similar way related to $\overline{\alpha}$ through the factor intensity matrix $c_w(w_B)$.[60] Consider what happens if we now allow for factor movements in addition to trade. Markusen and Svensson (1985) assumed that the mobile factor is capital, but the analysis goes through for labor as well. Instead of capital owned by residents of one country being invested abroad, we then simply have cross-border migrant stocks, as in the previous subsection. The outcome is that the correlation (across goods) between the extent of a country's (Hicks-neutral) technological superiority and the volume of its goods exports is stronger than with trade in goods alone. In other words, trade volumes (net exports) for the same vector of technological superiority $\overline{\alpha}$ are *on average* larger with such factor mobility than without. It is in this sense that trade and factor mobility are complements if countries feature different technologies. Notice, however, this does not mean larger net export quantities for *all* goods, a subtlety already noted in Markusen (1983).

3.3 International convergence

At first sight neoclassical theory seems to suggest that factor movements should generally lead to factor price convergence. However, we have already seen above that this is no foregone conclusion; see our above remarks on "factor price insensitivity". In this subsection, we first add a few further remarks on this issue, still maintaining the assumption of a convex technology, to be followed by a brief treatment of convergence for a case where non-convexities lead to agglomeration.

3.3.1 Convex technology

If we assume *complete* factor mobility, then on a fundamental level the question of complete convergence boils down to whether there exists a unique world equilibrium with factor price equalization and diversification in factors, meaning that all countries host all factors. For well-behaved neoclassical models, the answer should be yes. But the process

[60] See Ethier (1982). Markusen and Svensson (1985) derived an even stronger result by restricting the technological superiority vector $\overline{\alpha}$ so that the two countries are in effect equal in demand.

of convergence much depends on the underlying forces for factor movements. For instance, in the Markusen (1983) model where the cause of factor movements lies in Hicks-neutral technological superiority, complete factor price convergence in a world with trade requires more factor movements than if the underlying cause of factor price gaps is different factor endowments with identical technology. More specifically, while factor price equalization with identical technology is perfectly possible without hitting the far extreme of complete specialization, this is not true for technology-driven factor movements. The reason is that with Hicks-neutral technological superiority the zero profit conditions with equal factor prices can be satisfied with equality in *both* countries for at most *one* good.[61] To put it more generally, there is a general presumption that if countries command technological superiority, then convergence of factor prices needs a much larger cross-country flow of quantities than with common technologies across all countries.[62] In the extreme case where a country is superior in *all* activities, all factors would flow to this country without ever reaching factor price convergence.

More interesting, however, is the question of *local* convergence, i.e., asking whether movements of some factor k between countries A and B caused by $w_A^k < w_B^k$ lead to factor price adjustments in the two countries, such that $\mathrm{d}(w_A^k > w_B^k) < 0$. Again, for the standard two-by-two model this seems fairly trivial, provided that the international gaps in factor returns are due to factor endowment gaps, and not to technological superiority. Using A and B to denote two countries, we may state that $w_A^L > w_B^L$ implies $w_A^k < w_B^k$, and $\mathrm{d}v_A^L = -\mathrm{d}v_B^L > 0$ as well as $\mathrm{d}v_A^K = -\,\mathrm{d}v_B^K < 0$. Importantly, this assumes full equalization of all goods prices between the two countries through trade, whence differences in factor returns are always differences in s returns. The outcome of such movements, according to standard properties of the two-by-two model, is $\mathrm{d}w_A^L < 0$ and $\mathrm{d}w_B^K > 0$, and vice versa for country B.

A straightforward extension of this idea to higher dimensions is to ask whether any non-zero-valued vector $w_A - w_B$ leads to specific factor movements $\mathrm{d}v_A = -\,\mathrm{d}v_B$, which then cause factor price adjustments in both countries, such that

$$\mathrm{d}(w_A - w_B)\cdot(w_A - w_B) \leq 0. \tag{18.14}$$

If so, then we may speak of "average convergence" across all factors. However, it follows from the above that we may not in general expect such convergence to occur. Suppose

[61] Suppose country A commands Hicks-neutral superiority in sector i over country B, with parameter α^i, where $\alpha^i \neq \alpha^j$ for any i and j. We may allow superiority to be negative, $\alpha^i < 1$, for some sectors and positive for others. Then, the zero profit conditions in an equilibrium with free and costless trade in goods that leads to factor price equalization, $w_A = w_B = w$, read as follows: $p^i\alpha^i \geq c^i(w)$ for country A and $p^i \geq c^i(w)$ for country B, with the usual complementary slackness conditions. In these equations, $c^i(w)$ gives country B's unit-cost function in sector i, and $c^i(w)/\alpha^i$ does the same for country A. It is obvious that with $\alpha^i \neq \alpha^j$ these conditions can be fulfilled with equality for at most one good.

[62] This is, again, a reaction of the Le Chatelier–Samuelson principle.

that $dv_A = S(w_A - w_B)$ captures the response of the factor inflow (outflow) in country A (B) to the vector of factor price differences, $w_A - w_B$. The condition (18.5) then reads as

$$(dw_A - dw_B) \cdot S(w_A - w_B) \leq 0. \quad (18.15)$$

One may list a few plausible properties of S, such as positive diagonal elements of the derivatives matrix S_w and the condition that $dv_A^K > 0$ only if $w_A^k - w_B^k > 0$, where k indexes factors. But this is not enough to guarantee convergence in the sense of (18.15). To put it differently, conditions (18.14) and (18.15) jointly impose an implicit convergence condition on the function $S(w_A - w_B)$, describing behavior of factor owners, that must be satisfied for average cross-country convergence of all factor returns. Inevitably, at this level of generality, this condition must appear somewhat arcane.

Potentially, the decision by a factor owner to supply her factors across country borders is a very complex affair. Arguably, it is considerably more complex for labor movements than for capital movements. The reason is that, barring cross-border commuting, labor migration implies movement of both the location of factor use and the factor owner. In short, it involves movement of people, in addition to movement of factor inputs. Therefore, it is likely to involve considerations that go beyond simple wage comparisons, and it almost certainly goes beyond considering "own effects" in the function $S(w_A - w_B)$, as we know from the works of Borjas (1987) and Stark (1993), among others. Without going into detail, our conclusion at this stage is that, once we go beyond the simple two-by-two case, neoclassical theory does *not* generally suggest that factor movements should cause local convergence in the average sense of inequality (18.14). Importantly, however, theory does not suggest that there is any *force of divergence* either.

3.3.2 Increasing returns: new economic geography

Allowing for non-convexity means that we allow for one form or another of increasing returns to scale. Intuitively, the general presumption is that this potentially destroys whatever convergence there might be without such scale economies. The most prominent case in point is, of course, the theory of "new economic geography". This theory modifies the neoclassical approach to trade and factor movements by allowing economies of scale in order to model forces of agglomeration. Intuition and quick inspection of data tell us that such forces have always played a big role in the distribution of economic activity in space, but until recently they have not been addressed in formal analysis using neoclassical models of trade and factor movements. While this theory usually does not frame its notion of space along the dimension of country borders, it is all too obvious that many of its insights are relevant also for the interrelationship between trade and factor movements across countries.

New economic geography models focus on a particular form of scale economies that leads to so-called backward and forward linkages. The important point here is that in the Marshallian dichotomy these economies do not constitute externalities, but are internal to the firm, modeled through a fixed cost of production. Hence they require a departure also from the paradigm of perfect competition.[63] Most models of new economic geography assume monopolistic competition along the lines suggested by Dixit and Stiglitz (1977), which does not involve complex strategic interaction among firms and which features zero profits in equilibrium. A further key departure from the assumptions that we have so far made is that trade of manufacturing goods is subject to "iceberg-type" transport costs. In this subsection, we assume that these transport costs are given (or that they vary) in exogenous fashion. In particular, they are assumed exogenous to migration. This assumption will be relaxed in Section 4 below.

The canonical model of the new economic geography was developed by Krugman (1991b) and is now known as the "core-periphery model".[64] In some sense this model is diametrical to the models of trade and migration that we have used above. The numbers of factors and goods is reduced to two, and it features specific parameterization of production and preferences, thus placing less emphasis on generality. This cost is justified by sharp predictions, although closed form solutions are usually not available. The model assumes two factors that are completely specific to two sectors. One is the so-called numéraire sector (agriculture) featuring constant returns to scale and perfect competition, and the other is manufacturing which produces under increasing returns and monopolistic competition. In its simplest form, the model also assumes two regions, which for the present purpose may be seen as our two countries A and B. Agricultural goods are traded between regions without cost, while manufactures are tradable subject to transport costs (as opposed to revenue-generating barriers like a tariff or a quota). It is crucial that such transport costs are modeled in "iceberg form". This approach, due to Samuelson (1952), is almost ubiquitous in modern trade literature.[65] Importantly, manufacturing labor is assumed to be completely mobile between regions (countries), while agricultural labor is assumed immobile.[66] This is a further important departure from the above analysis where we have assumed factor-specific costs of cross-border movement.

[63] Perfect competition could be maintained if one assumes economies of scale to be external to the firm (Marshallian scale economies). However, as is well known, this typically leads to multiple trading equilibria with vastly different patterns of specialization, at least if we abstract from costs of transport or other so-called real trade costs. For this reason, the literature mostly shied away from this modeling approach. For a recent approach that greatly reduces the scope for multiplicity of equilibria by deviating from the simple perfect competition pricing rule, see Grossman and Rossi-Hansberg (2010).

[64] More elaborate versions of this model are found in Fujita et al. (2001, Chapter 5), and in Fujita and Thisse (2002, Chapter 9).

[65] Taken literally, the assumption means that when shipping a certain amount of a good to a distant market, a fraction less than one of what has been produced will end up being delivered. More fundamentally, it means that (i) the technology of transporting goods uses the same factors (technology) as the technology of production, and (ii) transport costs are variable in nature, effectively increasing marginal cost of production by a constant factor.

[66] The story can also be told in terms of skilled and unskilled workers respectively; see Fujita and Thisse (2002).

The potential of divergence in this world of the new economic geography is best understood by considering deviations from a completely symmetric equilibrium where the two countries appear as clones of each other. In a neoclassical world with conventional properties, two countries who are clones would not trade with each other. Nor would we expect any incentive for factor movements between such countries. With monopolistic competition and product differentiation, we do observe trade, but this is intra-industry trade based on consumers' desire for product variety, which means that consumers in either country consume all varieties produced worldwide. But we would still not expect any incentive for factor movements, for wage rates are the same in both countries, $w_A^f = w_B^f$, where f indicates the factor specific to the manufacturing sector.

However, the presence of real trade cost now makes all the difference. The difference is not that such an equilibrium entails different factor prices, but that it need not be *stable*. Given mobility of the factor specific to manufacturing, a stable equilibrium may involve a large (in the extreme case complete) concentration of the entire world endowment of this factor in one of the two countries, which will then also pay a much higher real wage rate. Factor mobility may thus unleash a *force of divergence*.

However, instability of a symmetric equilibrium is a possibility, not a foregone conclusion. What are the economic mechanisms of divergence and what determines their relative weight in the adjustment? This can be seen without going into further model details by considering the effects of a deviation from a symmetric equilibrium that are caused by moving a unit of the factor specific to manufacturing from one country to the other, say from A to B. Krugman (1991b) identified three effects. First, there is the conventional force from relative scarcity of sector-specific factors, which should benefit manufacturing factor owners in A and hurt those in B, where manufacturing has now become a larger sector relative to agriculture. This force is conducive to stability of the symmetric equilibrium, as it tends to depress w_B^f relative to w_A^f. In some sense, it is comparable to the force of diminishing marginal returns in the conventional neoclassical model. Notice that all penalties of higher dimensions that we have addressed above are ruled out in this model of the new economic geography: There are only two factors, one specific to agriculture, the other specific to manufacturing.

But there are two further forces, deriving from economies of scale and transport costs, both of which are destabilizing in nature. The first is what Krugman (1980) has called the "home market effect". Compared to country B, factors working in country A's manufacturing sector are now less productive in *serving* markets, because a larger share must be served at a distance, incurring transport cost. This must work towards an increase in w_B^f, relative to w_A^f, thus contributing to instability. And finally, if manufacturing factor owners live where their factors work, then those now living in B benefit from *being served* locally for a larger share of the differentiated manufacturing goods that they consume, because country B now hosts a larger share of worldwide manufacturing factors. Notice that this effect only works if factor owners migrate with the location of their factor use.

It will typically be present for labor migration, but not for footloose capital. Thus, for migration, there is a destabilizing force from both the perspective of serving markets as well as from the perspective of being served from markets.[67]

What determines the strength of these destabilizing forces? Obviously, the size of transport costs matters. To see this, first note that whenever instability obtains, by construction of our argument there will be two symmetric stable equilibria. Moreover, if transport cost are zero, then the location of mobile factors does not matter, provided that there is no cost of moving for manufacturing workers, as assumed. Thus, for transport costs in the vicinity of zero, the symmetric equilibrium cannot be stable. At the other extreme, if transport costs are infinite, then there is no trade. In this case there is no equilibrium other than a completely symmetric equilibrium.[68] By continuity, there must be a magnitude of trade costs that separates the two worlds of stability and instability respectively. Thus, without factor movements we have a symmetric equilibrium in a world which is symmetric to start with, and which becomes potentially non-symmetric only due to factor mobility. Whether or not it does, however, depends on the size of transport costs.

What does the new economic geography suggest about the relationship between migration and trade? There are two ways to approach this question. One is to compare alternative stable equilibria with different degrees of concentration for a world with completely symmetric distribution of immobile factors across countries, and to see how different levels of migrant *stocks* relate to the volume of trade. The other is to focus on the adjustment process leading to such an equilibrium and see how migration *flows* relate to associated changes in the level of trade. We briefly sketch answers found for either of these two approaches.

Take the stock view first. In terms of the above technology, the stable equilibria will involve different levels of cross-border stocks of migrants, and in the simple model described above a country with only one mobile factor, i.e., manufacturing labor, a country will unambiguously end up either having a stock of emigrants or a stock of immigrants of manufacturing labor. Different equilibria will also involve different volumes of trade, and one may now ask whether a larger stock of migrants will also involve a higher volume of trade. In the simple model there are at most three stable equilibria: a completely symmetric equilibrium, with a zero cross-country stock of manufacturing migrants, and two opposite core–periphery equilibria, with all manufacturing labor concentrated in one of the two countries (the core), and the other country (periphery) appearing as a pure emigration economy, with all of its manufacturing labor having emigrated to the core.

One might be tempted to expect more trade in the agglomeration equilibria than in the symmetric equilibrium. However, this need not be the case. In the symmetric

[67] In the new economic geography literature, these two perspectives have become known as "forward" and "backward linkages" respectively.

[68] Remember that both countries are endowed with equal amounts of agricultural factors. By continuity, there must be a magnitude of transport costs.

equilibrium with zero migrant stocks, the volume of intra-industry trade reaches its maximum level, while the volume of inter-industry trade is zero. The opposite is true for the core–periphery equilibrium. Hence, without further knowledge about structural features of the economies involved, we must conclude ambiguity when looking at overall trade. When looking at intra-industry trade, we may conclude complementarity. When looking at intra-industry trade, we may conclude substitutability.[69]

But what if we look at the relationship between trade and migration in the adjustment dynamics? As we have mentioned above, the adjustment dynamics of the new economic geography models typically implies that the symmetric equilibrium breaks down once the level of trade costs falls below a critical level. Passing this critical level from above, a small reduction in trade cost and an associated increase in the volume of trade will be associated with first-time movements of manufacturing labor. Moreover, it is the presence of trade and trade costs that install the force of agglomeration and divergence in factor movements in the first place. In this sense, then, we may unambiguously conclude that the new economic geography predicts complementarity between trade and factor movements.

4. MIGRATION AND THE FORMATION OF PHYSICAL CAPITAL

As outlined in Section 2, the first era of globalization in the nineteenth century was characterized by simultaneous capital and labor flows from Europe to the US (Solimano and Watts, 2005). In the process of enlargement of the European Union (EU) to the East in 2004, first, labor was primarily migrating from Eastern EU countries to the UK and Ireland. More recently, after the transitional agreements ran out, Eastern Europeans used the novel opportunity to migrate also to other Western EU countries like Germany (D'Auria et al., 2008). Interestingly, at least until the financial crisis started in the year 2007, capital may have been accumulated faster in the East (Jevčák et al., 2010). Another interesting case is the German unification (Sinn, 2002). As documented by Burda (2006), capital was flowing from West to East, whereas there was substantial migration from the East to the West of Germany.

In this section, we explore the relationship between migration flows and capital formation from a dynamic perspective. We ask how international (or interregional) labor market integration affects both private capital investment and labor migration over time.[70]

4.1 Neoclassical models with capital adjustment costs

We first explore the determinants and effects of factor mobility from a neoclassical perspective (constant returns to scale and perfect markets) in which the dynamics are

[69] For reasons of space, we cannot go into details here. More details can be found in Helpman and Krugman (1985), who traced out loci of equal trade in endowment space.

[70] Due to space constraints, we focus on private rather than public capital investment. Grossmann and Stadelmann (2011, 2012) developed a model in which migration lowers the optimal level of (productivity-enhancing) public capital investment in source economies, whereas the opposite effect arises in host economies.

governed by capital adjustment costs. We start with a single-sector framework before distinguishing tradable and non-tradable goods.

4.1.1 Single-sector setup

Rappaport (2005) employed the following continuous-time framework to investigate whether labor market integration speeds up the process of economic convergence of a capital-poor economy to a large economy, which has reached the long-run equilibrium. In both economies there is a representative firm that produces a homogeneous (numéraire) good according to a linearly homogeneous production function $Y = F(K, L) \equiv Lf(k)$, where Y is output, L is a homogeneous labor input, K is the capital stock, and $k \equiv K/L$ is the capital–labor ratio. We assume that $f(\cdot) \equiv F(\cdot, 1)$ is strictly concave. The capital stock accumulates according to

$$\dot{K} = I - \delta K, \tag{18.16}$$

where I is gross investment in terms of the numéraire good and $\delta > 0$ is the depreciation rate. A dot on a variable denotes its derivative with respect to time. The time index, t, is omitted when obvious. The initial capital stock is given, $K_0 > 0$. Installing an amount I of new capital requires incurring capital adjustment costs $IG\,(I/K)$, where G is an increasing and convex function.[71]

Capital is mobile internationally. The capital-poor economy is small, whence the interest rate, $r > 0$, is exogenously given from the world capital market. Thus, domestic savings have no effect on capital accumulation, which is determined by investment demand of the representative firm.[72] It maximizes the net present value of its future cash flows, i.e., at time $t = 0$, it solves

$$\max_{\{L_t, I_t\}_{t=0}^{\infty}} \int_0^{\infty} e^{-rt} \left\{ F(K_t, L_t) - w_t L_t - I_t \left[1 + G\left(\frac{I_t}{K_t} \right) \right] \right\} dt \text{ s.t. } (18.16) \tag{18.17}$$

and boundary conditions, where w is the wage rate.

Denote by q the multiplier to constraint (18.16), i.e., the shadow price of capital. The first-order condition for (18.17) with respect to I then implies that $q = 1 + G(I/K) + I/K\,G'(I/K) \equiv \tilde{q}(I/K)$, where $\tilde{q}' > 0$.[73] Thus, $I/K = \tilde{q}^{-1}(q) \equiv \iota(q)$. Writing (18.16) in per-capita terms and using $I = \iota(q)K$, we have

[71] See Abel (1982) and Hayashi (1982).

[72] For this reason, we abstain from specifying intertemporal preferences of consumers.

[73] The current-value Hamiltonian function associated with (18.17) reads as

$$\mathcal{H} = F(K, L) - wL - I \cdot \left[1 + G\left(\frac{I}{K} \right) \right] + q \cdot (I - \delta K).$$

The first-order conditions with respect to control variables L and I are given by $F_1(K, L) = w$ and $\partial \mathcal{H}/\partial I = 0$. With respect to state variable K, we have $-\partial \mathcal{H}/\partial K = \dot{q} - rq$.

$$\dot{k} = \left(\iota(q) - \delta - \frac{\dot{L}}{L}\right)k. \tag{18.18}$$

In a steady state with $\dot{K} = 0$, we have $I/K = \delta$, according to (18.16). Thus, the steady state shadow price of capital is given by $q = \tilde{q}(\delta) \equiv \bar{q}$. Combining $I/K = \iota(q)$ with the first-order condition with respect to the capital stock, q evolves according to

$$\dot{q} = (r + \delta)q - f'(k) - \iota(q)^2 G'(\iota(q)). \tag{18.19}$$

In the steady state, we have $\dot{q} = 0$ and $I/K = \iota(q) = \delta$. Thus, (18.19) implies that the steady-state capital–labor ratio reads as $k = (f')^{-1}(r\bar{q} + \delta + \delta G(\delta)) \equiv \bar{k}$. The wage rate is given by the marginal product of labor, $w = f(k) - kf'(k) \equiv \tilde{w}(k)$. Thus, in steady state, we have $w = \tilde{w}\left(\bar{k}\right) \equiv \bar{w}$.

Immigration has no impact on the long run values of the capital–labor ratio and the wage rate in the large economy. To capture the notion of the small economy being initially capital-poor, we assume $k_0 = K_0/L_0 < \bar{k}$, whence $w_0 < \bar{w}$.

Each worker inelastically supplies one unit of labor. Workers from the capital-poor economy migrate as long as the migration benefit exceeds migration costs. Note that the net present value of future wages abroad is given by $\bar{w}/r$. Also note that, when staying at home, the net present value of future wages at time t reads as $\Omega_t \equiv \int_t^\infty e^{-r(\tau - t)} w_\tau d\tau$, i.e.,

$$\dot{\Omega} = r\Omega - \tilde{w}(k). \tag{18.20}$$

Similarly to Braun (1993), suppose the migration benefit, B, is an increasing function of the ratio of the net present value of future wages abroad to that at home; $B = b(\bar{w}/r\Omega)$, with $b' > 0$. Moreover, suppose that the migration cost, C, increases proportionally with the emigration rate, $-\dot{L}/L$. One reason could be congestion effects born by migrants at the destination if labor flows in too rapidly (a form of labor adjustment costs). Formally, $C = -\frac{1}{\mu}\frac{\dot{L}}{L}$, where the parameter μ measures the degree of labor market integration. In equilibrium, the migration benefit equals migration costs, $B = C$. Thus,

$$-\frac{\dot{L}}{L} = \mu b\left(\frac{\bar{w}}{r\Omega}\right). \tag{18.21}$$

We are now ready to ask how the variables (k, q, Ω, L) evolve over time for $k_0 < \bar{k}$, according to the dynamical system (18.18)–(18.21). This sheds light on the dynamic effects of a comprehensive integration of a capital-poor economy into the global economy. Examples encompass the integration of East Germany into the West German market in the 1990s (Sinn, 2002) and the enlargement of the EU to Eastern European countries in the 2000s. Rappaport (2005) examined the transitional dynamics of system (18.18)–(18.21) numerically, assuming that the production function F is of the Cobb–Douglas type, function G (capturing adjustment cost per unit of installed capital) is linear and migration benefit function b is logarithmic. As the marginal return to capital, $f'(k)$, is high when $k_0 < \bar{k}$ but wages are low $\left(w_0 = \widetilde{w}(k_0) < \overline{w}\right)$, during the transition to the steady state, labor emigrates $\left(\dot{L} < 0\right)$ and capital accumulates $\left(\dot{K} > 0\right)$, i.e., $I/K = \iota(q) > \delta$. This is consistent with the observation in Europe after the fall of the Iron Curtain, characterized by relatively fast capital accumulation in the East and labor migration from Eastern to Western Europe.

An interesting question is whether an increase in labor mobility (increase in μ) raises the speed of wage convergence. Faster emigration raises wage rates, *ceteris paribus*, but also turns out to reduce the shadow price of capital, due to the complementarity between labor and capital, which slows down capital accumulation. Rappaport (2005) demonstrated that, as a result, there is little quantitative difference in the convergence process when labor mobility increases at moderate values of μ.

4.1.2 Tradable and non-tradable Goods

Adjustment costs from migration as reflected in (18.21) may partly be justified by the notion that migration flows change interregional differences of house prices. In fact, there is convincing evidence that immigration raises housing costs (e.g., Saiz; 2003, 2007; Nygaard, 2011; Jeanty et al., 2010; Gonzalez and Ortega, 2013). Following Grossmann et al. (2012), we now incorporate the channel from migration to housing costs. We introduce, in addition to a tradable goods sector, a non-tradable consumption goods sector that uses land intensively and could be interpreted as housing sector. The sectors are indexed by T and N respectively.

We again consider migration from the perspective of a small economy (not necessarily in its steady state before labor market integration) to or from a large economy that is and remains in steady state. To allow for potential supply responses to migration-driven changes in housing demand, suppose that (residential) capital accumulates subject to capital adjustment costs. Again, there are no market imperfections. Time is discrete. Analogously to (18.16), the capital stock in sector $j \in \{T, N\}$ evolves according to

$$K_{t+1}^j = I_t^j + \left(1 - \delta^j\right) K_t^j. \tag{18.22}$$

$K_0^j > 0$, where I^j is gross investment in terms of the tradable good (chosen as numéraire) and $\delta^j > 0$ is the depreciation rate in sector j. Again, firms maximize the net present value of future profits and face capital adjustment costs. Analogously to the one-sector model, the total cost (including adjustment cost) per unit of installed capital in sector j (in terms of the numéraire) is $1 + G^j(I^j/K^j)$, where G^T and G^N are increasing functions.

Output levels of the tradable and non-tradable good, Y^T and Y^N respectively, are given by neoclassical production functions F^T, F^N: we have $Y^T = F^T(K^T, L^T)$ and $Y^N = F^N(K^N, L^N; Z)$, where K^j and L^j are the amounts of physical capital and labor used in sector $j \in \{T, N\}$, and Z is the input (as well as supply) of a fixed factor, called "land", in the non-tradable goods sector.

Individuals live for two periods ("working age" and "retirement") in overlapping generations. They are identical with respect to their labor endowment, but may differ in their land endowment. They draw utility from consumption of both goods in both periods of life and can save for retirement at the given world market interest r. Let p_t denote the (relative) price of the non-tradable good ("house price") in period t. Indirect lifetime utility in both economies of an individual born in t with wage income w_t but no other source of income can be written as $V(w_t, p_t, p_{t+1})$, where V is increasing in w and decreasing in the house price in both periods of life.

There are neither institutional migration barriers nor psychological migration costs. Individuals decide at the beginning of the first period whether to stay or to migrate, seeking to maximize utility. Denote by $\overline{w}^*$ and $\overline{p}^*$ the steady-state value of the wage rate and the house price in the foreign economy respectively. Thus, $V^* \equiv V(\overline{w}^*, \overline{p}^*, \overline{p}^*)$ is the steady-state utility of a worker abroad with wage income $\overline{w}^*$. An individual who does not own land has a higher incentive to migrate than an individual with income from landholding, because only the wage income potentially changes when migrating and the marginal utility of income is declining. A landless individual is indifferent between staying and migrating if $V(w_t, p_t, p_{t+1}) = V^*$. If there is a sufficiently high fraction of such workers in the population, this condition must hold in equilibrium with integrated labor markets. Prior to migration, the number of old natives, $L_{-1} > 0$, is given. In equilibrium, we have $L^N + L^T = L$, where L is endogenously determined when labor markets are integrated. In addition to possible wage differences, bilateral migration flows depend on the (initial) difference in the population density. Intuitively, an increase in population density raises the house price, because of a land dilution effect. In equilibrium, both house prices and wage rates may differ across regions even in the long run.

If both economies are in steady state prior to labor market integration, opening up the labor market induces capital and labor to flow in the same direction. In the destination economy, labor market integration leads to an increase in house prices. Because a higher house price raises the shadow price of residential capital, it triggers capital accumulation. Moreover, the price of land rises along with immigration during the entire transition path. Thus, immigration aggravates the welfare differences that arise from differences in the ownership of land. In the absence of wage effects of immigration,[74] individuals born in the destination country with labor income only unambiguously lose from labor market integration, whereas landowners may win. Conversely, outward migration slows down the residential capital accumulation but may benefit native workers through lower housing costs.

If the initial capital stock is sufficiently low and the initial population density is sufficiently high, then labor market integration triggers outward migration at the same time as capital accumulates. Capital accumulation leads to a reversal of migration flows during

[74] For instance, in the special case where tradables are produced using labor only, $Y^T = aL^T$, $a > 0$, wage rates would be constant ($w_t = a$ for all t). In this case, welfare effects run through changes in the house price and, by affecting lifetime income of landowners, through the price of land.

the further transition. This development is consistent with the "natural experiment" of the German reunification, where labor emigrated massively from the east to the west of Germany in the 1990s along with capital formation in the eastern part, whereas more recently some regions in the east of Germany experienced net immigration.

In sum, a neoclassical framework with factor adjustment costs is capable of explaining interregional movements of capital and labor in the same or in opposite directions, depending on initial conditions and the point in time of the transition. Moreover, in a two-sector framework with endogenous housing supply, there may be non-monotonic transitions of population density. Nevertheless, the *causal* effect of immigration (emigration) on residential capital investment and house prices, in response to interregional labor market integration, is always positive (negative). House and land prices permanently rise with higher population density even though housing supply adjusts over time.

4.2 Increasing returns and agglomeration effects

We have seen that initial conditions (i.e., "history") entirely determine factor flows in neoclassical growth models in response to integration shocks. If we allow for increasing returns and agglomeration effects, expectations matter as well, potentially leading to multiple equilibria.

Burda and Wyplosz (1992) considered human capital externalities, inspired by Lucas (1988, 1990), in a two-region model with adjustment costs and mobility of both capital and labor across regions. They applied their model to the case of German unification. Allowing for bilateral factor movements as well, Faini (1996) and Reichlin and Rustichini (1998) captured "learning-by-doing" externalities from physical capital, inspired by Arrow (1962) and Romer (1986).

Consider the following stylized two-country, one-sector framework in discrete time which incorporates both sources of increasing returns. There is a unit mass of identical final goods producers. Final output (the numéraire good) of a domestic firm is produced according to

$$Y = AK^{\alpha}L^{1-\alpha}, \tag{18.23}$$

$\alpha \in (0, 1)$, where K is physical capital, L is labor input, and A is the TFP level. The TFP level depends on the (average) level of human capital of the individuals in the economy, h,[75] and the average level of physical capital of the final goods producers, $\overline{K}$, according to

[75] There has been some debate about whether human capital externalities are important empirically. Although Acemoglu and Angrist (2001) and Ciccone and Peri (2006) are sceptical, whereas the more recent literature finds quite strong support (e.g., Iranzo and Peri, 2009; Hunt and Gauthier-Loiselle, 2010; Gennaioli et al., 2011).

$$A = h^{\beta}\overline{K}^{\gamma}, \tag{18.24}$$

$\beta \geq 0, \gamma \in [0, 1-\alpha)$, i.e., physical capital externalities are limited. In equilibrium, $K = \overline{K}$.

We consider the case where physical capital is perfectly mobile, whereas after labor market integration there is still less than full labor mobility. Denote foreign variables with an asterisk. Suppose that workers want to migrate if and only if the wage rate abroad, w^*, relative to the wage rate at home, w, exceeds some threshold $\rho > 1$. Otherwise, nobody migrates. To capture labor adjustment costs that prevent a complete exodus of labor in just one period, we assume that the number of immigrants per period as a fraction of the native population must not exceed $\lambda \in (0\ ,\ 1)$. Also for simplicity, suppose that firms do not face capital adjustment costs.

Perfect capital mobility implies that the returns to physical capital are equalized: $r = r^*$, where

$$r \equiv \left.\frac{\partial Y}{\partial K}\right|_{K=\overline{K}} = \alpha h^{\beta} k^{\alpha+\gamma-1} L^{\gamma} \tag{18.25}$$

according to (18.23) and (18.24). Thus,

$$\frac{k}{k^*} = \left(\frac{h}{h^*}\right)^{\frac{\beta}{1-\alpha-\gamma}} \left(\frac{L}{L^*}\right)^{\frac{\gamma}{1-\alpha-\gamma}}. \tag{18.26}$$

Suppose that, to begin with, the labor market is closed internationally and the two labor forces are of equal size in period 0, $L_0 = L_0^*$. Assume, moreover, that the domestic labor force is more skilled initially, $h_0 > h_0^*$. For $\beta > 0$ (i.e., with human capital externalities), the domestic economy has the higher capital–labor ratio, $k_0 > k_0^*$, as hypothesized in Lucas (1988).

The wage rate is equal to the marginal product of labor:

$$w = \left.\frac{\partial Y}{\partial L}\right|_{K=\overline{K}} = (1-\alpha) h^{\beta} k^{\alpha+\gamma} L^{\gamma}. \tag{18.27}$$

Computing the relative wage w/w^* from (18.27) and substituting (18.26), we find

$$\frac{w}{w^*} = \frac{k}{k^*} = \left(\frac{h}{h^*}\right)^{\frac{\beta}{1-\alpha-\gamma}} \left(\frac{L}{L^*}\right)^{\frac{\gamma}{1-\alpha-\gamma}}. \tag{18.28}$$

Thus, if $L_0 = L_0^*$, the wage rate is initially higher for the country with a higher human capital level whenever $\beta > 0$; that is, if $h_0 > h_0^*$, then $w_0 > w_0^*$. If labor markets are liberalized in period 1, multiple equilibria may arise. To see this, suppose that $\frac{w_0}{w_0^*} = \left(\frac{h_0}{h_0^*}\right)^{\frac{\beta}{1-\alpha-\gamma}} < \rho$. Provided that nobody migrates in period 1, we also have $\frac{w_1}{w_1^*} < \rho$.That is, initial wage differences are not large enough to induce workers to migrate to the high-wage country. Thus, a situation without migration is an equilibrium. However, for $\gamma > 0$, it is possible that there is an equilibrium with migration as well. To see this, suppose that the maximum amount of λL_0 workers immigrate from abroad. Consequently, the average domestic human capital level drops to

$$h_1 = \frac{h_0 + \lambda h_0^*}{1+\lambda} < h_0, \tag{18.29}$$

whereas $h_1^* = h_0^*$. Using (18.28) and (18.29), $L_1 = (1+\lambda)\, L_0$, $L_1^* = L_0^* - \lambda L_0$ and $L_0 = L_0^*$, the relative wage abroad in period 1 reads as

$$\frac{w_1}{w_1^*} = \left[\frac{\left(\frac{h_0}{h_0^*} + \lambda\right)^{\beta} (1+\lambda)^{\gamma-\beta}}{(1-\lambda)^{\gamma}}\right]^{\frac{1}{1-\alpha-\gamma}}. \tag{18.30}$$

If $\gamma > 0$, $[w_1]/[w_1^*]$ may be increasing in the fraction of immigrants in the total population, λ. On the one hand, immigration from the foreign country with lower skill endowment depresses the average human capital level in the destination country. This, in turn, reduces migration incentives by lowering TFP (Burda and Wyplosz, 1992). On the other hand, immigration induces physical capital inflows, due to the complementarity between labor and physical capital. If capital externalities on TFP are sufficiently high (for instance, if $\gamma \geq \beta$), this effect on relative wages dominates the human capital erosion effect. In this case, if λ is high enough, $w_1/w_1^* > \rho$ such that maximum migration is a second equilibrium outcome, in addition to the equilibrium without migration. If initial human capital differences are sufficiently high such that $w_0/w_0^* > \rho$, an equilibrium without migration may not exist in any period, such that the low-income country may vanish in finite time.

Our simple model illustrates some general lessons from models with migration under increasing returns. First, like the literature on new economy geography, they help to explain core–periphery patterns. More generally, initial conditions ("history") matter for the range of equilibrium outcomes that can rationally be expected (see also Krugman, 1991a). If initial differences in productivity levels (i.e., in the average level of human capital) across regions are sufficiently large, an equilibrium without migration may not exist, whereas with modest initial productivity differences, either migration or no migration are potential equilibrium outcomes. In the case of multiple equilibria, the evolution of the economy depends on expectations, in addition to history.

In an interesting recent paper, Schäfer and Steger (2014) proposed a dynamic multiregion setting with increasing returns to study the effects of a simultaneous integration of both capital and labor markets. They showed that non-monotonic adjustments paths for the capital stock and the labor force may arise subsequent to opening up borders to factor flows. Their contribution highlights the interaction of history and expectations for regional development.[76] It could be interesting for future research to analyze the interaction between migration and capital flows as driven by both history and expectations in a spatial model with trade costs.

4.3 Empirical evidence

As discussed in this section so far, theory suggests that interregional flows of capital and labor in response to labor market integration are jointly driven by differences in initial conditions like productivity, the stock of capital, and population density. In the case of increasing returns, they are also potentially driven by expectations. As capital and labor may flow in the same or in opposite directions, simple statistical analysis on the relationship between capital and labor flows could be quite uninformative or misleading. One robust and testable prediction of the previous analysis is that the *causal* effect of immigration (emigration) on capital accumulation and house prices is positive (negative) in the short run as well as in the long run.

In fact, the effect of immigration on house prices at the regional level is well-established. For instance, Jeanty et al. (2010) found that a 1 percentage point increase in population growth leads to a 0.24% increase in housing costs in the metropolitan area of Michigan. A particularly interesting study is provided by Gonzalez and Ortega (2013) who, in addition to the effect on house prices, estimated the effect of regional immigration on residential construction for Spain. Following Card (2001), they constructed an instrument for the change in regional population size that is based on past immigration.[77] Their instrumental-variable estimates suggested substantial and positive causal effects of immigration on both house prices and residential investment.

A different strand of empirical literature asks whether the stock of immigrants from a certain country determines capital inflows from and capital outflows to that country. The key theoretical idea is that immigrants mitigate informational frictions, thereby reducing the well-known home bias of financial and capital flows (e.g., Coeurdacier and Rey, 2013).[78] Buch et al. (2006) employed panel data for the time period 1991–2002 on both stocks of immigrants and stocks of inward foreign direct investment (FDI) from foreign countries in the 16 German federal states. They find that an increase in the total stock of

[76] It is related to Baldwin and Martin (2004), who discussed international agglomeration effects from capital mobility in a new economic geography model with endogenous growth; however, they abstracted from labor mobility. Ottaviano and Thisse (2004) and Breinlich et al. (2013) provided excellent surveys on dynamic new economic geography models.

[77] Denote by $m_{i,j,t}$ the number of individuals in region i born in country j in year t and by $M_{j,t} := \sum_i m_{i,j,t}$ the total number of individuals born in country j in year t. The predicted stock of migrants in region i in year $t > 0$, based on year 0, then is $S_{i,t} := \sum_j \frac{m_{i,j,0}}{M_{j,0}} M_{j,t}$. Denoting the population size in region i in year $t-1$ by $Pop_{i,t-1}$, the instrument for the change in population size of region i is constructed as $\frac{S_{i,t} - S_{i,t-1}}{Pop_{i,t-1}}$.

[78] Okawa and van Wincoop (2012) discussed how standard gravity forces used to analyze the determinants of international trade (see Section 6) can be used to analyze international financial flows as well.

immigrants in a German state significantly raises the stock of inward FDI, whereas a higher domestic labor force has no significant effect. Moreover, immigration raises inward FDI from the same country of origin as the immigrants, if it is a high-income OECD country. As immigrants to Germany from high-income countries on average tend to be relatively skilled, this suggests that high-skilled immigrants are important to create ties to foreign companies investing in Germany. Conversely, the impact of an increase in inward FDI on immigration from the same source country is generally insignificant. It is positive in the east of Germany if and only if the source country has high income. In sum, the evidence is consistent with agglomeration effects of high-skilled immigration.

Similarly, analyzing bilateral international data, Kugler et al. (2013) found that a higher stock of immigrants has a positive impact on cross-border flows of financial capital from the sending to the host country of migration. The effect is particularly large when the immigrants are high-skilled and two countries do not have a common language or do not share the similar colonial/legal origin. This suggests that immigrants are particularly important for cross-border financial flows when informational problems are severe.

Conversely, Kugler and Rapoport (2007) investigated the impact of a higher stock of immigrants in the US on FDI financed by US firms in the immigrants' country of origin. They found that the stocks of both low-skilled and high-skilled immigrants in the year 1990 have a significant effect on the growth rate of outward FDI between the years 1990 and 2000. As pointed out by Kugler and Rapoport (2007), this may suggest that low-skilled immigrants signal labor force quality to US investors abroad and high-skilled immigrants contribute to the creation of international business networks. Javorcik et al. (2011) also studied the effects of a higher stock of immigrants to the US on the stock of outward FDI. Their instrumental-variable estimates account for potential endogeneity problems which may arise, for instance, when FDI to foreign countries induces migration from subsidiaries to the US headquarter of a multinational company. They estimate that a 1% increase in the stock of skilled immigrants causally raises the stock of US outward FDI by about 0.5%.[79]

5. HIGH-SKILLED MIGRATION AND PRODUCTIVITY GROWTH

Whereas the previous section has examined the interaction between migration on physical capital formation, particularly high-skilled migration, is potentially important for the

[79] An earlier literature has analyzed the effect of ethnic Chinese networks abroad (i.e., Chinese emigration) on inward FDI. Gao (2003) suggested that an increases in the ethnic Chinese population share in the source country by 1 percentage point raises the cumulative FDI inflow to China between 1984 and 1997 by at least 3.7%. Tong (2005) estimated that a 1% increase in the product of the numbers of Chinese emigrants in two countries in 1990 increases the contemporaneous stock of bilateral FDI in 1990 by at least 0.38%.

formation of knowledge capital and productivity growth as well. For instance, since many decades the US has attracted talented people from abroad to places like Silicon Valley and to elite universities. As documented by Saint-Paul (2004), among others, high-skilled immigrants often come from other advanced regions like Europe.[80] While we would expect high-skilled immigrants to contribute to innovation and productivity growth in their host countries, brain drain may be detrimental in their home countries. In this section, we discuss both hypotheses, starting with the perspective of host countries.

5.1 Knowledge capital formation

Analyzing the impact of high-skilled immigration on innovation and the accumulation of knowledge capital requires a general equilibrium perspective. Endogenous growth theory is particularly well suited. An important debate related to whether skilled immigration fosters R&D-driven growth is whether there exist scale effects either on the growth rate or on the level of per-capita GDP. We briefly discuss the mechanisms that may give rise to scale effects in models of both horizontal and vertical innovation.

5.1.1 *Product innovation*

We first turn to continuous-time models where productivity growth is driven by innovations that lead to new capital goods.

The Romer–Jones model Consider a large economy with population size $L_t = \overline{L}e^{nt}$ at time $t \geq 0$. The population growth rate, $n \geq 0$, is constant. We capture immigration by a one-shot increase in population size, i.e., by an increase in $\overline{L}$. There is a representative household, who owns the average amount of assets, and inelastically supplies one unit of labor to the production of a homogeneous consumption good (the numéraire) or to an R&D sector. We assume that the household has an infinite time horizon and chooses her consumption path based on the standard intertemporal utility function

$$U = \int_0^\infty \frac{(c_t)^{1-\sigma} - 1}{1-\sigma} \mathrm{e}^{-(\theta - n)t} \mathrm{d}t, \tag{18.31}$$

θ, $\sigma > 0$, where c is consumption per capita.

[80] Another example is Switzerland, known for its financial industry and pharmaceutical sector. For instance, after bilateral migration with the EU15 countries has been liberalized, the net immigration flow from Germany alone to Switzerland (which had a population size of 7.8 million in 2009) exceeded 100,000 people in the period 2007–10 (www.bfs.admin.ch). The majority of immigrants from Germany hold a university degree.

There is a representative firm in the final goods sector that produces according to

$$Y = \left(L^Y\right)^{1-\alpha}\int_0^A x(i)^\alpha \mathrm{d}i, \tag{18.32}$$

$\alpha \in (0, 1)$, where L^Y is labor input in final goods production and $x(i)$ is the quantity of capital good $i \in [0, A]$. There is perfect competition except in the capital goods sector, in which there are monopolistically competitive single-good firms. One unit of foregone consumption can be transformed into one unit of a capital good. The physical capital stock then is $K = \int_0^A x_i \mathrm{d}i$. In symmetric equilibrium where $x(i) = K/A$ for all i, we find that per capita income reads $y \equiv Y/L = k^\alpha (Al^Y)^{1-\alpha}$, where $k \equiv K/L$ and $l^Y \equiv L^Y/L$. That is, holding the allocation of resources devoted to the final goods sector constant (i.e., holding k and l^Y constant), per-capita income is increasing in the number of capital goods, A. Similar to "love of variety" monopolistic competition models in new trade theory and the literature on new economic geography, this kind of specialization gain is an implication of decreasing marginal returns to each capital good assumed in (18.32).

When $L^A = L - L^Y$ workers are allocated to R&D, the number of capital goods, which measures the economy's knowledge stock, changes according to

$$\dot{A} = \lambda A^\phi L^A, \tag{18.33}$$

$\lambda > 0$, $\phi \leq 1$. If $\phi > 0$ there is an "intertemporal knowledge spillover" from previous R&D. In his seminal paper on endogenous technical change, Romer (1990) assumed (implicitly) that $\phi = 1$ and $n = 0$, whereas Jones (1995) analyzed the case where $\phi < 1$ and $n > 0$. In steady state, there is a common time-invariant growth rate of the knowledge stock (A), income per capita (y), the capital–labor ratio (k) and per capita consumption (c), denoted by $g = \dot{A}/A = \dot{y}/y = \dot{k}/k = \dot{c}/c$.

In the Romer model, with $\phi = 1$, $\dot{A}/A = \lambda L^A$ is increasing in the number of R&D workers, L^A. Romer (1990) showed that, in the steady state, L^A is increasing in population size, $\overline{L}$. In this sense, immigration of workers capable of performing R&D tasks would permanently raise the economy's long-run growth rate, g. This has been referred to as "strong scale effect".

However, the assumption $\phi = 1$ may be criticized as a knife-edge case. Assuming $\phi < 1$ dramatically changes the outcome. It is easy to see that $g = \dot{A}/A = \lambda A^{\phi-1} L^A$ is time-invariant if and only if the growth rate of R&D employment, $n^A \equiv \dot{L}^A/L^A$, is time-invariant. The long-run growth rate reads as $g = n^A/1 - \phi$. In fact, one can show that the long-run allocation of labor is independent of population size such that R&D employment grows at the same rate as population size, $n^A = n$. Hence, in contrast to the Romer model, the economy's growth rate in the Jones model does not depend on $\overline{L}$. However, as discussed in Jones (1999, 2005), the scale effect now shows up in levels of the variables of interest, rather than in their growth rates. To be precise, a one-shot increase in population size, $\overline{L}$, raises the de-trended level of the knowledge stock, $\widetilde{A}_t \equiv A_t/e^{gt}$, thus raising the level of per capita income, $\widetilde{y}_t \equiv y_t/e^{gt}$, in the long run (as $t \to \infty$). This property is typically referred to as "weak scale effect".[81] To conclude, in this class of horizontal innovation models, international migration of skilled labor leads to divergence rather than convergence of per-capita income across economies.

Directed technical change As emphasized in Acemoglu (1998, 2002), an increase in the size of the high-skilled relative to the low-skilled population may determine whether innovations are directed to new capital goods, which are complementary to skilled labor, or directed to capital goods complementary to unskilled labor. Thus, selective immigration policy towards high-skilled labor may affect the direction of technological change.

We briefly illustrate the idea by following Acemoglu (2002). There is skilled and unskilled labor, in amounts H and L respectively. Both types of population grow at the same rate, $n \geq 0$. High-skilled and low-skilled immigration is captured by an increase in initial population sizes, $\overline{H}$ and $\overline{L}$ respectively. Final output (the numéraire) is produced under perfect competition according to the CES production function

$$Y = \left(\gamma (X_L)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma)(X_H)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{18.34}$$

$\gamma \in (0, 1)$, $\varepsilon > 0$. X_L and X_L are composite intermediate inputs, which are also produced under perfect competition with differentiated capital goods ("machines") as well as with low-skilled and high-skilled labor respectively. Formally, we have

$$X_L = \left(L^X\right)^{1-\alpha} \int_0^{A_L} x_L(i)^\alpha \mathrm{d}i, \tag{18.35}$$

$$X_H = \left(H^X\right)^{1-\alpha} \int_0^{A_H} x_H(i)^\alpha \mathrm{d}i, \tag{18.36}$$

$\alpha \in (0, 1)$, where $x_L(i)$ and $x_H(i)$ are inputs of machines that are complementary to low-skilled labor, L^X, and high-skilled labor, H^X, respectively. In each of the two machinery sectors there is a monopoly firm with constant marginal costs of unity in terms of the numéraire.

Similar to the Romer–Jones model, the mass ("number") of machines, A_L and A_H, expands through horizontal innovations according to

$$\dot{A}_L = \lambda_L (A_L)^\phi H_L^A, \tag{18.37}$$

$$\dot{A}_H = \lambda_H (A_H)^\phi H_H^A, \tag{18.38}$$

[81] An increase in $\overline{L}$ also raises welfare U. In the long run, the utility integral (18.31) is finite if $\theta > n + (1-\sigma)g > 0$, which also ensures that the transversality condition for the problem of the representative consumer holds.

$\lambda_L, \lambda_H > 0$, $\phi \leq 1$, where H_L^A and H_H^A are the number of scientists directed to innovations that are complementary to low-skilled and high-skilled labor in manufacturing respectively. In labor market equilibrium, $H^X + H_L^A + H_H^A = H$ and $L^X = L$. Unsurprisingly, the scale effects properties of the Romer–Jones model with respect to high-skilled immigration still apply. We now focus on the effect of selective immigration on the composition of R&D activity.

Given competitive input markets, prices of the composite inputs are given by marginal products, $P_L \equiv \partial Y/\partial X_L, P_H \equiv \partial Y/\partial X_H$. Using (18.34), this gives us relative intermediate goods demand:

$$\frac{X_H}{X_L} = \left(\frac{1-\gamma}{\gamma}\right)^{\varepsilon} \left(\frac{P_H}{P_L}\right)^{-\varepsilon}. \tag{18.39}$$

According to (18.35), the inverse demand for machine i in the low-skilled intensive sector is $p_L(i) \equiv \alpha P_L([L^X]/[x_L(i)])^{\alpha-1}$. Thus, the profit-maximizing price is $p_L(i) = 1/\alpha$, implying $x_L(i) = x_L = (\alpha^2 P_L)^{\frac{1}{1-\alpha}} L^X$. Using the latter in (18.35) gives us $X_L = A_L L^X (\alpha^2 P_L)^{\frac{\alpha}{1-\alpha}}$. Analogously, $x_H(i) = x_H = (\alpha^2 P_H)^{\frac{1}{1-\alpha}} H^X$ and $X_H = A_H H^X (\alpha^2 P_H)^{\frac{\alpha}{1-\alpha}}$. Denote by π_H and π_L the profits of machine producers who employ skilled and unskilled labor respectively. Since a constant mark-up implies that π_H and π_L are proportional to output, x_H and x_L, respectively, we have

$$\frac{\pi_H}{\pi_L} = \left(\frac{P_H}{P_L}\right)^{\frac{1}{1-\alpha}} \frac{H^X}{L^X}. \tag{18.40}$$

Moreover, relative supply of composite inputs is

$$\frac{X_H}{X_L} = \frac{A_H H^X}{A_L L^X} \left(\frac{P_H}{P_L}\right)^{\frac{\alpha}{1-\alpha}}. \tag{18.41}$$

Equating the right-hand sides of (18.39) and (18.41) leads to a negative relationship between the relative price of the two composite inputs, $[P_H]/[P_L]$, and relative "efficiency units" of labor, $A_H H^X / A_L L^X$:

$$\left(\frac{P_H}{P_L}\right)^{\frac{1}{1-\alpha}} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{\alpha+\varepsilon(1-\alpha)}} \left(\frac{A_H H^X}{A_L L^X}\right)^{-\frac{1}{\alpha+\varepsilon(1-\alpha)}}. \tag{18.42}$$

Incentives to innovate in a certain direction depend on relative profits, π_H/π_L.

Now consider a selective immigration policy towards skilled labor, such that $\overline{H}/L$ rises. Consequently, the (steady-state) labor allocation will change such that relative employment of skilled labor in production, H^X/L^X, rises. This has two counteracting effects on π_H/π_L. First, according to (18.40), for a given relative price P_H/P_L, relative profits in the high-skilled intensive sector rise ("market size effect"). Second, however, according to (18.42), the relative price for the high-skilled intensive good falls due to a change in relative output ("price effect"). If the composite inputs are "good substitutes",

$\varepsilon > 1$, the first effect on relative profits dominates and selective migration changes the steady-state composition of machines by raising A_H/A_L; if they are "good complements", $\varepsilon < 1$, the second effect dominates. In the former case, skilled migration may trigger "skill-biased technological change" with potentially positive effects on the wage premium on being skilled.

5.1.2 Vertical innovation

We now turn to a class of models with quality improvements (vertical innovations) of differentiated capital inputs as in Young (1998).

We return to a discrete-time notation. The working-age population size, L, grows at a constant rate $n, L_t = (1+n)^t \overline{L}$. Immigration is again captured by an increase in $\overline{L}$. Let us modify the production function to

$$Y = BZ^{1-\alpha-\beta}\left(L^Y\right)^{\beta}\int_0^N A(i)^{1-\alpha} x(i)^{\alpha}\mathrm{d}i, \tag{18.43}$$

$\alpha, \beta \in (0,1), \alpha+\beta \leq 1$, where Z is land input, L^Y is labor input, and $x(i)$ and $A(i)$ are the quantity and a quality index of capital input $i \in [0, N]$ respectively. There is free entry of capital good producers who have to employ a fixed amount $f > 0$ of labor one period ahead production. Thus, the mass ("number") of firms, N, is endogenous. Marginal production costs are equal to the interest rate r that is given exogenously from the world capital market. We allow for adverse congestion effects from increasing density of the (working-age) population, $D \equiv L/Z$, on the productivity parameter B:

$$B = D^{-\eta}, \tag{18.44}$$

$\eta \geq 0$; B is taken as given by final goods producers.

By employing an amount $l_t(i)$ of R&D labor in period t, a capital input producer i affects quality in $t+1$ according to

$$A_{t+1}(i) = \overline{A}_t \Lambda(l_t(i)), \tag{18.45}$$

where $\overline{A}_t \equiv [1/N_t]\int_0^{N_t} A_t(i)\,\mathrm{d}i$ is the average product quality in period t. $\overline{A}$ measures the economy's knowledge stock. As will become apparent, the linear knowledge spillover sustains long-run growth. The function Λ is increasing, strictly concave, and fulfills $\Lambda(0) = 1$. $\overline{A}_0 > 0$ and $L_0^Y \in (0, \overline{L})$ are historically given.

Producer i faces an (inverse) demand function $p(i) = \partial Y/\partial x(i)$ and charges a mark-up equal to $1/\alpha$. Using (43), we can solve $\partial Y/\partial x(i) = r/\alpha$ for $x(i)$ and substitute the resulting expression into (43). Also inserting (44) and using definition $l^Y \equiv L^Y/L$ yields the following expression for per-capita income:

$$y \equiv \frac{Y}{L} = \left(\frac{\alpha^2}{r}\right)^{\frac{\alpha}{1-\alpha}} \left(l^Y\right)^{\frac{\beta}{1-\alpha}} D^{\frac{\beta-\eta}{1-\alpha}-1} N\overline{A}. \tag{18.46}$$

As shown in the Appendix, there is a symmetric equilibrium in which the R&D labor input of each firm, denoted by $\tilde{l}$, is time-invariant, equal among firms, and, importantly, independent of population size, L_t. It is solely determined by the function Λ in R&D technology (18.45) and by the fixed labor requirement f. This result is an implication of free entry of intermediate goods firms. The equilibrium number of intermediate goods firms (N) increases proportionally with population size (L), leaving R&D input per firm unaffected (see e.g., Young, 1998). In equilibrium, $\overline{A}$ evolves according to $\overline{A}_{t+1} = \overline{A}_t \Lambda\left(\tilde{l}\right)$, independently of population size. This result is implied by the assumption that the intertemporal spillover effect is driven by average product quality, $\overline{A}$, thus being independent of the number of firms in equilibrium. Consequently, there is no scale effect in growth rates. Moreover, at any point in time, the equilibrium allocation of labor is independent of population size, i.e., the fraction of labor devoted to manufacturing, l^Y, is independent of L.

According to (18.46), these properties imply that the impact of immigration (increase in $\overline{L}$), on per-capita income, y, is positive (negative) if $\beta > (<)\eta$. There are three kinds of scale effects in levels. First, since the equilibrium number of intermediate goods firms, N, is proportional to L, immigration raises the level of TFP through specialization gains of the sort already discussed in the Romer–Jones model.[82] Second, there is a congestion effect from higher population density, D, on productivity B if $\eta > 0$. Third, if $\alpha + \beta < 1$, there is a land dilution effect, since a larger population size reduces land input per capita (see also Acemoglu and Johnson, 2007). The latter two effects combined imply that scale effects may be negative, in contrast to standard models that only feature specialization gains.[83]

5.1.3 Multi-region models

So far we have considered the effects of immigration on knowledge capital formation without allowing for interactions between regions. Lundborg and Segerstrom (1998) proposed a quality-ladder endogenous growth model with North–South trade and migration from the South to the North. In their framework, immigration in the North lowers Northern wages and therefore spurs R&D employment. This is growth-enhancing but welfare-reducing. Baldwin and Forslid (2000) analyzed the role of

[82] See Grossmann (2009) for a R&D-based growth model with entrepreneurial firms, which neither features such specialization gains nor, as a result, positive scale effects.

[83] See Grossmann (2013) for further discussion. In view of (18.46), the facts that l^Y is time-invariant in the long run and $\overline{A}_{t+1} = \overline{A}_t \Lambda\left(\tilde{l}\right)$, the steady-state growth rate of income per capita is given by $g = (1+n)^{\frac{\eta-\beta}{1-\alpha}} \Lambda\left(\tilde{l}\right) - 1$. Thus, if $\eta > \beta$, g is decreasing in the population growth rate, n.

endogenous horizontal innovations in the spatial framework of Krugman (1991b) for the stability of equilibrium. Interregional knowledge spillovers are necessary for the stability of a symmetric equilibrium. The economy's knowledge stock is higher in a core–periphery equilibrium (i.e., in which workers have migrated).

From a theoretical point of view, allowing for labor mobility in endogenous growth models with more than two regions is challenging. Desmet and Rossi-Hansberg (2014) provided a novel approach in which there is free mobility of labor across sectors and across a continuum of regions ordered in a unit interval. Firms compete for land, which is necessary for production. Moreover, they can invest in productivity-enhancing innovation at their location. Technology spatially diffuses after a one-period delay subject to costs of communication and transport. The analysis suggests that R&D investment is higher in locations with more employment. Moreover, the rate of productivity growth is increasing in the costs of spatial technology diffusion. The reason is that these costs are conducive to concentration in the presence of congestion in land use.

Future research should certainly extend such analyses of dynamic spatial models to analyze in more depth the interaction between migration and the formation of knowledge capital. The present literature is still at an early stage and provides little guidance for structural estimations, particularly in view of technical difficulties and the inherent potential of multiple equilibria in models with increasing returns.

5.1.4 Empirical evidence

Jones (1995, 2005) has questioned the property of strong scale effects arising in the first class of endogenous growth models, according to which a higher population size (of skilled workers) should raise the economy's growth rate. He pointed out that the hypothesis is at odds with the post-World War II experience of advanced countries where the number of R&D workers has risen substantially while TFP growth has remained remarkably stable. However, employing US panel data for the period 1940–2000, Hunt and Gauthier-Loiselle (2010) found that an increase in the share of immigrant college graduates by 1 percentage point raises patents per capita by 9–18%. Consistent with at least the weak form of positive scale effects from migration, by using international data on bilateral migration stocks, Grossmann and Stadelmann (2013) argued that high-skilled migration has a positive (albeit small) effect on the relative destination-to-source level of both income and TFP.

Ortega and Peri (2011, 2014) disentangled the effects of migration and trade on both GDP per capita and TFP. They took into account that openness to migration and openness to trade are highly related and both kinds of bilateral flows are determined by geographical and cultural distance between countries. Ortega and Peri (2011) exploited the year-to-year variation by using a longitudinal dataset including the OECD and their main trade and migration partners. Their instrumental variable estimations suggest that, in the short run, immigration has a negative effect on TFP but a positive effect on the employment rate. The result on TFP is consistent with short-run congestion effects. Ortega and Peri (2014) showed, however, that in the long run and in contrast to trade, immigration has a positive and robust effect on GDP per capita. Interestingly, the mechanism works mainly through the effect on TFP and can be attributed to the impact of immigration on the diversity of productive skills and innovation activity.

In sum, while congestion effects may play some role in the short run, there is clear evidence for a positive long-run effect of immigration on the formation of knowledge capital and per-capita income.

5.2 High-skilled migration and human capital formation

In an important early contribution, Bhagwati and Hamada (1974) analyzed migration effects in a static context with rigid wages and endogenous education. They argued that outward migration has adverse effects on per-capita income and fosters unemployment. More recently, scholars have directed attention to potentially positive effects of emigration of skilled workers on human capital formation, which could result in a gain for the source economy (e.g., Mountford, 1997 Stark et al., 1997 Beine et al., 2001, 2008). We illustrate the basic mechanism for this possibility in a simple dynamic framework before discussing empirical evidence.

5.2.1 A simple dynamic model

Consider a perfectly competitive developing economy. We analyze capital flows along with migration and the education decision by employing the standard notion that physical capital, K, and raw labor, L, are good substitutes. To capture this assumption in its sharpest form, we assume that K and L are perfectly substitutable such that output depends on a simple composite of both factors, $X = K + L$:

$$Y = AF(X, H) = AHf(\kappa), \tag{18.47}$$

where $A > 0$ is a TFP parameter, H is the number of skilled workers remaining in the country after emigration, and $\kappa \equiv X/H$. F is a linearly homogeneous function and $f(\kappa) \equiv F(\kappa, 1)$ is increasing and strictly concave. As is typically assumed in the "brain gain" literature, only skilled labor may migrate. Due to immigration quotas abroad, an individual expects to migrate with a probability $p \in (0,1)$, which *ex post* is the fraction of migrants among the skilled population. There are no mobility costs. International integration of labor markets for skilled workers is modeled as an exogenous increase in p. Physical capital is internationally mobile and the interest rate, $r > 0$, is exogenous.

As the marginal product of capital, $Af'(\kappa)$, is equal to the interest rate, r, we have $\kappa = (f')^{-1}(r/A) \equiv \widetilde{\kappa}(A)$,[84] where $\widetilde{\kappa}' > 0$. Thus, the wage rate per unit of skilled labor reads as

$$w_H = A[f(\widetilde{\kappa}(A)) - \widetilde{\kappa}(A)f'(\widetilde{\kappa}(A))] \equiv \widetilde{w}_H(A). \tag{18.48}$$

The function $\widetilde{w}_H(A)$ is increasing in the TFP parameter A. Moreover, the wage rate for unskilled labor is given by $w_L = r$, due to perfect substitutability with physical capital. Thus, for a given level of TFP, the wage rates for both types of labor do not depend on domestic employment of workers, H or L. Any employment change triggers adjustment of the physical capital stock such that the marginal product of skilled and unskilled labor is unaffected. Hence, any effect of labor market integration on the wage rate of skilled labor, w_H, must stem from adjustments in TFP.

Each period, a mass one of natives is born. Each individual lives two periods in overlapping generations. In the first period, individuals decide whether to become skilled, whether to migrate, and how much to save for old age. In the second period, individuals are retired and consume their savings. Preferences of individual i born in period $t = 1, 2, \ldots$ are characterized by the intertemporal utility function

$$U_t(i) = u(c_{1t}(i)) + \theta u(c_{2t+1}(i)), \tag{18.49}$$

where $c_{1t}(i)$ and $c_{2t+1}(i)$ are consumption levels in the first and second periods of life respectively. The instantaneous utility function u is increasing and concave. For simplicity, we assume that the time preference rate is related to the interest rate according to the standard condition $\theta(1+r) = 1$. Thus, optimal savings of an individual with first-period (labor) income $y_t(i)$ imply $c_{1t}(i) = c_{2t+1}(i) = y_t(i)/1+\theta$. Intertemporal lifetime utility is increasing in income; we have $U_t(i) = (1+\theta)u(y_t(i)/1+\theta) \equiv V(y_t(i))$ with $V' > 0$.

Unskilled workers inelastically supply one unit of labor, whereas a skilled individual i supplies $1 - e(i)$ units of labor. Time costs of education, e, are distributed according to the cumulative distribution function $\Phi(e)$. Denote by w^* the wage rate of skilled migrants per unit of labor supplied abroad. We assume that w^* is time-invariant, $w^* > w_H = \widetilde{w}_H(A)$ and $w^* > w_L (= r)$. An individual i acquires schooling if and only if

$$p . V((1-e(i))w^*) + (1-p) \cdot V((1-e(i))\widetilde{w}_H(A)) \geq V(w_L) = V(r). \tag{18.50}$$

Thus, an individual i becomes skilled when individual time cost $e(i)$ is below some threshold level $\bar{e} = \bar{e}(A, p, w^*)$, which rises in the level of TFP, A, the emigration quota, p, and the foreign wage rate, w^*. Thus, the share of natives who acquire schooling,

[84] For notational simplicity, we suppress the interest rate, r, in functions we define in the remainder of this section.

$$s = \Phi(\bar{e}(A, p, w^*)) \tag{18.51}$$

is increasing in A, p and w^*. An increase in the emigration quota, p, raises the education incentive by raising the prospect of receiving a higher wage abroad ($w^* > w_H$), conditional on being skilled.

Denote the share of skilled workers in total population after emigration took place by h. Suppose that TFP is determined according to $A_{t+1} = h_t$. The initial level of TFP, $A_0 > 0$, is given. Note that $H = (1-p)s$ is the economy's number of skilled workers after migration took place and $1 - ps$ is the size of the population remaining in the country after migration. Thus, the evolution of TFP over time is governed by

$$A_{t+1} = h_t = \frac{(1-p)s_t}{1-ps_t} = \frac{(1-p)\Phi(\bar{e}(A_t, p, w^*))}{1-p\Phi(\bar{e}(A_t, p, w^*))} \equiv \widetilde{h}(A_t, p, w^*). \tag{18.52}$$

The share of skilled workers after migration, $h = \widetilde{h}(A_t, p, w^*)$, is increasing in the contemporaneous TFP level, A.

Can labor market integration for skilled workers raise the TFP level in the source country, therefore boosting the wage rate of skilled workers, $w_H = \widetilde{w}_H(A)$? Suppose that u and Φ are such that $\widetilde{h}(0, \cdot) > 0$ and $\partial h/\partial A < 1$. In this case, there exists a unique and stable long-run equilibrium TFP level. Consequently, during transition as well as in steady state, the equilibrium level of A is increasing in the migration quota p, if $\widetilde{h}(A_t, p, w^*)$ is increasing in p. There are two counteracting effects. First, an increase in p lowers the economy's share of skilled workers after migration, $h = (1-p)s/1 - p \cdot s$, for a given share of skilled natives, s. This captures the standard "drain effect". Second, a higher migration quota raises $s = \Phi(\bar{e}(A, p, w^*))$. If this "gain effect" dominates the "drain effect", labor market integration benefits skilled workers remaining in the source country.

If the effect of an increase in immigration quota p on the equilibrium share of skilled natives (s) is large enough, skilled employment, $H = (1-p)s$, may be increasing in p. Similarly, the physical capital stock, $K = H\widetilde{\kappa}(A) - L$,[85] may increase in p for three reasons that are related to an increased schooling incentive. First, because skilled labor is complementary to physical capital. Second, because inflows of physical capital substitute for the declining number of unskilled workers, $L = 1 - s$. And third, because of the intertemporal effect of higher education on the TFP level, reflected in (18.52).

To sum up, if emigration prospects of skilled workers in developing countries are uncertain due to immigration quotas in advanced countries, better emigration prospects foster incentives to acquire schooling. The drain effect on the average human capital stock from higher outflows may then be dominated by an increase in the number of skilled natives. In this case, higher brain drain may go along with inflows of physical capital.

[85] Recall that $\kappa = K + L/H$ by definition and $\kappa = \widetilde{\kappa}(A)$.

5.2.2 Empirical evidence

Beine et al. (2001, 2008) (provided empirical evidence that strongly suggests that a higher emigration rate of skilled workers (the stock of tertiary educated emigrants divided by the size of the skilled population) who live in OECD countries positively affects human capital formation. Whereas Beine et al. (2001) reported that the effect is higher for poor countries, Beine et al. (2008) did not find support for non-linearities. Beine et al. (2008) (instrumented the skilled emigration rate by total population size (capturing that immigration quotas in OECD countries are relatively higher for smaller source countries) and by the total stock of migrants (capturing network effects of migrants). According to their estimates, doubling the instrumented emigration rate of skilled workers in 1990 raises the pre-migration share of skilled workers in 2000 relative to the one in 1990 by 5%.

Based on this estimate, Beine et al. (2001, 2008) provided simulation results on the counterfactual share of skilled workers and compared it with the actual share. Their results suggest that only some countries, typically those combining low human capital levels and low emigration rates, may experience gains from increased migration prospects, albeit very small ones. Importantly, the majority of developing countries loses, sometimes quite substantially. One may conclude that for developing source countries the gain effect of higher immigration quotas for skilled labor in advanced destination countries is typically almost equal to or smaller than the drain effect.

6. MIGRATION IN THE GRAVITY EQUATION OF TRADE

A substantial empirical literature studies the link between international trade and international migration. Usually it aims at identifying the channels that may rationalize the estimated relationship, and/or testing for causation running from bilateral migration (stocks or flows) to bilateral trade flows. The underlying formal structure for this literature is that of the so-called gravity equation: a log-linear relationship linking the trade flows between two countries to economic determinants, political variables, and geography. Using standard linear econometric models, the gravity equation is easy to implement empirically. Moreover it is fairly general, since it can be derived from a broad class of models that differ with respect to details of their microstructure. It is therefore no wonder that most of the empirical evidence on the relationship between migration and trade is based on the gravity model.[86] Before we discuss the empirical evidence, a short review of the conceptual foundation of the gravity equation is warranted.

[86] The focus of this section is on the analysis of aggregate bilateral trade data for large cross-sections or panels of country pairs. Increasingly, researchers also work with regional data for single countries. We will briefly present these newer studies in our discussion below.

6.1 Conceptual foundation of the gravity equation

Tinbergen (1962) expressed imports of country i to country j, M_{ij}, as directly proportional to the product of the GDPs of the two countries ($Y_i Y_j$) and indirectly proportional to geographical distance D_{ij}

$$M_{ij} = G\frac{Y_i Y_j}{D_{ij}^{\delta}}, \tag{18.53}$$

where G is a constant and δ is a parameter.[87] This formulation is formally akin to Newton's law of gravity in physics. In that context, M_{ij} is the force between two objects i and j, Y_i and Y_j are their respective masses, G is the gravitational constant, and δ is set equal to 2. Researchers have typically estimated (18.53) by adding a multiplicative error term and by applying OLS to the log-linearized model.

The estimation results obtained suggest that the size of markets—as captured by Yi and Y_j—and the force of trade costs—as captured by D_{ij}—matter strongly for the size of trade flows between countries. For example, regressing the log of exports on the log of GDPs and the log of distance in a sample of 114 countries for the year 2000 delivers coefficients on GDPs close to unity and an estimate of δ at -1.37.[88] The simple regression explains about 65% of the cross-sectional variation in trade flows; this is a very good fit for such a simplistic model.[89] Therefore, Anderson (2011) described the gravity equation "as one of the most successful empirical models in economics".

However, the theoretical rationale for the gravity equation was not very clear until Anderson and van Wincoop (2003) provided a rigorous underpinning in full general equilibrium. Earlier theoretical derivations have either not included a full treatment of trade costs or have not made use of the market clearing conditions. This chapter is not the right place for a survey of recent theoretical developments; see Anderson and van Wincoop (2004), Bergstrand et al. (2013) or the recent handbook chapter by Head and Mayer (2013). Here it suffices to mention that, as explained by Anderson (2011), the key simplification leading to tractable gravity representations is *modularity*. This means that the pattern of trade costs can be inferred from bilateral trade flows without at the same time having to explain total supplies of goods to *all* destinations or the total demand for goods from *all* origins. This property, often also referred to as trade *separability*, requires restrictions on demand or supply-side structure and assumptions on trade costs, in particular if there are multiple classes of goods. One frequent assumption that works is

[87] This section draws heavily on Felbermayr et al. (2010a) and Felbermayr and Toubal (2012).

[88] Coefficients on log GDPs are numerically close to 1, but statistically they are different from 1 at the 1% significance level.

[89] The goodness of fit rises above 80% if a more homogeneous sample of countries is used. It increases even more if additional determinants of trade costs such as dummies for common language or country contiguity are included.

that delivery of goods uses resources in the same proportion as the production of those same goods. Iceberg trade costs, as introduced by Samuelson (1952), meet this production proportionality criterion. The literature on the trade and migration nexus almost universally assumes that migration affects trade costs, and in that way the iceberg assumption is one of major importance in the present context.

The usual demand-side structure requires that cross effects in demand between classes of goods operate only through aggregate price indexes. This is the case whenever preferences or technology are homothetic and weakly separable across classes of goods defined by their location of production. This is the so-called Armington assumption. It describes a situation where goods are differentiated by place of origin, such as in the Anderson and van Wincoop (2003) perfect competition model. An often used model that also meets demand modularity is the Krugman (1980) model, where firms operate under monopolistic competition and where there is free entry that gives rise to a gravity equation that is isomorphic to the one derived under perfect competition. In both models, the elasticities of trade flows with respect to iceberg trade costs is given by $1 - \sigma$, where σ is the elasticity of substitution of varieties amongst one class of goods (i.e., within the same sector). A further important example is the Melitz (2003) model of heterogeneous firms, which gives rise to a similar gravity equation, provided that firm-level productivities are Pareto distributed. In that framework, the elasticity of trade flows with respect to trade costs is the negative of the Pareto shape parameter Chaney (2008). This result is obtained despite the existence of a product-level extensive margin in that framework.

Alternatively, one can restrict the supply side so that, in equilibrium, the share of goods traded between two countries is solely pinned down by the supply side. This is the gravity model derived by Eaton and Kortum (2002), which gives rise to a structure that is again mathematically isomorphic to the Armington model with the trade flow elasticity now given by the negative of the shape parameter of the Fréchet distribution describing country-level Ricardian productivity levels. Summarizing, the gravity model that we will use holds in models featuring an extensive margin along with the intensive margin; it holds also under monopolistic and perfect competition.

Given separability, multi-sector versions of the gravity model look also isomorphic to the one describing aggregate trade flows, with the only difference that sectoral rather than aggregate expenditure and revenue must be used. Moreover, since the basic framework features no dynamic link between trade and production (e.g., through adjustment in capital stocks or TFP), to turn the standard equation into a panel model one just has to add time indexes.[90]

The presence of migrants can promote trade between their source and their host countries in at least three ways. First, they might help overcome informal barriers to

[90] Olivero and Yotov (2012) offered a dynamic gravity model and discuss its proper estimation.

international trade related to language, culture, or institutions, they may facilitate the creation of business relationships, and they may make valuable information on foreign sales and sourcing opportunities more readily available (Rauch and Trindade, 2002; Combes et al., 2005; Heranda and Saavedra, 2005; Dunlevy, 2006). Thus, migrant networks mitigate *incomplete* information. Second, they can also attenuate frictions due to *asymmetric* information and the potentially disruptive opportunistic behavior that those frictions entail in the absence of enforceable property rights. These frictions can reduce the volume of transactions on a market beyond the socially desirable level. To the extent that ethnic networks provide an enforcement mechanism, for example by excluding members from their social and economic benefits, they can nudge the volume of trade closer to the social optimum. In that sense, migrant networks substitute for markets. These first two channels affect the effective transaction costs between two countries. The third channel, in contrast, relates to *preferences*: Migrants may boost trade if they derive higher utility from goods produced in their host countries (Gould, 1994; Head and Ries 1998; Girma and Yu, 2002; Wagner et al., 2002). Quantifying the relative importance of these mechanisms is important, since trade creation due to the alleviation of informational barriers and frictions constitutes a source of welfare gains for the host and source country. If trade is higher due to specific features of preferences, the endogeneity of the welfare criterion renders traditional welfare analysis impossible.

The literature on the trade–migration nexus has made increasing use of the gravity model of bilateral international trade. We follow Combes et al. (2005) and introduce a bilateral affinity parameter into the usual Dixit–Stiglitz utility function of the representative household. This parameter may depend on bilateral ethnic ties, thereby capturing the preference channel described above. We also allow bilateral trade costs to depend on migration; this is meant to account for the information channel described above.

More precisely, we assume that the representative agent in country i has a Dixit–Stiglitz utility function defined over domestic and imported varieties

$$U_i = \sum_{j=1}^{C}\sum_{h=1}^{n_j}\left(a_{ij}m_{ijh}\right)^{\frac{\sigma-1}{\sigma}} = \sum_{j=1}^{C} n_j\left(a_{ij}m_{ij}\right)^{\frac{\sigma-1}{\sigma}}, \sigma > 1,\ a_{ij} \geq 1, \qquad (18.54)$$

where m_{ijh} denotes consumption of good h sent from country j to country i, n_j is the number of varieties available from country j, C is the number of countries, and a_{ij} is the relative weight that consumers in i attach to goods delivered from country j. Let the consumer price in country i be given by $p_{ij} = p_j T_{ij}$, where p_j is the price accruing to producers in country j and $T_{ij} > 1$ indicates that T_{ij} units of a good need to be shipped from j for one unit to be delivered in i. Assuming that all varieties are produced with the same technology, so that they command the same price, we have $m_{ijh} = m_{ij}$. This implies the second equality above.

Maximizing (18.54) subject to the budget constraint $Y_i = \sum_{j=1}^{C} n_j p_{ij} m_{ij}$, country i's demand for a variety produced in country j is given by $m_{ij} = a_{ij}^{\sigma-1} p_{ij}^{-\sigma} P_i^{\sigma-1} Y_i$.

A higher preference parameter a_{ij}, leads to higher demand, while a higher price inclusive of trade costs p_{ij}, reduces demand. The aggregate price index is given by $P_i^{1-\sigma} = \sum_{j=1}^{C} a_{ij}^{\sigma-1} n_j p_{ij}^{1-\sigma}$.

Substituting $p_{ij} = T_{ij} p_j$, into optimal demand, total imports $M_{ij} = m_{ij} n_j p_{ij}$ can be expressed as

$$M_{ij} = \left(T_{ij} p_j\right)^{1-\sigma} n_j a_{ij}^{\sigma-1} P_i^{\sigma-1} Y_i, \tag{18.55}$$

which is a first gravity representation of bilateral trade. Note that trade depends on the price of single varieties (p_j) and on their number (n_j).

To make further progress, Anderson and van Wincoop (2003) proposed to use the market clearing conditions $y_i = \sum_{i=1}^{C} m_{ij} T_{ij}$, which make sure that production of a generic variety in country j (denoted by y_i) is exactly absorbed by demand for goods and transport services from all over the world. These C equations could in principle be used to solve for C endogenous prices p_j. But rather than working with explicit solutions, it is convenient to use an implicit solution to the market clearing conditions proposed by Anderson and van Wincoop (2003)

$$p_j^{1-\sigma} = \frac{Y_j}{Y_W n_j} \widetilde{P}_j^{\sigma-1}, \tag{18.56}$$

where (as using this expression in the aggregate price index shows) $\widetilde{P}_j$, is given by

$$\widetilde{P}_j^{1-\sigma} = \sum_{i=1}^{C} \frac{Y_i}{Y_W} \left(\frac{T_{ij}}{a_{ij}}\right)^{1-\sigma} \widetilde{P}_i^{\sigma-1}. \tag{18.57}$$

Expression (18.56) allows substituting $n_j p_j^{1-\sigma}$ out of the gravity equation (18.55). This yields

$$M_{ij} = \frac{Y_i Y_j}{Y_W} \left(\frac{T_{ij}}{a_{ij}}\right)^{1-\sigma} \left(\widetilde{P}_i \widetilde{P}_j\right)^{\sigma-1}. \tag{18.58}$$

To see that (18.56) is indeed consistent with market clearing, one can use the market clearing condition $Y_j = n_j \sum_{i=1}^{C} m_{ij} p_{ij}$, which states that the value of output in country j (GDP) must be equal to the value of sales. Substituting for m_{ij} and replacing $p_j^{1-\sigma}$ by the expression in (18.56) one obtains $Y_j = \left(Y_j / Y_W\right) \sum_{i=1}^{C} Y_i \left(\widetilde{P}_i \widetilde{P}_j\right)^{\sigma-1} \left(T_{ij} / a_{ij}\right)^{1-\sigma}$. Summing over all j, recognizing that $\sum_{j=1}^{C} Y_j = Y_W$, and rearranging, one obtains exactly (18.57).

Anderson and van Wincoop (2003) called $\widetilde{P}_i$ and $\widetilde{P}_j$ indices of inward and outward *multilateral* resistance respectively, because they depend on the trade costs of country i with all countries in the world. We are interested by the determinants of T_{ij} in general, and by the cost of obtaining information in particular. Following the literature, we assume that T_{ij} is a log-linear function of its determinants (see below). Also in line with almost all previous work, trade costs are modeled as incurring the use of real resources. While this choice subsumes *ad valorem* tariffs in the present context (as will become clearer below), it does not easily provide for a meaningful role of nominal exchange rates as this would require to model sticky nominal prices.

The central insight of Anderson and van Wincoop (2003) was that the volume of trade between i and j depends not only on the trade costs between i and j but on the entire distribution of trade costs between i and j and *all other countries* of the world. How strongly T_{ij} restricts trade between i and j depends on the costs that affect trade with alternative partners. Hence, in the estimation we have to deal with the $\widetilde{P}_i$ terms.

Equation (18.58) can be understood as a model of exports simply by exchanging indices i and j. One can also work with a measure of total trade by taking the geometric average of (18.58). This yields

$$M_{ij}^{1/2} M_{ji}^{1/2} = \left(\frac{T_{ij}}{a_{ij}}\right)^{\frac{1-\sigma}{2}} \left(\frac{T_{ji}}{a_{ji}}\right)^{\frac{1-\sigma}{2}} \left(\widetilde{P}_i \widetilde{P}_j\right)^{\sigma-1}. \tag{18.59}$$

The simple arithmetic mean $M_{ij} + M_{ji}$ leads to an additive structure that does not lend itself to log-linearization unless $T_{ij} = T_{ji}$ and $a_{ij} = a_{ji}$, which is a problematic assumption in the current context (see below).

6.1.1 The trade cost channel of migration

We assume that *ad valorem* trade costs T_{ij} depends on traditional factors such as transportation costs and variables describing the stance of trade policy or cultural proximity. The gravity literature discusses different ways to measure the former variables, usually using geographical distance, a dummy for a common border (adjacency), a dummy for the use of a common language, a dummy for joint membership in a free trade agreement (FTA) or in the World Trade Organization (WTO).

We posit that T_{ij}, depends on costs related to informational frictions, I_{ij} which may be affected by migrant networks, as described before. A migrant network is made up of bilateral links between agents of similar ethnic origin. Denote by I_{ij}, the availability of information on trading opportunities between i and j,

$$I_{ij} = \iota\left(N_{ij}, N_{ji}\right) \text{ with } i, j = 1, \ldots, C, \tag{18.60}$$

where N_{ij}, describes the stock of individuals from country (or ethnicity) j residing in i. The formulation (18.60) accounts for *direct* links, i.e., immigrants from j residing in i and emigrants from i residing in j may affect the availability and quality of information I_{ij} relevant for trade costs pertaining to imports of goods from j into i. For example, Combes et al. (2005) made this choice. In general, however, information relevant for trade between i and j can also be conveyed by *indirect* networks of ethnicity $k \neq i, j$. The most famous such network studied in the literature (e.g., Rauch and Trindade, 2002) is that of the Chinese, who have been found to increase trade between countries in which they reside, even if none of these countries is mainland China itself.

There is substantial uncertainty about how to model the dependence of T_{ij} on different proxies of trade costs. Denote by L_i the total resident population in country i, regardless of the ethnic group that residents may belong to. Then, $s_{ik}=N_{ik}/L_i$ denotes the probability that a randomly chosen individual residing in country i belongs to ethnicity k. Hence, $s_{ik}s_{jk}$ denotes the likelihood that two simultaneously drawn individuals residing in countries i and j respectively have the same ethnicity. So, $s_{ik}s_{jk}$ measures the probability of a co-ethnic contact and hence the strength of the link between i and j. In this concept, we regard k as the ethnic hub, and i, j as ethnic spokes. By construction, the hub is the country where the ethnicity k forms the majority.

Most of the literature disregards indirect links and assumes a functional form with constant elasticities

$$\ln I_{ij} = \ln \iota\left(N_{ij}, N_{ji}\right) = \overline{\mu}_1 \ln N_{ij} + \overline{\mu}_2 \ln N_{ji}, \tag{18.61}$$

where we expect the parameters $\overline{\mu}_1$ and $\overline{\mu}_2$ to be positive. Alternatively, one can also posit

$$\ln \iota\left(s_{ij}, s_{ji}\right) = \overline{\mu}_1 \ln s_{ij} + \overline{\mu}_2 \ln s_{ji} = \overline{\mu}_1 \ln N_{ij} + \overline{\mu}_2 \ln N_{ji} - \overline{\mu}_1 \ln L_i - \overline{\mu}_2 \ln L_j. \tag{18.62}$$

This latter specification has the plausible implication that information costs do not depend on the size of the two economies that form a trade relationship. In other words, the information-related tariff equivalent is invariant to a proportional increase in countries' total and foreign-born populations.[91] It is in line with the idea that the pro-trade effects of migrants' networks are the larger, the higher the probability to meet a migrant coming from a partner country. However, it imposes a stronger functional form, since it assumes that the elasticities on L_i. and L_j are the exact negatives of those on N_{ij} and N_{ji}. In practice, the terms $\overline{\mu}_1 \ln L_i$ and $\overline{\mu}_2 \ln L_j$ are often absorbed by importer and exporter fixed effects anyway, so that little is gained by imposing this additional structure.

The literature typically postulates a functional form of T_{ij} such as

$$T_{ij} = \overline{T}_{ij} I_{ij} D_{ij}^{\overline{\delta}} e^{\overline{\lambda}\left(1-LANG_{ij}\right)} e^{\overline{\gamma}\left(1-ADJ_{ij}\right)} e^{\overline{\pi}\left(1-FTA_{ij}\right)}, \tag{18.63}$$

where D_{ij} measures geographical distance, $LANG_{ij}$ is a dummy that takes value 1 if a sufficiently large portion of residents in the two countries i and j speak the same language, ADJ_{ij} is a contiguity dummy, and FTA_{ij} is a dummy that captures joint membership in a free trade agreement. Other determinants of trade costs (such as colonial history, membership in various bilateral or multilateral agreements, proxies for infrastructure, etc.) are

[91] Combes et al. (2005) specified the function $\iota(.,.)$ in levels rather than in shares. This implies that for a given composition of the workforce in the importer or exporter country, larger countries (who tend to receive and send more immigrants) have smaller iceberg trade costs.

easy to include in a similar log-linear fashion. $\overline{T}_{ij}$ captures unobserved and hence omitted determinants of bilateral trade costs. Substituting for I_{ij} and taking logs, we obtain

$$\ln T_{ij} = \xi_T' \mathrm{X}_{ij} - \nu_T^{im} \ln N_{ij} - \nu_T^{em} \ln N_{ji}, \tag{18.64}$$

where ξ_T is a vector of coefficients on controls X_{ij}, N_{ij} measures the strength of the immigrant network and N_{ji} that of the emigrant network on trade costs (expected to be positive). Evidence in favor of $\nu_T^{im} > 0$ and/or $\nu_T^{em} > 0$ would suggest that migrant networks lower informational or contractual costs, thereby encouraging trade through lower total trade costs.[92] This is the *trade cost channel* of networks.

6.1.2 The preference channel of migration

The preferences channel of migration works through the bilateral affinity parameter a_{ij} in the utility function (18.54). We assume that a_{ij} is, amongst other things, a function of the share of immigrants. It is plausible that immigrants have a special preference for varieties produced in their source countries. Also, there could be a demonstration effect by which natives become aware of foreign varieties. Since we work with a representative agent framework, a higher share of foreign-born individuals in the population means that preferences are more strongly tilted towards the host country of those individuals:

$$\ln a_{ij} = \xi_a' \mathrm{X}_{ij} + \nu_a N_{ij}, \tag{18.65}$$

where ν_a is expected to be positive.[93] The idea is that country i's cultural, political, or geographical proximity to country j increases the weight of goods imported from i. Evidence for $\nu_a > 0$ would be in line with the existence of a *preference effect* of ethnic networks. The formulation implies that there is no systematic bias for imports from any country unless there is a strictly positive stock of foreign-born individuals from that country residing in country i. This captures the home country bias that immigrants may have; it is also consistent with the idea that the presence of immigrants in some country may on its own tilt the preferences of natives towards goods typically consumed by those immigrants. Similar to Combes et al. (2005), this formulation disallows for a special preference for varieties produced in countries with a stock of expatriates. Realistically, immigrants may attach a particular weight to varieties produced in their host countries. In contrast, it is difficult to find a convincing reason why source country consumers should specially value goods produced in the country where emigrants reside.

Egger et al. (2012) argued that the functional forms implied by (18.64) and (18.65) are restrictive. In particular, they argue that additional trade due to the preference channel should be proportional to the number of immigrants as they create a market for imports.

[92] The superscripts *im* and *em* are used to distinguish the effects of immigrant and emigrant networks.
[93] The network channel cannot operate against the direction of trade whence $\nu_a^{em} = 0$ and we can drop indices on ν_a.

In contrast, when trade creation is due to a reduction of trade frictions, such as, e.g., of the informational type, the effect should be declining in the number of migrants as the marginal impact of additional migrants is less strong. In their empirical analysis, they apply semi-parametric methods and find some support for this conjecture. In this survey, however, we stick to parametric models.

6.1.3 Econometric issues

Estimation of an equation such as (18.58) poses a number of problems. First, while theoretical considerations imply the precise structure of (18.58), they tell us nothing about the correct specification of T_{ij}. The formulation (18.63) is common, but prone to specification error. The same holds true for (18.65). Also, one must make assumptions on the intercepts of the trade cost function $\left(\overline{T}_{ij}\right)$ and of the preference function $\left(\overline{A}_{ij}\right)$, which represent the unmodeled parts of costs and preferences. These are usually assumed to be included in the error term, i.e., they must be orthogonal to the other components of trade costs or preferences. Alternatively, in panel models, one can decompose the terms into a time-invariant part (fixed or random) and an error term. We will return to this issue below.

Moreover, the multilateral resistance terms $\widetilde{P}_i$ are essentially unobserved since they do not correspond to official CPI deflators. Anderson and van Wincoop (2003) showed how one can solve for the $\widetilde{P}_i$ terms numerically and use them in an iterative estimation strategy. They demonstrated that the failure to control for multilateral resistance typically biases the absolute value of estimated trade cost coefficients upwards. Rauch and Trindade (2002) recognized the problem of multilateral resistance (without mentioning the issue) by adding an ad-hoc remoteness term to their regressions. *Ex ante*, it is unclear whether this is sufficient to deal with omitted variable bias. In our regressions, we follow Feenstra (2004b) and Anderson (2011), who argued that the use of importer and exporter specific fixed effects in a simple OLS model leads to very similar results as the Anderson and van Wincoop (2003) strategy, but is technically much less demanding as well as more general as other sources of unobserved country-level heterogeneity may exist. We opt for this strategy, which is now common in virtually all gravity applications. Some researchers add separate importer and exporter fixed effects into the model; this is the most general specification, but comes with substantial loss of degrees of freedom. Alternatively, one can also use country dummies which, to the extent that trade costs are symmetric, yield identical results; see Baier and Bergstrand (2009) for a similar strategy.[94]

[94] There are alternative ways to deal with multilateral resistance indices. Baier and Bergstrand (2009) have proposed working with first-order expansions of the non-linear resistance terms and to directly control for them in the model. This has the advantage that the researcher can still identify the effects of country-specific variables. This is, however, not of central importance in the present context. Another way to deal with multilateral resistance terms is to employ an approach advocated by Combes et al. (2005), where trade flows M_{ij} and M_{ji} are divided by M_{ii} and M_{jj} respectively, and the resulting fractions are multiplied by each other. In the ensuing equation, the unobservable multilateral resistance terms will have dropped out.

Employing the specifications for T_{ij} and a_{ij} as given in (18.64) and (18.65) in (18.58) and using non-overlapping sets of country dummies μ_i and μ_j to control for the country-specific (multilateral) variables, we have

$$\begin{aligned}\ln M_{ij} = \ln\left(\frac{Y_i Y_j}{Y^w}\right) + (\sigma-1)\left(\xi_a' - \xi_T'\right)\mathrm{X}_{ij} + \\ (\sigma-1)\left[\left(\nu_a + \nu_T^{im}\right)\ln N_{ij} + \nu_T^{em}\ln N_{ji}\right] + \mu_i + \mu_j\varepsilon_{ij} \\ = \ln\left(\frac{Y_i Y_j}{Y^w}\right) + \overline{\xi'}\mathrm{X}_{ij} + \overline{\nu}^j \ln N_{ij} + \overline{\nu}^j \ln N_{ji} + \mu_i + \mu_j + \varepsilon_{ij}.\end{aligned} \tag{18.66}$$

Alternatively, one can also normalize trade flows by Y_iY_j so that the term ln (Y_iY_j) disappears from the right-handside of the equation. We will be interested in estimates of the parameters ν_a, ν_T^{im} and ν_T^{em}, which are consistently estimated by OLS under the assumption that $\{\mathrm{X}_{ij}$; ln N_{ij}, ln $N_{ji}\}$ are uncorrelated to ε_{ij}. We will return to the validity of this assumption below.

The inclusion of exporter and importer dummies μ_i and μ_j is easy to implement, but may have drawbacks. If the number of countries C is large but the sample is unbalanced in the sense that there are substantially fewer non-zero trade flows than the potential full number of trade flows (i.e., $C\,(C-1)$), identification of these dummies may rely on very few observations, reducing the efficiency of the procedure and making inference more difficult. One way to deal with this difficulty is to use country dummies that take value 1 if a country is part of a dyad, regardless of its role as either an exporter or an importer. This would be completely innocuous if $T_{ij}=T_{ji}$ and $a_{ij}=a_{ji}$, a strong assumption in our context. Nonetheless, using country dummies may be a sensible compromise when degrees of freedom are scarce.

Note that the calculation of the inward and outward multilateral resistance terms in (18.58) requires knowledge of trade costs between *all* trading countries. Even if one is interested in imports of a single country k from many source countries in the world, one needs estimates of the outward resistance terms of those source countries. This requires information about their trade costs with all other countries. In other words, to know the trade effects of immigration in one country one needs data on other countries' trade and immigration as well. One can of course estimate a model of $k's$ imports from many destinations and use destination dummies to take care of multilateral resistance. However, these dummies will reflect trade costs with k only, while in reality they should reflect trade costs with the whole world. This means that estimation of a model such as (18.66) based on trade data of only a single country is problematic. Nonetheless, for data reasons, this is what most of the literature has been doing so far.

When a full matrix of bilateral trade flows and of bilateral migration stocks is available, that is, if one has all countries' imports from all possible sources, it does not make sense to distinguish between imports and exports since i' imports from j are exactly $j's$ exports to i: However, it is meaningful to include measures of immigration and emigration in the estimated equation. Finding a positive coefficient on the stock of immigrants from j residing in i means that immigration positively affects imports of i from j. This is equivalent to saying that emigration from j to i fosters $j's$ exports to i. However, it is still meaningful to also include the stock of emigrants as they may facilitate trade as well. The literature considering immigration of a single country instead differentiates between import flows and exports.

Equation (18.66) is often estimated on pooled data. In that case, all variables in the model obtain a year or period index. The exporter and importer dummies now have to be interacted with time dummies v_t, so that (18.66) would include the terms $\mu_i \times v_t$ and $\mu_j \times v_t$ Separate inclusion of v_t is redundant. If applying panel techniques, the error term is usually decomposed as $\epsilon_{ijt} = \eta_{ij} + \varepsilon_{ijt}$. Most authors treat unobserved country-pair heterogeneity η_{ij} as fixed rather than random; this is indeed what the Hausman test typically suggests. In that case, η_{ij} (and all other observable time-invariant bilateral determinants of trade flows) can be eliminated either using a within transformation of the data or by first-differencing. If the number of time periods is two ($T=2$), the two methods yield identical estimates and inference: If $T>2$, the comparison depends on assumptions about ε_{ijt}. The fixed effects estimator is more efficient if ε_{ijt} is serially uncorrelated, while first-differencing is more efficient if ε_{ijt} follows a random walk (i.e., ε_{ijt}–ε_{ijt-1} is white noise). Baier and Bergstrand (2007) recommended first-differencing in the context of gravity equations. They argued that ε_{ijt} is likely to be serially correlated over time as unobserved determinants of trade flows are slow-moving over time. Also, trade flows as well as GDP levels are likely to be unit-root processes, which may lead to spurious regression in the fixed-effects model.

6.1.4 Aggregation

Anderson (2011) warned that estimation of (18.66) on aggregate data is problematic because of possible aggregation bias. The problem arises because of sectorally varying trade costs and sectorally varying elasticities of trade with respect to costs (Anderson and van Wincoop, 2004). This problem can be avoided by working with more disaggregate data. One can easily interpret (18.54) as the subutility index belonging to some specific sector (class of goods), and nest subutility indices, into, say, a Cobb–Douglas aggregator. Up to a constant multiplicative factor representing sectoral expenditure shares, bilateral trade flow equations for sub-aggregates (e.g., for groups of goods with different degrees of substitutability, indexed by a superscript s) will be formally similar to those derived from (18.54). Since sectoral output and expenditure data are not easy to come by for many levels of aggregation and countries, it is preferable to control for these variables by writing $\ln(Y_i^s Y_j^s) = \ln Y_i^s + \ln Y_j^s$ and let the country-sector dummies take care of them.

For example, Rauch and Trindade (2002) ran equation (18.66) on different dependent variables: First, they focused on trade in differentiated goods. These goods are highly heterogeneous with their characteristics typically depending on the producer who has monopoly on the production of a specific variant of the good. Most consumption goods such as apparel, appliances, or cars fall into this category. Second, there are goods for which either reference prices exist, or which are traded on organized exchanges. In both cases, the characteristics of the goods do not depend on the producer but are rather specific to the good itself. Standardized industrial inputs, or homogeneous products such as steel, wood, etc., fall into this category. Clearly, across those categories, the degree of product differentiation differs and so does the monopoly power of the producers. Hence, σ is probably low for differentiated goods, higher for reference-priced ones, and highest for exchange-priced ones. Also, the informational requirements for trade in differentiated goods are likely to be much higher than for homogeneous goods, so that ethnic networks should matter more for the former than for the latter. However, there are no clear predictions concerning the comparison between parameter estimates $\overline{\xi'}$ and $\{\overline{\nu}_T^{im}, \overline{\nu}_T^{em}\}$ obtained from these different regressions, where the bar refers to the coefficient multiplied by $(\sigma - 1)$, e.g., $\overline{\nu}_T^{im} = (\sigma - 1)\overline{\nu}_T^{im}$. For example, even if the trade cost and the preference channel could be separated, for a given strength of the network effect $\nu_T^k, k \in \{im, em\}$ the estimated coefficient $\overline{\nu}_T^k$ would be large for homogeneous goods since the degree of substitutability is high, and it would be low for differentiated goods. The opposite may be true if, for given σ, $\{\nu_T^{im}, \nu_T^{em}\}$ varies across the groups of goods. However, neither σ nor $\{\nu_T^{im}, \nu_T^{em}\}$ can be assumed constant over those sub-aggregates of goods so that the naive comparison of coefficients obtained from different regressions is problematic.

6.2 Empirical evidence: the effect of migration on trade

In the following we discuss the evidence reported in the literature. As a first step, we give a rough first browse over the main strands of thought in the literature. Then, we discuss how the key challenges have been dealt with in the literature. We focus on recent work that incorporates the lessons taught by Anderson and van Wincoop (2003). When possible, we contrast results from single-country multi-region models with many-country models. We start by discussing conditions under which consistent estimates of a model such as (18.66) is possible. Then we describe a number of factors that shape the form and size of the trade–migration nexus: product differentiation, immigrant education and occupation, intensive versus extensive margins of trade, foreign market characteristics, and indirect networks. What these different exercises have in common is their ambition to disentangle the transaction cost from the preferences channel of migration. However, so far no conclusive answer to this identification problem is provided in the numerous papers surveyed.

6.2.1 A quick browse over the existing body of research

The development of the empirical literature on the relationship between trade and migration has been driven by two phenomena: first, the emergence of more and better data, in particular on the distribution of migrants worldwide; second, improvements in the proper modeling of gravity models. These two issues are related: proper estimation of the multilateral resistance terms in the theory-founded gravity model requires employment of the full matrix of trade flows; this, in turn, also requires matching data on migrant stocks. While quality data on stocks of immigrants by country of origin have been and still are scarce, bilateral trade data for almost all country pairs in the world have existed since 1950.

Almost all papers in the literature focus on the effects of immigration, i.e., they set $\overline{\mu}_2$ to zero in (18.61). The first paper in the modern empirical tradition is Gould (1994). He studied the effect of immigration on trade between the US and 47 trading partners for the years of 1970–86. He estimated a gravity model that is surprisingly close to modern practice, but which relies on observed data to control for aggregate price indices rather than including an array of fixed effects. He did not exploit the panel nature of the data, treating it as repeated cross-sections, but did include the lagged value of trade flows on the right-handside. Using non-linear least squares methods, and ignoring the potential Nickel bias arising from the inclusion of a lagged endogenous variable on the right-handside of the regression equation, he distinguished between producer and consumer goods and between imports and exports. He found that immigrants increase trade, but the estimated effects are hard to compare to the subsequent literature that employed linear methods. However, on average they tend to be on the low side of later findings. Surprisingly, in his exercise, effects on exports are larger than on imports. This pattern would imply that preference effects are not important. Gould also found that the trade-enhancing effect of immigration is substantially larger for consumer goods as compared to producer goods. His interpretation is that consumer goods are more strongly differentiated than producer goods and therefore provide more.

Gould's seminal work has triggered a large literature. Many authors followed Gould by differentiating between differentiated and homogeneous goods and by distinguishing between imports and exports. Usually, papers study immigration into a single country, mostly into the US. They investigate American exports or imports from the immigrants' source countries, sometimes differentiating with respect to the goods covered (Dunlevy and Hutchinson, 1999; Mundra, 2005; Jansen and Piermartini, 2009). Using countrylevel data, there are also papers covering immigration into Canada (Head and Ries, 1998, Switzerland Tai, 2009, Germany Bruder, 2004, the UK Girma and Yu, 2002; Ghatak et al., 2009, Australia White and Tadesse, 2007 or New Zealand Bryant et al., 2004).

More recent studies also exploited the regional distribution of immigrants and look at the bilateral trade relationship between US regions (states) and foreign countries (Bardhan and Guhathakurta, 2004; Co et al., 2004; Dunlevy, 2006; Millimet and Osang, 2005; Bandyopadhyay et al., 2008; Tadesse and White, 2008). Herander and Saavedra (2005) analyzed the relative effects of state and nation-level migrant stocks for the US. Helliwell (1997) and Wagner et al. (2002) studied Canadian province-level trade flows.

There is also work on region-level trade and immigration for France (Combes et al., 2005; Briant et al., 2009, Italy Bratti et al., 2012, Spain Aleksynska and Peri, 2011, Sweden Hatzigeorgiou, 2010b, or Denmark Hiller, 2011).

Before the development of the Anderson and van Wincoop (2003) gravity equation (18.58), most papers used specifications that did not explicitly or implicitly (e.g., through the use of exporter and importer dummies) control for multilateral resistance terms (for example: Dunlevy and Hutchinson, 1999; Head and Ries, 1998; Girma and Yu, 2002). Since then it has become customary to include country dummies. Most papers reviewed above draw on pooled cross-sections and do not apply panel econometrics. With improved data, more and more studies exploit the time dimension (for example: Bandyopadhyay et al., 2008; Peri and Requena-Silvente, 2010). Generally, these modeling advances have narrowed the range of estimates that different authors found for different countries. A third, more recent but much smaller host of papers exploited more complete matrices of bilateral trade flows. That is, rather than studying trade of one immigrant destination country with respect to the rest of the world, these papers investigated bilateral trade and migration between many source and destination countries. Hatzigeorgiou (2010a) examined a cross-section of 75 countries in 2000, while Egger et al. (2012) worked with a set of 27 receiving OECD countries and 130 source countries. Felbermayr and Jung (2009) made use of a panel of countrypairs, covering North–South trade and migration links for the years 1990 and 2000. Parsons (2012) employed a large panel of bilateral trade flows and stocks of migrants covering the years 1960, 1970, 1980, 1990, and 2000. Equation (18.58) makes very clear that bilateral trade volumes depend on bilateral trade costs but also on multilateral resistance terms, which summarize trade frictions with all countries in the world. Consistent estimation of that gravity model, therefore, requires information on the whole set of trading partners for all countries. For these reasons, the increasing availability of full matrices of migration stocks is very welcome and will guide some of the following discussion in this section.

The smallest part of the literature deals with what one may call indirect trade effects of ethnic networks: the additional transactions between countries i and j driven by some ethnic group k that has residents in both places. The most famous paper in this area is Rauch and Trindade (2002), who studied the effect of the Chinese ethnic network on bilateral trade. They found that for trade in differentiated goods between countries with large ethnic Chinese populations (such as those in South-East Asia) the average trade increase attributable to ethnic Chinese networks is at least 60%. This is a large number that we will revisit below.

Wrapping up, the literature finds positive, statistically significant effects of migrant networks on trade. Figure 18.6 presents key results from the literature discussed above. It also plots the model export and import elasticities as reported in the meta-analysis of Genc et al. (2011) that is based on 48 studies containing about 300 estimates. The effect of immigrants on imports is typically estimated to be larger than the one of immigrants on

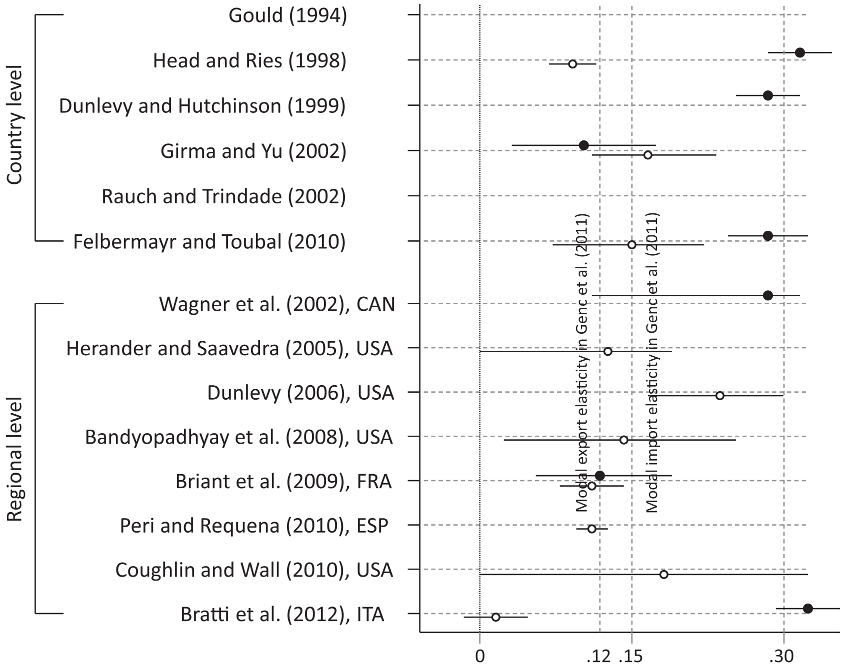

Figure 18.6 Summary of estimates in the literature. Note: Estimated elasticities of trade flows with respect to immigrant stocks in imports (black dots) or exports (white dots); 95% confidence intervals. *Figure adapted from Bratti et al. (2012).*

exports. Moreover, the trade–migration link appears stronger for goods whose trade is more likely to involve informational problems (differentiated goods) and for countries with weak institutions.

6.2.2 Dealing with endogeneity concerns

Probably the biggest single concern related to much of the papers mentioned above is that the network variable (the stock or share of migrants) may be correlated to trade shocks ϵ_{ij}: When this is the case, OLS estimation of (18.66) leads to biased and inconsistent results. Such endogeneity bias can arise from three sources. First, reverse causality: it is possible that some positive shock on the value of bilateral trade between two countries leads to more migration between the two places, for example, because the existence of some trade makes agents aware of bigger, hitherto hidden, trade potentials the realization of which makes migration for information arbitrage purposes worthwhile. Second, omitted variables: Hanson (2010) criticized the received literature on the migration–trade relationship by stating "*It is difficult to draw causal inference from these results, since immigration may be correlated with unobserved factors that also affect trade, such as the trading partners' cultural similarity or bilateral economic policies (e.g., preferential trade policies or investment treaties that raise* *the return to both migration and trade).*" In the same vein, Lucas (2006) argued "*reservations persist as to the potential for other, unobserved phenomena to be stimulating both trade and migration. . . . Overall the estimated effects seem improbably large . . .*". Third, measurement error: this is the least intellectually interesting but most likely highly relevant source of endogeneity bias.

The most convincing way to address the endogeneity concern is to look for some exogenous events that cause variation in bilateral migration stocks but have no direct effect on bilateral trade. Such natural experimental settings are rare in economics in general and in our area of interest in particular. To our knowledge, no study has yet proposed identification of the migration–trade nexus based on such an event. However, there do exist a few studies that propose instrumental variables. Also, there is a larger host of papers that exploit panel data to address the main concern voiced by Hanson (2010) or Lucas (2006) on omitted variables bias.

6.2.2.1 Controlling for unobserved heterogeneity

Applying the within-estimator or first-differencing the data, one can control for *all* time-invariant country-pair-specific determinants of both bilateral migration and bilateral trade. Next to the cultural or political determinants that could drive both migration and trade, country-pair effects also deal with initial conditions. This is important since it is well known that migrants tend to cluster where groups of their ethnicity already exist. It also deals with potential mismeasurement of true geographical barriers to mobility of goods or people by simple proxies such as great-circle distance between countries' economic centers of capital cities. Baier and Bergstrand (2007) documented very convincingly that country-pair-related heterogeneity can strongly distort estimates of gravity variables, such as that of free trade agreements.

There exist a couple of databases that report bilateral migrant stocks for a number of countries over time. Ozden et al. (2012) have presented the most comprehensive dataset so far, collecting data from national census, harmonizing it, and filling the gaps using alternative data sources and estimation. It comprises all countries in the world (226) and reports bilateral stocks of migrants based on the foreign-born concept for five completed census rounds, 1960, 1970, 1980, 1990, and 2000.

Parsons (2012) merged these data with trade data from Feenstra et al. (2005). The trade data however, span a smaller sample of countries than the migration data, so that the author ended up working with 178 countries at most. Moreover, the trade data do not contain the year 1960. The author used a model of the type described in equation (18.66) but augmented by country-pair fixed effects and exporter $\times$ year as well as importer $\times$ year dummies. However, his exercise led to disappointing results. In the presence of paireffects, the elasticity of immigrants on exports is -0.023, marginally significant with a standard error of 0.013. The elasticity of emigrants is -0.011, with a standard error of 0.012. So, it appears that migrant networks do not have an effect on trade, and if at all, that effect is negative. Note that Parsons did find positive network effects of plausible magnitudes in cross-sections for single years (with the sum of the emigrant and immigrant elasticities ranging between 0.10 and 0.13 across all years. The implication of his finding is that the effect of immigrant networks on trade obtained in his cross-sections is spurious: time-invariant, unobserved determinants of trade appear positively correlated with migration stocks, so that their omission wrongly attributes their trade-enhancing effect to migrants.

There are different conceivable reactions to this finding. First, methodological ones. Including pair-fixed effects along exporter and importer dummies interacted with year consumes a lot of degrees of freedom and asks very much from data that is probably severely affected by measurement error, so that inference is made very difficult. Also, given the structure of the data, it is not at all clear whether fixed effects estimation is the preferred strategy as compared to first-differencing. Second, the choice of trade data. The Feenstra et al. (2005) trade data have many advantages since they have been carefully cleaned. However, the coverage is substantially smaller than the IMF's Direction of Trade (DoT) dataset.

Table 18.2 shows regression results based on the Ozden et al. (2012) data merged with the DoT data. Paireffects are removed by first-differencing the data. All regressions include full sets of exporter and importer effects interacted with year dummies. Columns 1–5 present cross-sections for the years 1960, 1970, 1980, 1990, and 2000. The estimated elasticities for immigrants and emigrants alike are statistically significant at the 1% level (with a single exception in 1960) and of credible orders of magnitude. For example, in 2000, the estimated elasticity of 0.082 implies that doubling the number of immigrants in a country from some source country would increase imports of that country from that source by about 6% ($2^{0:082}=1.059$): Doubling the number of emigrants increases imports by about 3% ($2^{0.044}=1.031$): A one standard deviation increase in migration increases imports by 0.072 and 0.054 standard deviation respectively. These standardized beta coefficients can be compared to the one for distance: 0.306. The contribution of migration to the variation of trade is clearly smaller than that of distance but by no means trivial. Immigrants have a lower elasticity than emigrants; the difference of the estimated elasticities is different from zero at the 1% level. A similar finding was reported by Felbermayr and Toubal (2012) based on data for OECD countries for 2000 and using a slightly different specification (migrant shares rather than log level). This implies that migrants have a larger effect on imports than on exports. This is well in line with results documented in Figure 18.6.

Columns 6 and 7 in Table 18.2 report findings from the first-differenced panel model. Here, the elasticity of trade with respect to immigrants or emigrants is much smaller than in the cross-sectional exercise. The effect of emigrants on imports actually vanishes. However, a small positive and statistically significant effect of immigrants remains. Doubling the number of immigrants leads to an increase of imports by about 4%

Table 18.2 The effect of migration on trade: cross-sections versus panel estimates
Dependent variable: ln imports

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Census years	**1960**	**1970**	**1950**	**1990**	**2000**	**1970,1980,1990, 2000**	
			OLS			**FD panel**	
Ln immigrants	0.077***	0.112***	0.096***	0.093***	0.082***	0.059***	0.033*
	(0.015)	(0.014)	(0.015)	(0.012)	(0.011)	(0.017)	(0.016)
Ln emigrants	0.029***	0.042***	0.040***	0.045***	0.044***	−0.000	0.005
	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)
Ln GDP*i* GDP*j*	0.435***	0.547***	0.859***	0.600***	1.068***		
	(0.114)	(0.050)	(0.044)	(0.075)	(0.089)		
Free trade agreement (0,1)	0.294	0.427***	−0.007	−0.182*	0.395***	0.327***	0.318***
	(0.152)	(0.102)	(0.117)	(0.088)	(0.082)	(0.068)	(0.068)
Economic Integration (0,1)	−1.190***	−4.552***	−2.026***	−1.977***	−0.492***	0.710***	0 713***
	(0.208)	(0.355)	(0.253)	(0.233)	(0.112)	(0.089)	(0.086)
Currency union (0,1)	0.000	2.966***	0.581*	1.090***	−0.203	−0.193	−0.293**
	(0.000)	(0.330)	(0.236)	(0.212)	(0.118)	(0.110)	(0.109)
Ln distance	−0.511***	−0.701***	−0.930***	−1.085***	−1.263***		
	(0.054)	(0.053)	(0.056)	(0.050)	(0.046)		
Contiguity (0,1)	0.163	0.277	0.346*	0.465**	0.546***		
	(0.167)	(0.156)	(0.173)	(0.142)	(0.141)		
Common language (0,1)	0.556***	0.594***	0.607***	0.594***	0.574***		
	(0.097)	(0.090)	(0.088)	(0.083)	(0.075)		
Exporter x year dummies	YES	YES	YES	YES	YES	YES	
Importer x year dummies	YES	YES	YES	YES	YES	YES	
Country x year dummies							YES
R^2	0.719	0.716	0.732	0.715	0.741	0.298	0.243
N	1719	3367	3798	5687	7077	9777	9777

Note: Standard errors (in brackets) are clustered at the observation level (country pairs). *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. Own calculations.

($2^{0.059} = 1.042$). This is still no trivial effect. The differences relative to Parsons (2012) have two explanations: first, a larger dataset is used, and, second, rather than using a within-transformation, the data is first-differenced.

Due to a better data situation, panel techniques have been more prevalent in single-country multi-region studies than in cross-country ones. Bandyopadhyay et al. (2008) constructed a two-year panel of US state exports to show that, when the estimates control for state-country fixed effects, the estimate of the ethnic network falls by nearly half (from 0.27 to 0.14). Similarly, in Coughlin and Wall (2011) the panel estimate of the immigrant elasticity is only about 57% as big as the one obtained in a pooled cross-section (0.192 versus 0.335) and is only marginally significant statistically.

While the extensive use of dummy variables and first-differencing (or, equivalently, within-transforming) the data does help with omitted variable bias, there are still concerns about reverse causality. However, Wooldridge (2002) recommended a regression-based *F*-test for strict exogeneity; see also Baier and Bergstrand (2007), who discuss the suitability of this test in the gravity framework. Felbermayr and Jung (2009) applied the test in a two-period model of North–South trade and migration and found that it is not possible to reject strict exogeneity of migration, conditional, of course, on first-differencing the model and including the standard gravity controls also present in columns 6 and 7 of Table 18.2.

6.2.2.2 Instrumental variables strategies

Recently, some authors have used instrumental variables techniques in gravity models of the type given by (18.66). The most prominent example is Peri and Requena-Silvente (2010). This study uses data from a single host country, Spain. However, it has a panel dimension covering regional exports in Spain for the years 1995–2008. In that period, both trade and immigration increased substantially, the latter by an annual growth rate of 17%. Both across Spanish host regions and across source countries immigration displays strong variation. This makes Spain in that period an interesting case to study. The authors regressed the log of exports from some Spanish province to some foreign country at some point in time on the log stock of immigrants from that of foreign countries residing in that Spanish province in that year. They include country × year dummies and country × province dummies to account for characteristics of foreign markets and their evolution over time as well as for all time-invariant determinants of a province' s trade with some foreign country (such as historical migration patterns, historical ties, distance, etc.). These dummies absorb simple year, province, and country effects. Identification in the model is then based on variation of immigrant stocks within province–country pairs across time. The dataset is very large since it spans 50 provinces, 77 countries and 13 years. Peri and Requena-Silvente (2010) have filled zero trade flows by adding one to each export flow.

The authors instrumented the changes in immigrants in a particular province by imputing net inflows of immigrants based on historical data. More precisely, they used the distribution of immigrants by nationality and across provinces from 1993, i.e., before the strong rise in immigration to Spain, and attributed to each group in each province the net growth of immigrants from that nationality to Spain. This instrument has time variation because the overall immigration inflow varies over time, and it has cross-sectional variation because of the uneven distribution in 1993. If immigrants tend to move to regions where other individuals of their nationality already settle, the imputed inflow will correlate with the actual one. And since the instrument is based on the cross-province distribution of immigrants as of 1993, it will not be affected by trade shocks in the period considered.

In the first-stage regression, the instrument (imputed immigration) has a strong positive effect on actual immigration. Its coefficient is 0.55 with a t-value of 185; the R^2 of the regression is 0.85. The F-test produces a statistic of more than 300, which lets the instrument appear as very strong. In their preferred specification, the elasticity of exports with respect to the stock of immigrants is 0.11. Using the instrument described above, the elasticity drops to 0.05. That instrumentation reduces the estimate is exactly what one would expect, since the correlation of trade shocks with migration stocks is suspected to be positive, thus causing OLS to overestimate. The estimate is statistically significant at the 5% level. This elasticity is within the range of usual findings as summarized in Figure 18.6 but certainly on the low side.

The instrumental variables strategy has been used in subsequent research, for example by Bratti et al. (2012), who used province-level data for Italy, covering the period 2002–2009 during which Italy experienced a drastic increase in immigration. Their OLS results suggest that the elasticity of exports and imports with respect to immigration stocks was 0.058 and 0.344 respectively. Estimated at high precision, these elasticities are at the extreme ends of previous findings, (see Figure 18.6). The first stage of their two-stage least squares exercise again shows that the imputed evolution of provincial immigration stocks by country of origin correlates very well with the actual one; the reported F-statistic is a staggering 3871. The second stage of their IV exercise results in the effect of immigrants on exports becoming statistically indistinguishable from zero, while the effect of immigrants on imports increases by 60% to 0.548. The fact that instrumentation increases the estimates is counter-intuitive. However, measurement error in migrant stocks could well explain a downward bias of OLS estimates.

6.2.3 The role of product differentiation

Generally, in the gravity equations of the trade–migration nexus based on (18.58), the estimate of some trade cost-related variable—such as the log stock of immigrants in a country—reflects two elements: the elasticity of substitution across varieties and the effect of the immigrant network on iceberg trade costs. Moreover, the stronger the degree of product differentiation within a sector or product class, i.e., the lower the elasticity of substitution, the larger one would conjecture the trade cost effect of the network to be since informational needs are higher and potential informational asymmetries are stronger. Also, specific preferences for goods from their source countries can be assumed to be stronger when goods are more differentiable. Combining these considerations with our discussion of the gravity model (18.66) above, and writing square brackets to denote a functional dependence, the empirical estimate of the network $\overline{\nu}^i := (\sigma - 1)\left(\nu_a + \nu_T^{im}\right)$. In light of the above discussion, it is very reasonable to assume that ν_T^{im} depends negatively on σ. Hence, the estimate of $\overline{\nu}^i$ can be expected to be non-monotonic in σ with intermediate levels of differentiation leading to the largest estimated coefficient.

Nonetheless, the literature frequently disaggregates the trade data according to the degree of differentiation. This practice has started with the seminal work of Gould (1994), though he distinguished between consumer and producer goods. Also, Rauch and Trindade (2002) have prominently made the distinction operative by classifying products into three categories of goods ranked with declining degree of differentiation: goods traded on public exchanges (such as the London metal exchange or the Chicago board of trade) are homogeneous goods and the associated σ is high, goods for which reference prices exist are more easily differentiable and σ is of intermediate level, and the remaining goods also known as differentiated goods with low levels of σ. Peri and Requena-Silvente (2010) classified goods in a way directly related to estimates of σ taken from Broda and Weinstein (2006).

6.2.3.1 Cross-country evidence

Table 18.3 summarizes estimates provided by Felbermayr and Toubal (2012) for a sample of 29 OECD countries and the year 2000. Log imports are regressed on the shares of bilateral immigrant stocks in the destination country and the shares of bilateral emigrant stocks in the source country. Geographical and cultural proximity is controlled for by including variables such as the log of distance, a contiguity dummy, and a dummy for common language. Trade policy is controlled for by including information on bilateral trade agreements. Multilateral resistance terms and other country-level determinants of bilateral imports are taken care of by exporter and importer dummies. Column 1 reports the results of such a standard gravity model on total imports. Estimated coefficients on the non-migration-related variables have the right signs, and are, mostly, close to the usual magnitudes. For example, the elasticity of geographical distance is −0.9. EU membership increases bilateral trade by about 23% percent. More interestingly, however, both immigrants in country i from j and emigrants from country i in j increase $i's$ imports from j. The effect of the immigrant share is measured to be 0.278 and highly significant. Since the average share in the data is 0.181%, the average immigrant network increases imports by about 5.0% relative to a situation without such a network. The effect of the emigrant share is estimated at 0.154, which is equivalent to a trade-creating effect of about 2.8%. The estimates in the table are not immediately comparable with those reported in Figure 18.6 since those refer to log levels of migrants obtained in models with only

Table 18.3 The effect of emigrants and immigrants on imports across different classes of goods
Dependent variable: Ln imports

	(1)	(2)	(3)	(4)	(5)	(6)
	All migrants			High-skilled migrants		
	Total	Homog.	Diff.	Total	Homog.	Diff.
Share of migrants from South in North						
Immigrants	0.278***	0.319***	0.305***	0.694***	0.795**	0.759***
	(0.062)	(0.100)	(0.064)	(0.140)	(0.340)	(0.160)
Emigrants	0.154**	0.218**	0.184**	0.402*	0.554*	0.544**
	(0.075)	(0.110)	(0.072)	(0.210)	(0.310)	(0.210)
Geographical and cultural proximity						
Ln geographical distance	−0.891***	−1.504***	−0.852***	−0.881***	−1.492***	−0.841***
	(0.060)	(0.068)	(0.070)	(0.062)	(0.110)	0.072)
Contiguity	0.160	0.305**	0.007	0.271**	0.779***	0.130
	(0.120)	(0.140)	(0.140)	(0.120)	(0.200)	(0.130)
Common language	0.168	0.134	0.173	0.198	0.244	0.198*
	(0.120)	(0.140)	(0.110)	(0.120)	(0.210)	(0.110)
Trade policy						
Both countries in EU	0.246*	0.414***	0.243*	0.230*	0.213	0.255*
	(0.140)	(0.160)	(0.130)	(0.140)	(0.270)	(0.130)
Both countries in NAFTA	1.014***	0.217	1.512***	0.171***	−0.626	1.656***
	(0.400)	(0.320)	(0.450)	(0.450)	(0.730)	(0.510)
Accession treaties	0 227	0.023	0.178	0.204	−0.464	0.153
	(0.180)	(0.220)	(0.150)	(0.180)	(0.370)	(0.160)
R^2	0.915	0.880	0.920	0.913	0.800	0.920

Note: OECD countries only, year 2000. Balanced sample of 536 dyads. Robust standard errors in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively. All regressions include exporter and importer effects, and a constant. Synthesis of results documented in Felbermayr and Toubal (2012).

one migrant variable on the right-hand side. The standardized beta coefficients to the estimates (0.278 for immigrants and 0.154 for emigrants) are 0.08 and 0.05 respectively.[95] Not surprisingly, compared to the effect of distance, which has a beta coefficient of −0.57, the importance of migration is relatively small.

The fact that immigrants matter more for imports than emigrants may be informative about the role of the information relative to the preferences channel. Immigrants as well as emigrants may help overcome lack of information or informational asymmetries. However, different from emigrants, immigrants may have a special preference for goods from their home countries, so that the difference between the estimated coefficients may be the first evidence for the existence of a preference channel.

Turning to imports of homogeneous (exchange traded) goods and differentiated goods (according to the classification of Rauch, 1999) the authors found that immigrants and emigrants still matter, and the obtained coefficients compare in a fairly similar fashion. However, to back out the trade cost savings expressed as tariff equivalents implied by these network effects, one needs to divide the estimates by $\sigma - 1$. Taking σ from the survey by Anderson and van Wincoop (2004) to be equal to 5 for differentiated goods but 10–20 for homogeneous goods, the estimates shown in Table 18.3 imply trade costs savings that are at least twice as large for differentiated goods as for homogeneous goods (in *ad valorem* tariff equivalents).

6.2.3.2 Region-level evidence

Peri and Requena-Silvente (2010) ran separate regressions for trade flow aggregates consisting of highly differentiated, medium differentiated and low differentiated products. As in the exercise on cross-country OECD data described above, point estimates of log immigrant stocks do not differ much across these goods classes. They are almost identical for the high and the low differentiated goods (0.097 and 0.098 respectively) and highest for medium goods (0.122). However, the trade costs savings implied by these estimates again differ by about a factor of 2.5 across the high and the low differentiated goods.[96]

The results of the two selected studies therefore confirm earlier findings of Rauch and Trindade (2002): migrant networks are more important for goods featuring low degrees of differentiation.

[95] Beta coefficients are defined as the estimated coefficient times the standard deviation of its corresponding independent variable divided by the standard deviation of the dependent variable, which transform the estimated coefficients into units of sample standard deviation. This allows one to compare the power of covariates in explaining the dependent variable. For example, a beta coefficient of $\hat{\beta}$ on some independent variable x would signal that a one-standard-deviation increase in x results in a $\hat{\beta}$ -standard-deviations increase in the independent variable.

[96] Details depend on the assumed value of the elasticity of substitution. Peri and Requena-Silvente (2010) worked with fairly low values of σ (products with σ estimated by Broda and Weinstein (2006) to be higher than 3.5 are already classified as low differentiated). However, the σ– 1 correction factor implied by these numbers is at least 2.5 higher for high as compared to low differentiated goods.

6.2.4 The roles of immigrant education and occupation

Much of the literature takes the strength of a network created by migrants as proportional to the number of individuals involved. It does not account for the heterogeneity within the migrant population with respect to educational achievement or occupations. Both are likely to matter, though. If immigrants are indeed instrumental in overcoming informational frictions between their host and source countries they must at least be able to secure themselves jobs in their countries of residence that allow them to capitalize their knowledge. Many unskilled immigrants occupy jobs that do not provide them with this option, such as in the construction, retailing, or cleaning sectors. Better education is likely to help them leverage their informational advantages so that they can become effective in facilitating international trade. For example, one would not expect significant reductions of trade costs from migrants if migrants do not master the language of their host country. Better education correlates with integration into the host country's society in general and into its labor markets in particular.

If more skilled and more able individuals are more likely to select into emigration and into export-related occupations, then the observed correlation between migration and trade may be a byproduct of self-selection of migrants. Relaxing policy barriers to immigration with the idea of increasing trade flows may then only be successful if the policies target specific types of migrants: those with the right skills.

More educated migrants could, in principle, also strengthen the preferences channel simply because they command higher incomes. There is very little work so far that differentiates between skill or occupation; in particular, to our knowledge a region/level analysis based on microdata is still missing. More work in this area is very much welcome.

6.2.4.1 Immigrant education

Columns 4–6 in Table 18.3 present estimates from Felbermayr and Toubal (2012) that isolate the role of tertiary-educated migrants. Compared to the estimate in column 1 the estimated effect of migration comes out with doubled magnitude in column 4. Hence, a one-point increase in the share of high-skilled migrants has twice as strong a trade-creating effect as an equivalent increase in the total share of migrants (with, supposedly, an even stronger difference when compared to unskilled migrants.) However, mostly because the share of high-skilled migrants in the population of the destination country is only about a quarter of the total share on average, the beta coefficients are 0.06 for immigrants and 0.04 for emigrants, which is similar in size to the effect found for total migration.

Since the effects of immigrants and emigrants on imports are statistically identical both when the sum of migrants is considered or only the highly educated ones, one can restrict the two effects to be similar and re-estimate the models. This yields common estimates of 0.185 and 0.612 respectively, both statistically significant at the 1% level. With an elasticity of substitution of $\sigma = 6$ for total bilateral trade, the *ad valorem* tariff equivalent of increasing the share of migrants by 1 percentage point is 3.7% (0.185/(6 − 1)). Considering high-skilled migrants only, one obtains a tariff equivalent of about 12.2% (0.612/(6 − 1)). Columns 5 and 6 look at the groups of homogeneous and differentiated goods and confirm what we have seen before: point estimates across product classes are very comparable. But since the underlying elasticities of substitution differ, the trade-creating effect of high-skilled migrants is again at least twice as strong for differentiated goods than for homogeneous goods.

The strong trade-creating effect of high-skilled migration confirms within our broad cross-country OECD sample earlier results of Herander and Saavedra (2005). The trade-promoting effects of migrant networks is the larger, the better the ability of that group to receive and process information on trading opportunities.

6.2.4.2 Immigrant occupation

Aleksynska and Peri (2011) provided the first study that differentiates immigrants according to their occupations in the host country. They used a new dataset provided by the OECD that covers 89 destination countries and up to 233 countries of origin. The data refer to the years 2000–02 and provide information on immigrants' age, gender, schooling level, labor market status and occupation at the one-digit ISCO classification. These variables are merged with standard trade data and gravity covariates.

Their gravity equation follows (18.66). It contains the log of the total immigrant stock from country j in country i, but also includes the log share of immigrants in a specific occupation group with special affinity to trade facilitation. They focus on ISCO group 1, which includes senior government officials, officials in special interest organizations, and managers of enterprises as well as on groups 5 and 9, which contain sales persons of all levels and demonstrators.

In regressions containing both the log total level of immigration and the log share of immigrants in business occupations, Aleksynska and Peri (2011) documented the following results. The elasticity of imports with respect to the total immigrant stock is 0.27. This is in the range of estimates shown in Figure 18.2. On top of this, the coefficient on the share of migrants in business activities is also positive and significant at the 5% level. This implies that immigrant business people have an effect above that of the total migrants. Everything else being equal, an increase in the share of immigrants employed in the business network occupations by 1% increases imports by approximately 0.6%. Including the share of migrants in ISCO groups 5 and 9 does not yield significant results.

The authors also included the shares of immigrants with primary, secondary and tertiary education into their regression along with the share of immigrants in business network occupations. They found a significant trade-enhancing effect of the highest education group, but no effect of other skill classes. They continued to report a strong effect of the business network shares. Differentiating between educational categories within their occupation classes, they argued that both high education of immigrants per se and their occupation in business activities contribute to their trade-creating effect.

6.2.5 Extensive versus intensive margins

Recent work on the theoretical underpinnings of the gravity equation when firms are heterogeneous in terms of their productivity (Melitz, 2003) stresses the different roles of fixed market access costs versus variable (iceberg) trade costs in explaining global trade patterns. In the presence of such fixed costs, only the more productive firms may find it worthwhile to export to foreign markets. The paper by Chaney (2008) derived a gravity equation for the Melitz (2003) model where firms are assumed to sample their time-invariant productivities from a Pareto distribution. He showed that a reduction in fixed bilateral costs of trade (market access costs) should not have any impact on the *intensive margin* of trade, that is, on exports of a given variety (produced by a specific firm). Instead, it would increase total exports by allowing more varieties to be exported (i.e., more firms to become exporters) and, thus, have an effect on the *extensive* margin. A reduction in variable trade costs, in contrast, affects both the intensive and the extensive margins. This suggests that an analysis of the margins of trade creation through immigrant networks can shed light on the question whether networks reduce fixed or variable trade costs.

When fixed market access costs are too high even for the most productive producer to engage into exporting, the model of Melitz (2003) can also help understand the occurrence of country pairs that do not trade at all. Such zero-trade flows are of substantial importance in the data (see Felbermayr and Kohler, 2006a). This idea has been exploited by Helpman et al. (2008), who proposed a framework to estimate the gravity model in the presence of extensive and intensive margins of trade and that can be used on aggregate data rather than on firm-level transactions data.

6.2.5.1 Evidence based on aggregate region-level trade flows

Coughlin and Wall (2011) used data on manufacturing exports of 48 US states (Hawaii and Alaska are excluded) to 29 countries in 19 SIC industries for the years 1990 and 2000. Data on the stocks of immigrants are from the decennial census. Focusing on country–industry combinations for which exports were positive in at least one of the two years, the authors made 47,776 observations; 7,296 of these involved zero export flows. In this context, the extensive margin refers to aggregate data, that is, to a US state switching from zero exports to some country in some industry to positive exports, or the other way round. The intensive margin refers to changes in bilateral trade volumes conditional on them being positive.

Since the authors used panel data, they include country-pair effects into their regressions to control for unobserved sources of heterogeneity; see below for more discussion. They ran two types of regression: a fixed-effects logit model to estimate the extensive margin of ethnic networks and a conventional panel gravity model (where the least square dummy variables model is equivalent to a first-differenced model). Note that the fixed-effects logit model identifies the network effect of immigrants by relying on sector-country exports switching from zero to something positive or vice versa over time. The authors also ran a log-linear fixed effects model combining the two margins. For the log of exports to be defined for zero-trade flows, they added one to each export value in the data. They reported the following result: across both margins, the elasticity of total sectoral exports with respect to the immigrant stock is 0.192. Conditioning on positive exports, they found an elasticity of 0.139 for the intensive margin. In contrast, the probability model does not reveal any effect of the migrant network on the extensive margin of trade. That is, the presence of migrants from some country in some US state does not make it more likely to observe positive trade between that state and that country in some industry. This is a surprising result, since it is at odds with the idea that migrants lower the costs of information about foreign markets. In the model of Melitz (2003), such costs would take the form of fixed market access costs rather than of variable costs.

Parsons and Vézina (2013) exploited a natural experiment: the large inflow of refugees from Vietnam during and following the fall of Saigon in 1975 into the US. The US accepted those migrants for humanitarian reasons and distributed them across the states to share the burden. The US maintained a trade embargo with communist Vietnam from 1975 to 1994, so that it is very unlikely that trade-related reasons have influenced the distribution of refugees. Yet, for the period 1995–2010 (i.e., after the end of the embargo), the authors showed that the presence of those refugees has a strong and positive effect on state-level exports to Vietnam. Since the initial distribution of refugees is orthogonal to trade or even to the expectation of trade, it is safe to assume that the authors have established a truly causal effect.

6.2.5.2 Evidence based on firm-level transactions

One problem with the Coughlin and Wall (2011) study is that it draws on aggregate data and, therefore, has to interpret the extensive margin at the aggregate level. In contrast, the study of Peri and Requena-Silvente (2010), which we have already discussed in parts above, is based on individual transactions at the firm level. This allows for a more disaggregate view of the extensive margin that is defined as the number of transactions. The intensive margin, in turn, is defined as the average value of one transaction. Total exports of some Spanish province to some foreign country at some point in time can then be written as the product between the intensive and the extensive margins. When estimating the margins separately in log form using a gravity model similar to (18.66), the coefficients on the extensive and intensive margins equations have to add up to those obtained from a regression using log total trade on the left-hand side.

The authors found that immigrants affect exports much more strongly and consistently through the extensive margin than through the intensive margin. In their instrumental variables regressions (see below), migrants turn out to matter exclusively for the extensive margin. This is true across highly, medium, or low differentiated goods, but the strongest role of migrants at the extensive margin exists with highly differentiated goods. In light of the Melitz (2003) model, this implies that immigration to Spain reduces fixed bilateral trade costs rather than variable costs.[97]

To our knowledge, there does not yet exist a study that distinguishes between the extensive and intensive margins at the product level. The data for such an exercise exist, and it would be interesting to see whether more comprehensive data covering many source countries can resolve the contradiction in the findings of the papers discussed above.

6.2.6 The role of trade partner characteristics

A straightforward way to distinguish between the incomplete and the asymmetric information channels of the pro-trade effects of trade is to interact the network variable with trading country characteristics. Dunlevy (2006) used data on manufactures' exports at the level of US states to test the influence of the foreign-born on the bilateral exports of their states of residence to their countries of origin. He estimated Tobit models augmented by state and country fixed effects on a cross-section of around the year 1990.

Dunlevy (2006) found an overall export-enhancing effect of the immigrant network that is statistically significant and equivalent to an elasticity of 0.24, fitting well into the usual range of estimates. However, his contribution is to include interactions between the log of immigrants and source country characteristics, such as an index of corruption from the International Country Risk Guide, an index of institutional similarity that takes value 1 if the source country is Ireland, Canada, Australia, New Zealand or the UK and 0 otherwise, and finally two language dummies that record whether a country is English speaking or Spanish speaking. Clearly, in the presence of country dummies, the direct effects of these variables cannot be identified. However, he found strong evidence that higher corruption in the source country increases the elasticity of exports with respect to immigration. Moreover, the trade-creating effect is much more pronounced when the export partner is not an English- or a Spanish-speaking country. Institutional similarity has no measurable effect on the pro-trade effect.

When the language of the trade partner is one of the most common languages of the US, English or Spanish, the trade boost from the migrant network is reduced. This may reflect the fact that information is more easily and more cheaply available about market opportunities in these countries. The negative and significant interaction terms therefore suggest that networks do indeed affect trade by providing information. Corruption, in contrast, is more related to the lack of enforceability of contracts in the trading partner country, or to the general pervasiveness of opportunistic behavior. The positive and significant interaction term then informs about the relevance of the migrant network in

[97] In his analysis of Swiss data, Vézina (2012) also found that the extensive margin dominates.

mitigating asymmetric information and enabling transactions that would not have taken place due to lack of trust between two ethnically unrelated individuals.

6.2.7 Indirect network effects

6.2.7.1 The Chinese network

The most prominent evidence for market-replacing networks comes from the study by Rauch and Trindade (2002), who investigated how ethnic Chinese minorities residing in different countries promote trade between these countries. They characterized the three channels through which migrants can affect trade: resolution of incomplete information, mitigation of asymmetric information problems, and demand effects. In their study, they aimed at identifying the incomplete information channel by differentiating reference priced and non-reference priced goods. For the latter, informal information networks must be more important than for the former, where information on prices is easily available and quality is standardized. This is a neat argument; however, there are at least three caveats. First, to the extent that the presence of large ethnic Chinese populations in two countries also make their demand structures more similar, there could be Linder (1961)-type home market effects that result in more trade. Second, if differentiated goods have characteristics that are more difficult to include in contracts and that are more costly to enforce, networks that help to overcome asymmetric information may also promote trade more strongly for differentiated goods. Third, in gravity equations the trade cost effect of a network is typically confounded by the elasticity of substitution so that proper identification is hard; see Section 6.2.3 above.

The defining feature of the Rauch and Trindade (2002) paper is that they did not only consider the effect of the ethnic Chinese diaspora on bilateral trade between countries with Chinese minorities and homeland China, but also, and foremost, trade between country pairs not involving China itself. They exploited the fact that the ethnic Chinese network also spans countries different from China. They considered two cross-sections, one for the year 1980 and one for the year 1990, covering 57 countries in 1980 and 59 in 1990. Their specification differs from (18.66) in several respects: first, they estimated a threshold Tobit model to account for zero trade flows. Second, they used the log sum of imports and exports as the dependent variable. Third, they did not control for multilateral resistance terms and they did not include country dummies in their model. Their key independent variable is CHINSHARE, the product of the ethnic Chinese population shares for the two countries forming a pair. Their regressions drew on country pairs that have direct links to mainland China (where CHINSHARE is almost unity) and country pairs that have only indirect links to China (since they do not involve China itself).

They found that CHINSHARE has a strong positive effect on bilateral trade in both 1980 and 1990 and across goods classes consisting of goods traded on organized exchanges, goods for which reference prices exist, and goods characterized as differentiated. The point estimates reflect the expected ordering with respect to size: the coefficient for differentiated goods is typically about twice that for goods traded on exchanges. Next, the authors differentiated between strong ethnic networks (linking countries in both of which the Chinese ethnic minority makes up at least 1% of the population) and weak ones (the remainder). It turns out that the trade-creating effect of the Chinese network is an order of magnitude larger in the case of strong networks as compared to weak ones. More precisely, for differentiated goods, the percentage increase in bilateral trade attributable to ethnic Chinese networks is at least 60% in all models considered. Anderson and van Wincoop (2004) showed that this strong trade creation effect translates into an *ad valorem* tariff reduction equivalent of about 6%.

6.2.7.2 Indirect network effects

The empirical strategy of Rauch and Trindade (2002) reflects the methodological state of the art around the year 2000. Since then, a new consensus on how to estimate gravity models has emerged; see our discussion in Section 6.1. First, the gravity model suggests that the dependent variable should be the log of either imports or exports and not their sum. Second, the model should include exporter and importer dummies. Felbermayr et al. (2010b) revisited the evidence using such a revamped gravity framework. Rather than focusing on strong versus weak networks, they made a clearer distinction between direct and indirect links.

Table 18.4 shows regression results for aggregate trade (results differentiating across product groups look similar). Column 1 presents an OLS regression of log imports on the CHINSHARE variable. The point estimate obtained is very similar to the one found by Rauch and Trindade (2002), both for 1980 and 1990, who never ran regressions of aggregate trade, though. The estimate amounted to total trade creation of about 1.4% on average, assuming that CHINSHARE moves from zero to the sample average. However, when including exporter and importer in column 2 the point estimate falls by a factor of 5, with an associated average trade creation effect of merely 0.28%. The associated tariff equivalent is a mere 0.04%. Column 3 decomposes the total network effect into direct (involving mainland China) and indirect links (not involving China as a trade partner). The dummy variable DIR takes value 1 if the bilateral relationship involves China and 0 otherwise. Evaluated at the respective sample means, the effect of direct links is associated with substantial trade creation (4.7%), while trade creation due to indirect links is very minor (less than 0.1%). Columns 4–6 to (6) repeat this exercise for the year 1990, finding very similar results. This implies that the total CHINSHARE effect found by Rauch and Trindade (2002) is strongly dominated by direct network links. The indirect links, which are more likely to reflect informational issues rather than preferences, are much less important.

Estimates in Table 18.4 cannot be directly compared to Rauch and Trindade (2002), since they did not further differentiate between strong and weak networks and between

Table 18.4 The direct and indirect trade effects of the Chinese migration network
Dependent variable: Ln imports

	1980			1990		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
CHINSHARE	4.488***	0.893***		4.471***	1.262***	
	(0.941)	(0.339)		(0.642)	(0.478)	
CHINSHARE*(1-DIR)			1.137***			0.979***
			(0.434)			(0.371)
CHINSHARE*DIR			0.769*			1.402**
			(0.407)			(0.634)
Ln distance	−1.084***	−1.117***	−1.116***	−0.949***	−0.984***	−0.986***
	(0.068)	(0.061)	(0.061)	(0.047)	(0.042)	(0.042)
Contiguity	−0.0685	0.130	0.133	0.226	0.452**	0.449**
	(0.228)	(0.213)	(0.214)	(0.184)	(0.192)	(0.192)
Common language	0.557***	0.513***	0.514***	0.645***	0.549***	0.549***
	(0.124)	(0.124)	(0.124)	(0.091)	(0.092)	(0.093)
EEC	−0.226	−1.569***	−1.568***	0.154	−0.353**	−0.354**
	(0.167)	(0.175)	(0.175)	(0.115)	(0.148)	(0.149)
EFTA	0.656***	0.012	0.012	0.288**	−0.021	−0.021
	(0.168)	(0.193)	(0.170)	(0.141)	(0.160)	(0.161)
NAFTA						
MERCOSUR						
ASEAN						
Remoteness Index	YES			YES		
Exporter/importer effects		YES	YES		YES	YES
R^2	0.609	0.722	0.722	0.702	0.794	0.794
Tariff equivalent %						
CHIN	0.201	0.04		0.183	0.0518	
CHIN*1-DIR			0.0244			0.0195
CHIN*DIR			0.518			0.809

Note: N=2520 in 1980, N=2795 in 1990, and N=3259 in 2000. All regressions include a constant, the log product of both countries' GDPs, and a colony dummy. Observations clustered by unidirectional country pair. Robust standard errors in parenthesis. ***, **, * indicate significance at the 1%, 5%, and 10% level respectively. *Ad valorem* tariff equivalents % evaluated at the respective sample means. Elasticity of substitution is 8. Synthesis of results documented in Felbermayr et al. (2010).

different commodity groups. Felbermayr et al. (2010b) provided estimates that incorporate these features. They found that tariff equivalents are larger for differentiated than for homogeneous goods. The tariff equivalent of a strong network with differentiated goods is between 2% and 4% for the year of 1990. This is considerably lower than the 6% found by Rauch and Trindade (2002), but still considerable. The tariff equivalents are even more impressive when focusing on strong direct links. Here, in 1990 and for differentiated goods, they amount to between 5% and 11%. At the same time, even for strong networks, indirect links are of much lesser relevance. In 1990, they do not exceed 1% (again, for differentiated goods).

These findings suggest that indirect network effects of the Chinese diaspora are present in the data, but that they are considerably weaker than direct effects. When looking at other ethnic networks, and not differentiating between indirect and direct or between strong or weak networks, Felbermayr et al. (2010b) found evidence for several other co-ethnic networks. Average tariff equivalents as measures of the value of those networks, however, never exceed 0.1%. Interestingly, the highest value is associated with the Moroccan network, followed by the Ghanaian and Danish ones.

7. CONCLUSION

In this chapter we have set out to explore the interaction of migration, trade and capital movement as well as capital formation. Our goal was to shed light on the effects of this interaction on welfare as well as within-country income inequality and convergence between countries, and to identify possible patterns of causality between these three forms of economic globalization. Towards this end, we have taken four different perspectives. First, we have pursued a historical comparison between the nineteenth century era of "mass migration" and the recent decades of modern globalization. Second, we have taken a theoretical perspective through the lens of the factor proportions approach, focusing on the determinants and effects of migration between two economies that are already connected through free trade. Our third perspective, based on growth theory, has focused on the role that migration plays in the process of physical capital accumulation as well as for human capital formation. And finally, we have used the gravity approach to explore the relationship between migration and trade, emphasizing the role of migrant networks for trade costs.

Research by economic historians tells us that, some anomalies notwithstanding, the simultaneous surge of migration, trade and capital flows characteristic of nineteenth century globalization may be reconstructed along the lines of the factor proportions theory of trade and factor flows, provided that we take into account the technology factor. Contrary to widespread belief, globalization was not brought to a halt abruptly as a consequence of the Great War, but had started to be undermined much earlier by restrictive policies chosen in response to unwelcome trends in domestic income distribution, which may be understood along the lines of the factor proportions approach. This should act as a warning against "cold shoulder attitudes" vis-à-vis distributional trends in the present wave of globalization.

Interestingly, in the nineteenth century protective trade policies had been enacted long before restrictions on immigration. The same cannot be said for the more recent wave of economic globalization, where for decades now the rich part of the world has been running restrictive immigration policies, while at the same time pursuing liberalization policies with respect to trade and capital movements. Moreover, during these decades, international income convergence has been much more moderate than in the nineteenth century. Indeed, five decades of post-World War II economic globalization have not been able to reduce wage gaps between the richest and poorest countries of the world to anywhere near the gaps that had existed between the two sides of the "Atlantic economy" at the beginning of mass migration.

We have turned to a theoretical perspective in order to highlight the channels through which migration potentially affects domestic wages and international convergence of wages. Adopting a factor proportions approach, and allowing for migration to take place simultaneously for several types of labor, we were able to identify potential theoretical explanations for the inconclusive empirical evidence. A first conclusion to be drawn from this approach is that the factor price and welfare effects of migration very much depend on whether or not the sending and receiving countries of migration are connected by trade. Broadly speaking, the factor price effects are less pronounced in the empirically relevant case where migration takes place between countries that are open to trade on goods markets. At the same time, however, positive welfare effects for natives of the immigration country, the well-known immigration surplus, will arise only if immigration does affect factor prices and, thus, domestic income distribution. Therefore, immigration should not be expected to be beneficial for all individuals. Arguably, a broad consensus in favor of migration requires that policy explicitly addresses possible ways of compensation.

Any migration scenario relevant for the status quo must take into account preexisting stocks of migrants in the receiving country. We have developed a factor proportions model that allows us to do so when deriving welfare effects from migration. This departs from most of the existing literature, and it generates an important insight: On account of such migrant stocks, any change in wage rates constitutes opposite welfare effects for the sending and the receiving country, much like a terms of trade effect on goods markets. As with all terms of trade effects, this effect works in opposite directions in the two countries. This, together with the fact that one country's immigration surplus is the other country's emigration loss, points to a certain potential for conflict between countries in their attempts to influence the structure of migration flows to their own national advantage. Indeed, our analysis suggests that migration is a somewhat less benign form of globalization than trade in that it tends to have opposite aggregate welfare effects in the sending and the receiving country. This is reinforced if we take into account possible distortions in labor markets. Irrespective of these aggregate welfare effects, however, the factor proportions approach does stress that migration unquestionably confers significant gains on migrants themselves. All of this amounts to a pretty convincing argument for establishing something like an International Migration Organization that facilitates international coordination in matters of migration, for the sake of exploiting the huge potential that migration holds for alleviation of poverty.

A final conclusion that we draw from our factor proportions model relates to international convergence. Even if migration is driven by international wage gaps, and even if we rule out all economies of scale, we cannot expect migration to always exert a strong effect towards international convergence of wage rates. The reason is that migration is likely to involve simultaneous movements of different types of labor, whence the response of any one wage rate to the change in the corresponding labor supply may no longer be described as a move along a "downward-sloping" labor demand curve. The potential for anti-convergence effects of migration on wages is, of course, magnified if we allow for increasing returns to scale, such as in the New Economic Geography.

The dynamic interaction between migration and capital formation is all but trivial. In the fourth section, we first discussed the effects of labor market integration when the capital market is already integrated internationally and capital accumulation is subject to adjustment costs at the firm level. The analysis suggests that higher immigration (emigration) causes increased (reduced) capital accumulation. Nevertheless, if the capital stock and therefore wages are initially low, labor market integration leads first to emigration and decreased (but still positive) capital accumulation; that is, we may observe emigration and capital inflows at the same time. Later in the transition, there could be immigration (i.e., reversed migration) while the capital stock approaches its steady-state level. Distinguishing tradable and non-tradable goods allows us to consider novel welfare effects. If the non-tradable goods sector is land-intensive (capturing the housing sector), immigration may reduce individual welfare of individuals with low endowments of land. The reason is that immigration drives up the relative price for non-tradable goods, because land is subject to a dilution effect when population density rises. By contrast, landowners may benefit from immigration, due to an increase in the price of land. This analysis thus shifts the focus from wage effects of migration to distributional effects, which are related to unequal landownership among natives.

In the fifth section, we reviewed productivity effects of high-skilled migration, which are related to the formation of both human capital and knowledge capital. Standard endogenous growth theory suggests that immigration speeds up the innovation process and leads to scale effects with respect to either the growth rate or the level of GDP per capita. As a result, selective migration towards high-skilled labor may affect the direction of technical change towards innovations, possibly raising the productivity of skilled relative to unskilled labor. However, congestion effects from higher population density on total factor productivity may be a counteracting force to standard scale effects. We also discussed recent research on possibly positive effects of high-skilled immigration on the domestic human capital stock of source economies, in turn affecting physical capital flows. A less restrictive immigration policy in advanced countries may significantly support human capital formation in source countries by improving emigration prospects of skilled workers. However, empirical evidence suggests that such a net gain is observed for a minority of developing countries and occurs under specific circumstances only. If this is the case, however, higher emigration may go along with capital inflows. Future research should certainly dig deeper into the dynamic relationship between migration and capital flows. It would be fruitful to extend spatial models with trade costs to allow for increasing returns and multiplicity of equilibrium. In the case where both history (i.e., initial conditions) and expectations potentially matter for the dynamics, the current theoretical literature provides little guidance for structural estimations.

In the sixth section we reviewed the empirical literature exploring the links between migration and trade. The discussion is based on the so-called gravity model of trade, in which trade between pairs of countries is related to measures of their respective sizes, preferences, and trade costs. While migration has obvious effects on size variables, the more interesting mechanisms involve trade costs. Networks of migrants may help overcome incomplete information about trading opportunities. They may also act as vehicles to enforce non-opportunistic behavior when information between partners in a transaction is asymmetric. The trade cost channel is potentially confounded by a channel that runs through preferences, e.g., if immigrants overweight goods from their countries of origin in their spending. Against this backdrop, Section 6 revisited the identification of the overall trade-creating effect of migration and its breakdown into the trade channel and the preference channel. After providing a short summary of older work, we have discussed the conditions under which causal effects of migration on trade can be identified. The section also clarified the role of product differentiation for the size of estimated effects, discussed the role of immigrants' eduction and occupation, and highlighted the distinction between the intensive and the extensive margins. We described the role of trading partner characteristics, and distinguished between direct and indirect networks and their trade-enhancing potential.

We conclude that migrant networks do indeed foster trade, and that there the link has a causal interpretation. However, as econometric techniques have become more sophisticated, the estimated effects are smaller than was found in the earlier literature. There is still uncertainty concerning the clean distinction between trade cost and preference effects, and more work in this area would be valuable. Another direction for promising research lies in the joint empirical modeling of migration, capital, and trade flows. Finally, all studies surveyed allow only for a partial equilibrium interpretation of the effects of migrant networks on trade. Implementing these estimates in general equilibrium models would allow simulation of the indirect effects of those networks on the pattern of international trade and on the welfare.

APPENDIX. SOLUTION OF THE MODEL IN SECTION 5.1.2

We now prove the claims in Section 5.1.2 by fully solving the vertical innovation model. Combining $p(i) = r/\alpha$ with

$$p(i) = \alpha B \left(\frac{A(i)}{x(i)} \right)^{1-\alpha} \left(L^Y\right)^{\beta} Z^{1-\alpha-\beta} = \frac{\partial Y}{\partial x(i)} \tag{18.67}$$

and solving for $x(i)$ implies

$$x(i) = A(i) \left(B \frac{\alpha^2}{r} \left(L^Y\right)^{\beta} Z^{1-\alpha-\beta} \right)^{\frac{1}{1-\alpha}} \tag{18.68}$$

Taking wage rate w as given, *ex ante* of production, producer i chooses R&D labor input in period t to maximize

$$\Pi_{t+1}(i) \equiv \frac{\pi_{t+1}(i)}{1+r} - w_t l_t(i) - w_t f \tag{18.69}$$

where future profits $\pi_{t+1}(i) \equiv (p_t(i) - r) x_t(i)$ read

$$\pi_{t+1}(i) = (1-\alpha) r^{-\frac{\alpha}{1-\alpha}} \alpha^{\frac{1+\alpha}{1-\alpha}} B^{\frac{1}{1-\alpha}} A_t \Lambda(l_t(i)) \left(L^Y\right)^{\frac{\beta}{1-\alpha}} Z^{\frac{1-\alpha-\beta}{1-\alpha}} \tag{18.70}$$

according to $p(i) = r/\alpha$, (18.68) and (18.45). Using (18.70) in (18.69), the first-order condition for the optimal choice of R&D input $l_t(i)$ is

$$\chi \Lambda'(l(i)) B^{\frac{1}{1-\alpha}} \left(L^Y\right)^{\frac{\beta}{1-\alpha}} Z^{\frac{1-\alpha-\beta}{1-\alpha}} = w \tag{18.71}$$

$\chi \equiv \frac{1-\alpha}{1+r} r^{-\frac{\alpha}{1-\alpha}} \alpha^{\frac{1+\alpha}{1-\alpha}}$, where $w \equiv w/A$ is the productivity-adjusted wage rate. In equilibrium with free entry, $\Pi(i) = 0$ for all i. Thus,

$$\chi \Lambda(l(i)) B^{\frac{1}{1-\alpha}} \left(L^Y\right)^{\frac{\beta}{1-\alpha}} Z^{\frac{1-\alpha-\beta}{1-\alpha}} = (l(i) + f) w \tag{18.72}$$

Combining (18.71) and (18.72), the equilibrium R&D labor input of each firm, $\tilde{l}$, is time- invariant and uniquely given by

$$0 = \frac{\Lambda\left(\tilde{l}\right)}{\Lambda'\left(\tilde{l}\right)} - \tilde{l} - f \tag{18.73}$$

Moreover, the wage rate is given by $w = \partial Y / \partial L^Y = (1-\alpha) y / l^Y$. Using (18.44) and (18.46), we find

$$w = (1-\alpha) \left(\frac{\alpha^2}{r} \right)^{\frac{\alpha}{1-\alpha}} B^{\frac{1}{1-\alpha}} \left(\frac{Z}{L^Y} \right)^{\frac{1-\alpha-\beta}{1-\alpha}} N \tag{18.74}$$

Combining (18.71) and (18.74) then implies

$$N_{t+1} = \frac{\alpha \Lambda'\left(\tilde{l}\right)}{1+r} L_{t+1}^{Y} \tag{18.75}$$

In labor market equilibrium,

$$L_t^Y + \int_0^{N_{t+1}} \left(l_t(i) + f\right) \mathrm{d}i = L_t \tag{18.76}$$

Using $l(i) = \tilde{l}$ for all i, (18.73), (18.75) and $L_{t+1} = (1+n)L_t$, we see that the fraction of labor devoted to manufacturing evolves according to

$$l_{t+1}^{Y} = \xi.\left(1 - l_t^{Y}\right) \tag{18.77}$$

$\xi \equiv \frac{1+r}{\alpha \Lambda\left(\tilde{l}\right)(1+n)}$, where $l_0^Y = L_0^Y / L$ is given. Thus, l^Y is independent of population size, L, at all times. Moreover, the equilibrium number of firms is proportional to L, according to (18.75) and (18.77). Denote the steady-state value of l^Y by $\tilde{l}^Y$. According to (18.77), we have $\tilde{l}^Y = \xi / 1 + \kappa$. The steady state is globally stable if $\xi < 1$, which is well possible (recall that $\Lambda(0) = 1$ and $\Lambda' > 0$).

ACKNOWLEDGMENTS

Thanks are due to Katharina Erhardt, Bohdan Kukharskyy and Eva Spring for excellent research assistance.

REFERENCES

Abel, A.B., 1982. Dynamic effects of permanent and temporary tax policies in a q model of investment. J. Monetary Econ. 9 (3), 353–373.
Acemoglu, D., 1998. Why do new technologies complement skills? directed technical change and wage inequality. Q. J. Econ. 113, 1055–1089.
Acemoglu, D., 2002. Directed technical change. Rev. Econ. Stud. 69 (4), 781–809.
Acemoglu, D., Angrist, J., 2001. How large are human-capital externalities? Evidence from compulsory-schooling laws. NBER Macroecon. Annu. 15, 9–74.
Acemoglu, D., Johnson, S., 2007. Disease and development: The effect of life expectancy on economic growth. J. Polit. Econ. 115 (6), 925–985.
Aghion, P., Williamson, J.G., 1998. Growth, inequality, and globalization: theory, history, and policy. Cambridge University Press, Cambridge.
Aleksynska, M., Peri, G., 2011. Isolating the network effect of immigrants on trade. Working Papers 2011–26, CEPII research center.
Anderson, J.E., 2011. The gravity model. Annual Review of Economics 3, 133–160.
Anderson, J.E., van Wincoop, E., 2003. Gravity with gravitas: A solution to the border puzzle. Am. Econ. Rev. 93 (1), 170–192.
Anderson, J.E., van Wincoop, E., 2004. Trade costs. J. Econ. Lit. 42 (3), 691–751.
Arrow, K.J., 1962. The economic implications of learning by doing. Rev. Econ. Stud. 29 (3), 155–173.

Aydemir, A., Borjas, G.J., 2007. Cross-country variation in the impact of international migration: Canada, mexico, and the united states. J. Eur. Econ. Assoc. 5 (4), 663–708.

Aydemir, A., Borjas, G.J., 2011. Attenuation bias in measuring the wage impact of immigration. J. Labor Econ. 29 (1), 69–112.

Baier, S.L., Bergstrand, J.H., 2007. Do free trade agreements actually increase members' international trade? J. Int. Econ. 71 (1), 72–95.

Baier, S.L., Bergstrand, J.H., 2009. Bonus Vetus OLS: A simple method for approximating international trade-cost effecs using the gravity equation. J. Int. Econ. 77 (1), 77–85.

Baldwin, R.E., Forslid, R., 2000. The core–periphery model and endogenous growth: Stabilizing and destabilizing integration. Economica 67 (267), 307–324.

Baldwin, R.E., Martin, P., 2004. Agglomeration and regional growth. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics. Elsevier, pp. 2671–2711, Volume 4, Chapter 60.

Bandyopadhyay, S., Coughlin, C.C., Wall, H.J., 2008. Ethnic networks and us exports. Rev. Int. Econ. 16 (1), 199–213.

Bardhan, A.D., Guhathakurta, S., 2004. Global linkages of subnational regions: Coastal exports and international networks. Contemp. Econ. Pol. 22 (2), 225–236.

Beine, M., Docquier, F., Rapoport, H., 2001. Brain drain and economic growth: Theory and evidence. J. Dev. Econ. 64 (1), 275–289.

Beine, M., Docquier, F., Rapoport, H., 2008. Brain drain and human capital formation in developing countries: Winners and losers. Econ. J. 118 (528), 631–652.

Bergstrand, J.H., Egger, P., Larch, M., 2013. Gravity redux: Estimation of gravity-equation coeffcients, elasticities of substitution, and general equilibrium comparative statics under asymmetric bilateral trade costs. J. Int. Econ. 89 (1), 110–121.

Berman, E., Bound, J., Griliches, Z., 1994. Changes in the demand for skilled labor within U.S. manufacturing: Evidence from the annual survey of manufacturers. Q. J. Econ. 109 (2), 367–397.

Berman, E., Bound, J., Machin, S., 1998. Implications of skill-biased technological change: International evidence. Q. J. Econ. 113 (4), 1245–1279.

Berry, R.A., Soligo, R., 1969. Some welfare aspects of international migration. J. Polit. Econ. 77 (5), 778–794.

Bhagwati, J., Hamada, K., 1974. The brain drain, international integration of markets for professionals and unemployment: A theoretical analysis. J. Dev. Econ. 1 (1), 19–42.

Borjas, G.J., 1987. Self-selection and the earnings of immigrants. Am. Econ. Rev. 77 (4), 531–553.

Borjas, G.J., 1999. The economic analysis of immigration. In: Ashenfelter, O., Card, D. (Eds.), In: Handbook of Population Economics, vol. 3A. Elsevier, pp. 1697–1760.

Borjas, G.J., 2003. The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. Q. J. Econ. 118 (4), 1335–1374.

Borjas, G.J., Freeman, R.B., Katz, L.F., 1997. How much do immigration and trade affect labor market outcomes? Brookings Paper. Econ. Activ. 1997 (1), 1–90.

Borjas, G.J., Grogger, J., Hanson, G.H., 2012. Comment: On estimating elasticities of substitution. J. Eur. Econ. Assoc. 10 (1), 198–210.

Bratti, M., De Benedictis, L., Santoni, G., 2012. On the pro-trade effects of immigrants. IZA Discussion Papers 6628, Institute for the Study of Labor (IZA).

Braun, J., 1993. Essays on Economic Growth and Migration. Ph. D. thesis, Harvard University, Cambridge MA.

Breinlich, H., Ottaviano, G.I.P., Temple, J.R., 2013. Regional growth and regional decline. In: Aghion, P., Durlauf, S.N. (Eds.), In: Handbook of Economic Growth, Volume 2, North Holland.

Briant, A., Combes, P.-P., Lafourcade, M., 2009. Product complexity, quality of institutions and the pro-trade effect of immigrants. CEPR Discussion Papers (7192).

Broda, C., Weinstein, D.E., 2006. Globalization and the gains from variety. Q. J. Econ. 121 (2), 541–585.

Bruder, J., 2004. Are trade and migration substitutes or complements? The case of Germany. Working paper, University of Rostock.

Bryant, J., Genc, M., Law, D., 2004. Trade and migration to New Zealand. Treasury Working Paper Series 04/18, New Zealand Treasury.

Buch, C.M., Kleinert, J., Toubal, F., 2006. Where enterprises lead, people follow?: Links between migration and fdi in Germany. Eur. Econ. Rev. 50 (8), 2017–2036.
Burda, M., 2006. Factor reallocation in eastern Germany after reunification. Am. Econ. Rev. 96 (2), 368–374.
Burda, M., Wyplosz, C., 1992. Human capital, investment and migration in an integrated europe. Eur. Econ. Rev. 36 (2–3), 677–684.
Burstein, A., Vogel, J., 2012. International trade, technology, and the skill premium. Working Paper.
Card, D., 1990. The impact of the Mariel boatlift on the miami labor-market. Ind. Labor Relat. Rev. 43 (2), 245–257.
Card, D., 2001. Immigrant inows native outows, and the local labor market impacts of higher immigration. J. Labor Econ. 19, 22–64.
Chaney, T., 2008. Distorted gravity: The intensive and extensive margins of international trade. Am. Econ. Rev. 98 (4), 1707–1721.
Chiquiar, D., Hanson, G.H., 2005. International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. J. Polit. Econ. 113 (2), 239–281.
Chiswick, B.P., 1999. Are immigrants favorably self-selected? Am. Econ. Rev. 89 (2), 181–185.
Chiswick, B.P., 2000. Are immigrants favorably self-selected? An economic analysis. In: Brettell, C.D., Hollifield, J.F. (Eds.), Migration Theory: Talking Across Disciplines. Routledge.
Ciccone, A., Peri, G., 2006. Identifying human-capital externalities: Theory with applications. Rev. Econ. Stud. 73 (2), 381–412.
Cline, W., 1997. Trade and Income Distribution. Institute for International Economics, Washington.
Co, C., Euzent, P., Martin, T., 2004. The export effect of immigration into the usa. Appl. Econ. 36 (6), 573–583.
Coeurdacier, N., Rey, H., 2013. Home bias in open economy financial macroeconomics. J. Econ. Lit. 51, 63–115.
Combes, P.-P., Lafourcade, M., Mayer, T., 2005. The trade-creating effects of business and social networks: evidence from france. J. Int. Econ. 66 (1), 1–29.
Coughlin, C.C., Wall, H.J., 2011. Ethnic networks and trade: Intensive versus extensive margins. Econ. Lett. 113 (1), 73–75.
D'Auria, F., Morrow, K.M., Pichelmann, K., 2008. Economic impact of migration flows following the 2004 EU enlargement process – a model based analysis. European Commission, Economic Papers 349.
Davis, D.R., 1998. Does European unemployment prop up American wages? National Labor Markets and Global Trade. Am. Econ. Rev. 88 (3), 478–494.
Davis, D.R., Weinstein, D.E., 2002. Technological superiority and the losses from migration. NBER Working Paper 8971.
Deardorff, A.V., 1979. Weak links in the chain of comparative advantage. J. Int. Econ. 9 (2), 197–209.
Deardorff, A.V., 2000. Factor prices and the factor content of trade revisited: What is the use? J. Int. Econ. 50 (1), 73–90.
Desmet, K., Rossi-Hansberg, E., 2014. Spacial Development. Am. Econ. Rev. 104 (4), 1211–1243.
Diamond, J.M., 1997. Guns, germs and steel: the fate of human societies. Cape, London.
Dixit, A.K., Norman, V.D., 1980. Theory of international trade: a dual, general equilibrium approach. Cambridge University Press, Cambridge.
Dixit, A.K., Stiglitz, J.E., 1977. Monopolistic competition and optimum product diversity. Am. Econ. Rev. 67 (3), 297–308.
Dornbusch, R., Fischer, S., Samuelson, P.A., 1977. Comparative advantage, trade, and payments in a Ricardian model with a continuum of goods. Am. Econ. Rev. 67 (5), 823–839.
Dornbusch, R., Fischer, S., Samuelson, P.A., 1980. Heckscher-ohlin trade theory with a continuum of goods. Q. J. Econ. 95 (2), 203–224.
Dunlevy, J.A., 2006. The inuence of corruption and language on the protrade effect of immigrants: Evidence from the American states. Rev. Econ. Stat. 88 (1), 182–186.
Dunlevy, J.A., Hutchinson, W.K., 1999. The impact of immigration on American import trade in the late nineteenth and early twentieth centuries. J. Econ. Hist. 59 (04), 1043–1062.

Dustmann, C., Fabbri, F., Preston, I., 2005. The impact of immigration on the British labour market. Econ. J. 115 (507), F324–F341.
Eaton, J., Kortum, S., 2002. Technology, geography, and trade. Econometrica 70 (5), 1741–1779.
Egger, P., von Ehrlich, M., Nelson, D., 2012. Migration and trade. World Econ. 35 (2), 216–241.
Epifani, P., Gancia, G., 2008. The skill bias of world trade. Econ. J. 118, 927–960.
Ethier, W.J., 1982. The general role of factor intensity in the theorems of international trade. Econ. Lett. 10, 337–342.
Faini, R., 1996. Increasing returns, migrations and convergence. J. Dev. Econ. 49 (1), 121–136.
Fan, C.S., Stark, O., 2011. A Theory of Migration as a Response to Occupational Stigma. Int. Econ. Rev. 52 (2), 549–571.
Feenstra, R., 2004a. Advanced international trade: theory and evidence. Princeton University Press, Princeton, NJ.
Feenstra, R., 2004b. Advanced International Trade: Theory and Evidence. Princeton University Press.
Feenstra, R., Hanson, G.H., 1996. Foreign investment, outsourcing and relative wages. In: The Political Economy of Trade. Policy: Papers in Honor of Jagdish Bhagwati. MIT Press.
Feenstra, R., Hanson, G.H., 1997. Foreign direct investment and relative wages: evidence from Mexico's maquiladoras. J. Int. Econ. 42, 371–393.
Feenstra, R., Hanson, G.H., 1999. The impact of outsourcing and high-technology capital on wages: Estimates for the United States, 1979–1990. Q. J. Econ. 114 (3), 907–940.
Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C., Mo, H., 2005. World trade flows: 1962–2000. NBER Working Papers 11040, National Bureau of Economic Research, Inc.
Felbermayr, G.J., Jung, B., 2009. The pro-trade effect of the brain drain: Sorting out confounding factors. Econ. Lett. 104 (2), 72–75.
Felbermayr, G.J., Kohler, W., 2006a. Exploring the intensive and extensive margins of world trade. Review of World Economics 142 (4), 642–674.
Felbermayr, G.J., Kohler, W., 2006b. Immigration and wages in general equilibrium: A theoretical perspective. In: Foders, F., Langhammer, R. (Eds.), Labor Mobility and the World Economy. Springer-Verlag, pp. 51–80.
Felbermayr, G.J., Kohler, W., 2007. Immigration and native welfare. Int. Econ. Rev. 48 (3), 731–760.
Felbermayr, G.J., Geis, W., Kohler, W., 2010a. Restrictive immigration policy in Germany: Pains and gains foregone? Review of World Economics 146 (1), 1–21.
Felbermayr, G.J., Toubal, F., 2012. Revisiting the trade-migration nexus: Evidence from new OECD data. World Dev. 40 (5), 928–937.
Felbermayr, G.J., Jung, B., Toubal, F., 2010b. Ethnic networks, information, and international trade: Revisiting the evidence. Annales d'Economie et de Statistique 97–98, 41–70.
Freeman, R.B., 2006. People flows in globalization. J. Econ. Perspect. 20 (2), 145–170.
Friedberg, R.M., 2001. The impact of mass migration on the Israeli labor market. Q. J. Econ. 116 (4), 1373–1408.
Friedberg, R.M., Hunt, J., 1995. The impact of immigrants on host country wages, employment and growth. J. Econ. Perspect. 9 (2), 23–44.
Fujita, M., Thisse, J.-F., 2002. Economics of agglomeration: cities, industrial location, and regional growth. Cambridge University Press, Cambridge.
Fujita, M., Krugman, P.R., Venables, A.J., 2001. The spatial economy: cities, regions, and international trade. MIT Press, Cambridge, MA.
Galbraith, J.K., 1979. The nature of mass poverty. Harvard University Press, Cambridge, MA.
Gao, T., 2003. Ethnic Chinese networks and international investment: Evidence from inward FDI in china. J. Asian Econ. 14 (4), 611–629.
Genc, M., Gheasi, M., Nijkamp, P., Poot, J., 2011. The impact of immigration on international trade: A meta-analysis. IZA Discussion Papers 6145, Institute for the Study of Labor (IZA).
Gennaioli, N., Porta, R.L., de Silanes, F.L., Shleifer, A., 2011. Human capital and regional development. NBER working papers, National Bureau of Economic Research, Inc.
Ghatak, S., Pop-Silaghi, M.I., Daly, V., 2009. Trade and migration flows between some CEE countries and the UK. J. Int. Trade Econ. Dev. 18 (1), 61–78.

Girma, S., Yu, Z., 2002. The link between immigration and trade: Evidence from the United Kingdom. Review of World Economics 138 (1), 115–130.

Gonzalez, L., Ortega, F., 2013. Immigration and housing booms: Evidence from Spain. J. Reg. Sci. 53, 37–59.

Gould, D.M., 1994. Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. Rev. Econ. Stat. 76 (2), 302–316.

Grogger, J., Hanson, G.H., 2011. Income maximization and the selection and sorting of international migrants. J. Dev. Econ. 95 (1), 42–57.

Grossman, G.M., Rossi-Hansberg, E., 2010. External economies and international trade redux. Q. J. Econ. 125, 829–858.

Grossmann, V., 2009. Entrepreneurial innovation and economic growth. J. Macroecon. 31 (4), 602–613.

Grossmann, V., 2013. Structural change, urban congestion, and the end of growth. Rev. Dev. Econ. 17 (2), 165–181.

Grossmann, V., Stadelmann, D., 2011. Does international mobility of highskilled workers aggravate between-country inequality? J. Dev. Econ. 95 (1), 88–94.

Grossmann, V., Stadelmann, D., 2012. Does high-skilled migration affect publicly financed investments? Rev. Int. Econ. 20 (5), 944–959.

Grossmann, V., Stadelmann, D., 2013. Wage effects of high-skilled migration: International evidence. World Bank Econ. Rev. 27 (2), 297–319.

Grossmann, V., Schäfer, A., Steger, T.M., 2012. Migration, capital formation, and house prices. CESifo Working Paper No. 4146.

Hanson, G.H., 2006. Illegal migration from Mexico to the United States. J. Econ. Lit. 44 (4), 869–924.

Hanson, G.H., 2009a. The economic consequences of the international migration of labor. Annual Review of Economics 1 (1), 179–208.

Hanson, G.H., 2009b. The Economics and Policy of Illegal Immigration in the United States. Migration Policy Institute, Washington, DC.

Hanson, G.H., 2010. International migration and the developing world. In: Handbook of Development Economics, pp. 4363–4414, Chapter 66.

Harrison, A., McLaren, J., McMillan, M., 2011. Recent perspectives on trade and inequality. Annual Review of Economics 3 (1), 261–289.

Hatton, T.J., 2007. Should we have a WTO for international migration? Econ. Pol. 22 (50), 339–383.

Hatton, T.J., Williamson, J.G., 1998. The age of mass migration: causes and economic impact. Oxford University Press, New York.

Hatton, T.J., Williamson, J.G., 2005. Global migration and the world economy: two centuries of policy and performance. MIT Press, Cambridge, MA.

Hatton, T.J., Williamson, J.G., 2006. International migration in the long run: Positive selection, negative selection, and policy. In: Foders, F., Langhammer, R. (Eds.), Labor Mobility and the World Economy. Springer-Verlag, pp. 1–31.

Hatzigeorgiou, A., 2010a. Does immigration stimulate foreign trade? Evidence from Sweden. J. Econ. Integrat. 25, 376–402.

Hatzigeorgiou, A., 2010b. Migration as trade facilitation: Assessing the links between international trade and migration. The B.E. Journal of Economic Analysis and Policy 10 (1), 24.

Hayashi, F., 1982. Tobin's marginal q and average q: A neoclassical interpretation. Econometrica 50 (1), 213–224.

Head, K., Mayer, T., 2013. Gravity equations: Workhorse, toolkit, cookbook. In: Helpman, E., Gopinath, G., Rogoff, K. (Eds.), In: Handbook of International Economics, Volume IV. Elsevier.

Head, K., Ries, J., 1998. Immigration and trade creation: Econometric evidence from Canada. Can. J. Econ. 31 (1), 47–62.

Helliwell, J.F., 1997. National borders, trade and migration. Pac. Econ. Rev. 2 (3), 165–185.

Helpman, E., Krugman, P., 1985. Market structure and foreign trade : increasingreturns, imperfect competition, and the international economy. MIT Press, Cambridge, MA.

Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: Trading partners and trading volumes. Q. J. Econ. 123 (2), 441–487.

Hendricks, L., 2002. How important is human capital for development? Evidence from immigrant earnings. Am. Econ. Rev. 92 (1), 198–219.
Herander, M.G., Saavedra, L.A., 2005. Exports and the structure of immigrant-based networks: The role of geographic proximity. Rev. Econ. Stat. 87 (2), 323–335.
Hiller, S., 2011. The export promoting effect of emigration: Evidence from Denmark. Working Papers 11–08, Aarhus University, Aarhus School of Business.
Hunt, J., Gauthier-Loiselle, M., 2010. How much does immigration boost innovation? Am. Econ. J. Macroecon. 2 (2), 31–56.
Iranzo, S., Peri, G., 2009. Schooling externalities, technology, and productivity: Theory and evidence from U.S. states. Rev. Econ. Stat. 91 (2), 420–431.
Jansen, M., Piermartini, R., 2009. Temporary migration and bilateral trade flows. World Econ. 32 (5), 735–753.
Javorcik, B.S., Özden, C., Spatareanu, M., Neagu, C., 2011. Migrant networks and foreign direct investment. J. Dev. Econ. 94 (2), 231–241.
Jeanty, P.W., Partridge, M., Irwin, E., 2010. Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. Reg. Sci. Urban. Econ. 40 (5), 343–352.
Jevčák, A., Setzer, R., Suardi, M., 2010. Determinants of capital flows to the new EU member states before and during the financial crisis. In: European Commission (Ed.), Capital flows to converging European economies – from boom to drought and beyond, Occasional Papers No. 75.
Jones, C.I., 1995. R&d-based models of economic growth. J. Polit. Econ. 103 (4), 759–784.
Jones, C.I., 1999. Growth: With or without scale effects? Am. Econ. Rev. 89 (2), 139–144.
Jones, C.I., 2005. Growth and ideas. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. Elsevier, pp. 1063–1111, Volume 1 of Handbook of Economic Growth, Chapter 16.
Kreickemeier, U., Wrona, J., 2011. Two-way migration between similar countries. University of Tübingen Working Papers in Economics and Finance, No 1, 2012.
Krugman, P., 1980. Scale economies, product differentiation, and the pattern of trade. Am. Econ. Rev. 70 (5), 950–959.
Krugman, P., 1991a. History versus expectations. Q. J. Econ. 106, 651–677.
Krugman, P., 1991b. Increasing returns and economic geography. J. Polit. Econ. 99 (3), 483–499.
Krugman, P., 1995. Growing world trade: Causes and consequences. Brookings Paper. Econ. Activ. 1995 (1), 327–377.
Krugman, P., 2000. Technology, trade and factor prices. J. Int. Econ. 50, 51–71.
Krugman, P., 2008. Trade and wages, reconsidered. Brookings Paper. Econ. Activ. 2008 (1), 103–137.
Kugler, M., Rapoport, H., 2007. International labor and capital flows: Complements or substitutes? Econ. Lett. 94 (2), 155–162.
Kugler, M., Levintal, O., Rapoport, H., 2013. Migration and cross-border financial flows. IZA Discussion Paper No. 7548.
Lalonde, R.J., Topel, R.H., 1997. Economic impact of international migration and the economic performance of migrants. In: Rosenzweig, M., Stark, O. (Eds.), Handbook of Population and Family Economics. Elsevier, pp. 799–850.
Leamer, E.E., 1997. In search of Stolper-Samuelson linkages between international trade and lower wages. In: Collins, S. (Ed.), Imports, Exports and the American Worker. Brookings Institution, pp. 141–214.
Leamer, E.E., 2000. What's the use of factor-contents? J. Int. Econ. 50 (1), 17–49.
Leamer, E.E., 2012. The Craft of Economics: Lessons from the Heckscher–Ohlin Framework. MIT Press, Cambridge, MA.
Leamer, E.E., Levinsohn, J.A., 1995. International trade theory: The evidence. In: Grossman, G., Rogoff, K. (Eds.), In: Handbook of International Economics, vol. III. Elsevier, pp. 1339–1394.
Legrain, P., 2009. Immigrants: Your Country Needs Them. Little, Brown, London.
Linder, S.B., 1961. An Essay on Trade and Transformation. Almqvist & Wiksell, Stockholm.
Lucas, R.E.J., 1988. On the mechanics of economic development. J. Monetary Econ. 22 (1), 3–42.
Lucas, R.E.J., 1990. Why doesn't capital flow from rich to poor countries? Am. Econ. Rev. 80 (2), 92–96.
Lucas, R.E.J., 2006. Migration and economic development in Africa: A review of evidence. J. Afr. Econ. 15 (2), 337–395.

Lundborg, P., Segerstrom, P.S., 1998. The growth and welfare effects of international mass migration. Working Paper Series 146, Trade Union Institute for Economic Research.
Markusen, J.R., 1983. Factor movements and commodity trade as complements. J. Int. Econ. 14 (3–4), 341–356.
Markusen, J.R., Svensson, L.E.O., 1985. Trade in goods and factors with international differences in technology. Int. Econ. Rev. 26 (1), 175–192.
Melitz, M.J., 2003. The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. Econometrica 71 (6), 1695–1725.
Millimet, D.L., Osang, T., 2007. Do state borders matter for U.S. intranational trade? The role of history and internal migration. Can. J. Econ. 40 (1), 93–126.
Mishra, P., 2007. Emigration and wages in source countries: Evidence from Mexico. J. Dev. Econ. 82 (1), 180–199.
Moraga, J.F.-H., 2011. New evidence on emigrant selection. Rev. Econ. Stat. 93 (1), 72–96.
Mountford, A., 1997. Can a brain drain be good for growth in the source economy? J. Dev. Econ. 53 (2), 287–303.
Mundra, K., 2005. Immigration and international trade: A semiparametric empirical investigation. Journal of International Trade and Economic Development 14 (1), 65–91.
Nygaard, C., 2011. International migration, housing demand and access to homeownership in the UK. Urban Stud. 48 (11), 2211–2229.
Obstfeld, M., Taylor, A.M., 2004. Global Capital Markets – Integration, Crisis, and Growth. Cambridge University Press, Cambridge.
OECD, 2011. International Migration Outlook 2011. OECD, Paris.
Ohlin, B.G., 1933. Interregional and international trade. Harvard Univesity Press, Cambridge, MA.
Okawa, Y., van Wincoop, E., 2012. Gravity in international finance. J. Int. Econ. 87, 205–215.
Olivero, M.P., Yotov, Y.V., 2012. Dynamic gravity: endogenous country size and asset accumulation. Can. J. Econ. 45 (1), 64–92.
O'Rourke, K.H., Williamson, J.G., 1999. Globalization and history: the evolution of a nineteenth century Atlantic economy. MIT Press, Cambridge, MA.
Ortega, F., Peri, G., 2011. The aggregate effects of trade and migration: Evidence from oecd countries. IZA discussion papers, IZA Discussion Paper.
Ortega, F., Peri, G., 2014. Openess and income: The role of trade and migration. J. Int. Econ. 92, 231–251.
Ottaviano, G.I.P., Peri, G., 2012. Rethinking the effect of immigration on wages. J. Eur. Econ. Assoc. 10 (1), 152–197.
Ottaviano, G.I.P., Thisse, J.-F., 2004. Agglomeration and economic geography. In: Henderson, J.V., Thisse, J.F. (Eds.), In: Handbook of Regional and Urban Economics, Volume 4. Elsevier, pp. 2563–2608, Chapter 58.
Ozden, C., Parsons, C.R., Schiff, M., Walmsley, T.L., 2012. Where on earth is everybody ? The evolution of global bilateral migration 1960–2000. World Bank Econ. Rev, forthcoming.
Panagariya, A., 2000. Evaluating the factor-content approach to measuring the effect of trade on wage inequality. J. Int. Econ. 50, 91–116.
Parsons, C.R., 2012. Do migrants really foster trade? The trade–migration nexus, a panel approach 1960–2000. Policy Research Working Paper Series 6034, World Bank.
Parsons, C.R., Vézina, P.-L., 2013. Migrant networks and trade: The Vietnamese boat people as a natural experiment. Oxford University, mimeo.
Passel, J.S., Cohn, D., 2009. A portrait of unauthorized immigrants in the United States. Pew Research Center Publications. Available online at www.pewhispanic.org.
Peri, G., Requena-Silvente, F., 2010. The trade creation effect of immigrants: evidence from the remarkable case of spain. Can. J. Econ. 43 (4), 1433–1459.
Rappaport, J., 2005. How does labor mobility affect income convergence? J. Econ. Dynam. Contr. 29 (3), 567–581.
Rauch, J.E., 1999. Networks versus markets in international trade. J. Int. Econ. 48 (1), 7–35.
Rauch, J.E., Trindade, V., 2002. Ethnic Chinese networks in international trade. Rev. Econ. Stat. 84 (1), 116–130.

Reichlin, P., Rustichini, A., 1998. Diverging patterns with endogenous labor migration. J. Econ. Dynam. Contr. 22 (5), 703–728.
Richardson, J.D., 1995. Income inequality and trade: How to think, what to conclude. J. Econ. Perspect. 9 (3), 33–55.
Romer, P.M., 1986. Increasing returns and long-run growth. J. Polit. Econ. 94 (5), 1002–1037.
Romer, P.M., 1990. Endogenous technological change. J. Polit. Econ. 98 (5), S71–S102.
Saint-Paul, G., 2004. The brain drain: Some evidence from European expatriates in the United States. IZA Discussion Papers 1310.
Saiz, A., 2003. Room in the kitchen for the melting pot: Immigration and rental prices. Rev. Econ. Stat. 85 (3), 502–521.
Saiz, A., 2007. Immigration and housing rents in American cities. J. Urban Econ. 61 (2), 345–371.
Sala-i-Martin, X., 2006. The world distribution of income: Falling poverty and . . . convergence, period. Q. J. Econ. 121 (2), 351–397.
Samuelson, P.A., 1952. The transfer problem and transport costs: The terms of trade when impediments are absent. Econ. J. 62 (246), 278–304.
Schäfer, A., Steger, T., 2014. Journey into the unknown? Economic consequences of factor market integration under increasing returns to scale. Rev. Int. Econ, submitted.
Sinn, H.-W., 2002. Germany's economic unification: An assessment after ten years. Rev. Int. Econ. 10 (1), 113–128.
Solimano, A., Watts, N., 2005. International migration, capital flows and the global economy: A long run view. Serie Macroeconomia del Desarollo 35.
Stark, O., 1993. The migration of labor. Blackwell, Cambridge, MA.
Stark, O., Taylor, J.E., 1991. Migration incentives, migration types: The role of relative deprivation. Econ. J. 101 (408), 1163–1178.
Stark, O., Helmenstein, C., Prskawetz, A., 1997, August. A brain gain with a brain drain. Econ. Lett. 55 (2), 227–234.
Tadesse, B., White, R., 2008. Cultural distance and the us immigrant–trade link. World Econ. 31 (8), 1078–1096.
Tai, S., 2009. Market structure and the link between migration and trade. Review of World Economics - Weltwirtschaftliches Archiv 145 (2), 225–249.
Taylor, A.M., Williamson, J.G., 1997. Convergence in the age of mass migration. Eur. Rev. Econ. Hist. 1 (1), 27–63.
Tinbergen, J., 1962. Shaping the World Economy; Suggestions for an International Economic Policy. Twentieth Century Fund, New York.
Tong, S.Y., 2005. Ethnic networks in fdi and the impact of institutional development. Rev. Dev. Econ. 9 (4), 563–580.
Topel, R.H., 1997. Factor proportions and relative wages: The supply-side determinants of wage inequality. J. Econ. Perspect. 11 (2), 55–74.
United Nations Secretary General, 2013. International migration and development. In: United Nations - General Assembly, United Nations.
Vézina, P.-L., 2012. How migrant networks facilitate trade: Evidence from Swiss exports. Swiss Journal of Economics and Statistics, forthcoming.
Wagner, D., Head, K., Ries, J., 2002. Immigration and the trade of provinces. Scot. J. Polit. Econ. 49 (5), 507–525.
White, R., Tadesse, B., 2007. Immigration policy, cultural pluralism and trade: Evidence from the white Australia policy. Pac. Econ. Rev. 12 (4), 489–509.
Williamson, J.G., 1997. Globalization and inequality: Past and present. World Bank Res. Obs. 12 (2), 117–135.
Wong, K., 1995. International trade in goods and factor mobility. MIT Press, Cambridge, MA.
Wooldridge, J., 2002. Econometric analysis of cross section and panel data. MIT Press.
World Bank, 2006. Global Economic Prospects: Economic Implications of Remittances and Migration. World Bank, Washington, DC.
Young, A., 1998. Growth without scale effects. J. Polit. Econ. 106 (1), 41–63.