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

Gabriel Felbermayr
Volker Grossmann
Wilhelm Kohler

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Gabriel Felbermayr
University of Munich, Ifo Institute and GEP Nottingham

Volker Grossmann
University of Fribourg, IZA and CESifo

Wilhelm Kohler
University of Tübingen, CESifo and GEP Nottingham

ABSTRACT

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

In this paper, we provide an overview of the relationship between international migration and international trade as well as capital movements. After taking a brief historical perspective, we first investigate migration flows between two countries in a static, neoclassical context. We allow for a disaggregated view of migration that distinguishes between different types of labor and emphasizes the distinction between migration flows and pre-existing stocks. We focus on different welfare channels, on internal income distribution, international income convergence and on whether migration and trade are substitutes or complements. Complementarity/substitutability hinges on whether countries share the same technology, and the pivotal question is whether or not technology is convex. Generally, under substitutability between trade and migration and with convex technology, globalization tends to lead to convergence. Moreover, under non-convex technology trade and migration tend to be complements. Turning to dynamic models with capital adjustment costs and capital mobility, the same is true for the relationship between migration and capital flows. Nevertheless, in neoclassical models, we may observe emigration at the same time as capital accumulates during the transition to a steady state. Moreover, we can explain reverse migration. We also touch upon the effects of migration on the accumulation of both knowledge and human capital, by invoking endogenous growth theory. Finally, we review the empirical literature exploring the link 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. We revisit the identification of the overall trade-creating effect of migration and its break-down into the trade channel and the preference channel. We clarify the role of product differentiation for the size of estimated effects, discuss the role of immigrants' education and occupation, and emphasize direct and indirect networks and their trade-enhancing potential.

JEL Classification: F1, F2, F4

Keywords: migration, international trade, capital movements, capital formation, globalization

Corresponding author:

Volker Grossmann
University of Fribourg
Bd. de Pérolles 90
1700 Fribourg
Switzerland
E-mail: volker.grossmann@unifr.ch

Contents

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. To put it in modern jargon, the "very long-run" history of globalization features a complementarity, indeed a causal relationship between migration and trade, meaning that migration leads to (more) trade between the sending and receiving countries. The key force underlying this complementarity was that in their "new countries" migrants eventually ended up producing goods which were in short supply in their "old countries", mostly for reasons of nature and climate.[1] Since trade is a precondition for capital movements, a similar "very long-run" complementarity relationship also holds for capital movements and both migration and trade.

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. In this chapter, we want to give an overview of what modern economic analysis tells us about this relationship.

Currently, an estimated three percent of the world population live outside their countries of birth. This is commonly regarded as a low figure. But what is the benchmark against which to judge? Perhaps more informative is a comparison of living conditions in different parts of the world. Using data from the World Bank World Development Indicators and calculating bottom and top percentiles we obtain a first and very rough impression of the amount of inequality between countries. In 2007, the 25th

---

[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* Aug 4th, 2012.

Figure 1. International income inequality and convergence

GDP per capita, PPP (constant 2005 int. $)
Deviations from unweighted mean

Data: World Bank, World Development Indicators
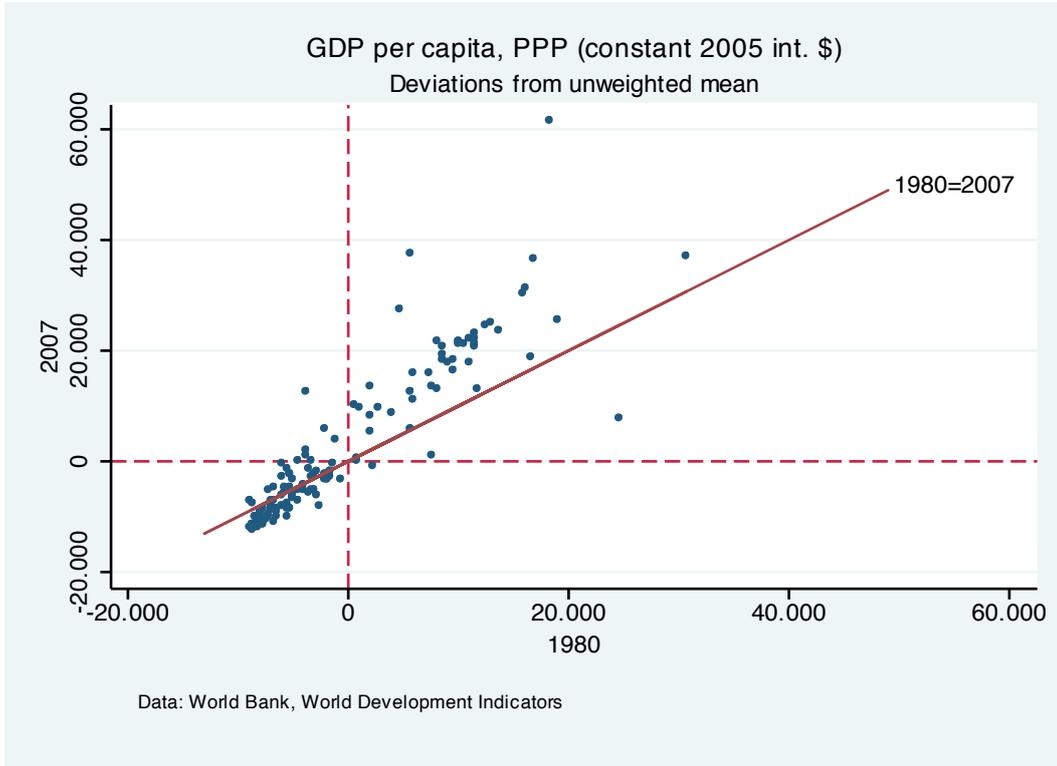
percentile of GDP per capita (at international PPP) as a fraction of the 75th percentile is a mere 0.125, *down* from 0.138 in 1980. For private household expenditure, the fraction was 0.140 in 2007, *down* from 0.179 in 1980.[2] Figures of a similar magnitude have been presented by Freeman (2006), based on occupation-specific wages taken from the NBER Wages around the World data base. What is striking is both, the degree of *inequality*, as well as *lack of convergence*. Figure 1 gives a somewhat more comprehensive impression by plotting country-specific deviations from an unweighted mean for 1980 against 2007. Data points above (below) the line in the positive (negative orthant indicate divergence.[3]

If this quick impression of international inequality is any indication of country-specific determinants of differences in worker's marginal productivity in different coun-

---

[2]We should hasten to add an important remark of qualification. Our focus here is not global inequality, which would require looking at internal distribution of income within countries. The point that we are trying to make here is that enormous income gaps still exist between countries that will serve as powerful incentives for future migration, despite all migration flows that we have observed in the past. Needless to say that 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).

[3]A similar picture is obtained using household expenditure per capita.

2

tries of this world, then international migration seems like an important key to improving living conditions of the poorer half of the world's population - without any need of international transfers, through a mere increase in the *world-wide efficiency of factor use*.[4] It also gives an impression of the emigration incentives that individual workers and families must feel, now and in the near future, in many poor countries of this world.

However, at the present, would-be migrants face stiff immigration restrictions in most rich countries. Indeed, looking at policies pursued in the developed world, one is tempted to say that international labor migration is something like a big *missing element* of present day globalization. *Goods markets* are characterized by many decades of multilateral and regional trade liberalization. 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 20th century apart from its 19th century counterpart.

The view of migration as a missing element, while correct when looking at present policies, does not do justice to the *sizable migration flows* that took place over the past 5 decades. Figure 2 presents a quick overview of the evolution of world trade, world-wide international migration and world-wide capital flows since World War II. There were sizable migration flows, in addition to a steady increase of world trade and an increasing importance of capital flows after the break down of the Bretton Woods system in the early 1970s. The figure demonstrates that the recent wave of economic globalization

---

[4]As we shall see below, empirical evidence suggests that only a small part of this international inequality is due to *country* characteristics, and a large part is due to *individual* characteristic like education and skills.

[5]This striking policy asymmetry is also emphasized by Freeman (2006).

Figure 2. Evolution of world trade, capital flows and migration

Exports of goods and services (% of GDP) [left axis]
Migrant stock (% of world population) [right axis]
Foreign direct investment, net inflows (% of GDP) [right axis]

Data: Worldbank, World Development Indicators
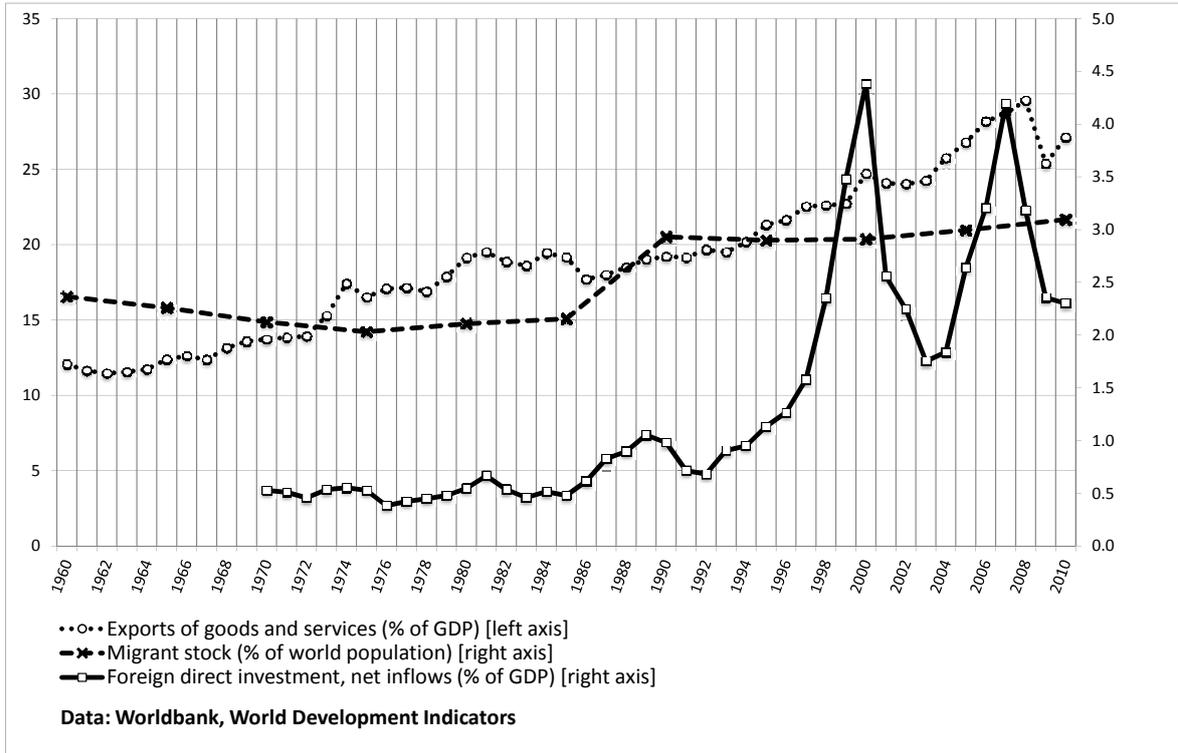
has featured a strong increase of all forms of internationalization. Obviously, it does not tell us anything about any one *causing* some other. Table 1 gives the present stocks of immigrants, relative to the population size as well as past net migration flows in the most important receiving countries

What are the dimensions in which countries (as opposed to individuals) may differ, and which may explain such inequality. This is probably one of the most intensively researched questions in economics. On a fundamental level, we may identify 6 *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 short-run change and policies, and our primary focus lies on migration, which is the most policy-restricted of all forms of internationalization, as emphasized above.

Against the backdrop of international income inequality and the associated inef-

4

Table 1. Immigration flows and stocks of immigrants for selected countries

Country	Net Migration Rates (5 year averages)							Stocks*
	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%
UK	0.07%	-0.17%	0.17%	0.35%	0.73%	1.61%	1.64%	11%
USA	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								

ficiency of world factor use, what is the appropriate policy stance vis à vis trade, migration and capital flows? Will enhanced trade, among poor countries or between poor and rich countries have a tendency to reduce international income gaps, thus also reducing the 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? Would more capital flowing from rich to poor countries help? Are the aforementioned income gaps likely to generate such capital movements?

As we shall detail below, *endowment-based* models of trade imply that trade and factor movements are *substitutes*. More specifically, trade and migration as well as capital movements are all working towards convergence of factor prices and towards a reduction of existing international income gaps. They also imply that more of any one of them, if it is in pursuit of arbitrage over existing international differences in goods or factor prices, implies less of at least one of the others.[6] Whatever substitutes for what, it has the same effect, qualitatively and - in extreme cases - even quantitatively, on factor prices. The deeper meaning of substitutability, stressed repeatedly by Ohlin (1933), one of the fathers of endowment-based trade theory, then is that inefficiencies in

---

[6]These statements indicate that complementarity may be discussed in a price or a quantity sense. We shall return to the exact way in which complementarity and substitutability between trade and factor movements may be defined in section 3 below.

Figure 3. Net migration and trade growth by country

Data: World Bank, World Development Indicators
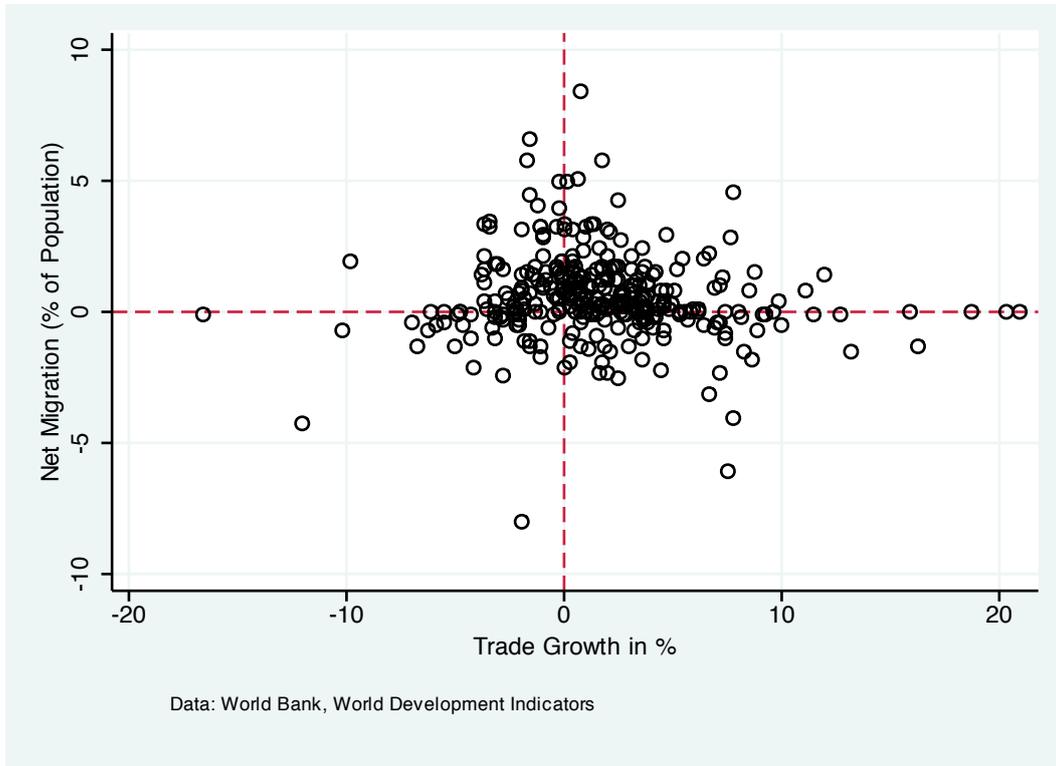
the world-wide use of different types of factors, as mirrored by factor price differences, may be removed, or at least reduced, by different forms of internationalization. The political challenge then is to rely on those forms that represent the least costly way towards a more efficient use of world factor endowments.

Endowment-based models typically assume identical technologies in all countries. However it is all too obvious that trade is determined by forces *other* than factor endowments, say international differences in technology, then there is a potential for trade and factor movements to become become *complements*, not just if looked at over the entire history of human settlement, but also over the long run of shorter horizons, say decades. However, as we shall see below, complementarity between trade and factor movements is not a forgone conclusion for *any* departure from the paradigm of endowment-based comparative advantage.

Looking at the overall evolution of trade and factor movements does not tell us anything about complementarity or substitutability. Quite obviously, the strong co-movement of trade, migration and capital flows evidenced in figure 2 reflects a strong co-movement in a downward direction of the barriers to all three forms of internation-

6

Figure 4. Net migration and international capital flows by country

Data: World Bank, World Development Indicators
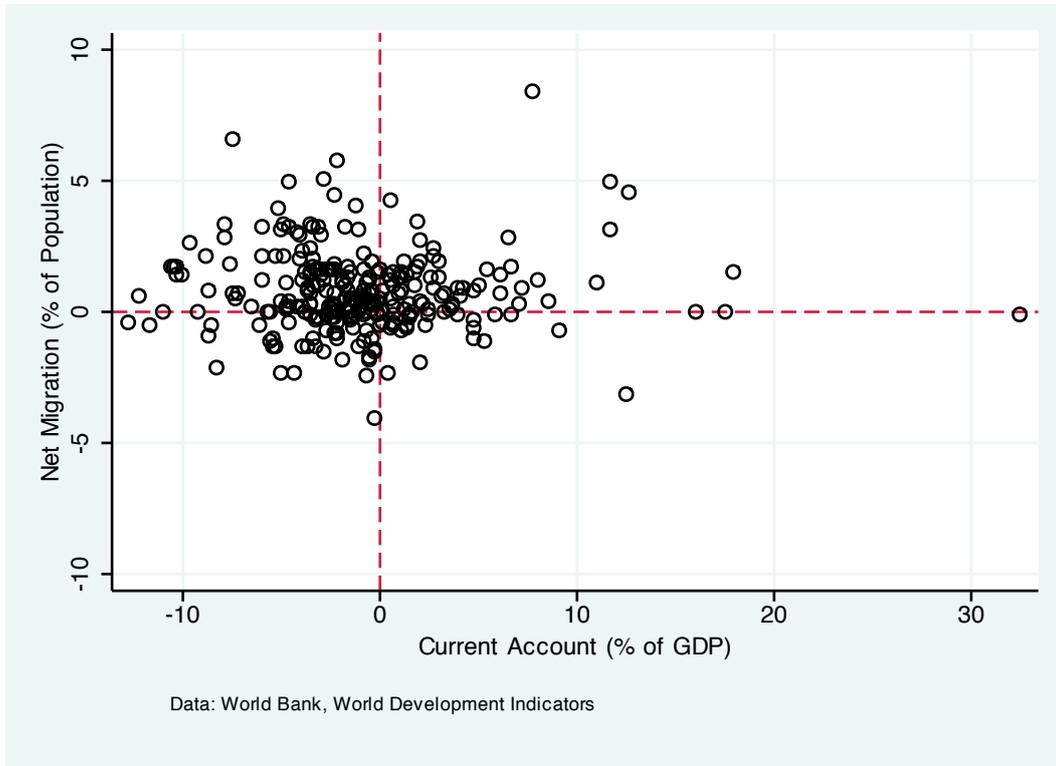
alization. Arguably, complementarity and substitutability should be addressed for an isolated increase in migration, say, holding fixed the barriers for the other types of internationalization. A first and quick attempt to resolve this problem is to look at *cross country comparisons*, since all countries will be affected symmetrically by secular changes in the different types of barriers involved. For instance, substitutability between *trade and migration* would imply that in a cross-country comparison, the countries experiencing high emigration or immigration rates would not at the same time be the ones that exhibit high growth rates in trade volumes. Figure 3 therefore presents a scatter plot of net-migration rates of individual countries in absolute values over all 5 year periods from 1960 to 2010 against corresponding 5-year-averages of annual growth rates of the same countries' merchandize trade, always taking gross trade (exports plus imports). The figure plots the pooled data. The figure quite clearly does not portray a situation of strong substitutability. And the situation does not change much if we we redraw the figure, lagging one or the other of the two series plotted by one 5-year period.

In a similar vein, if *migration and capital flows* were substitutes, then - other things

Figure 5. Capital flows and growth of trade by country

Data: World Bank, World Development Indicators
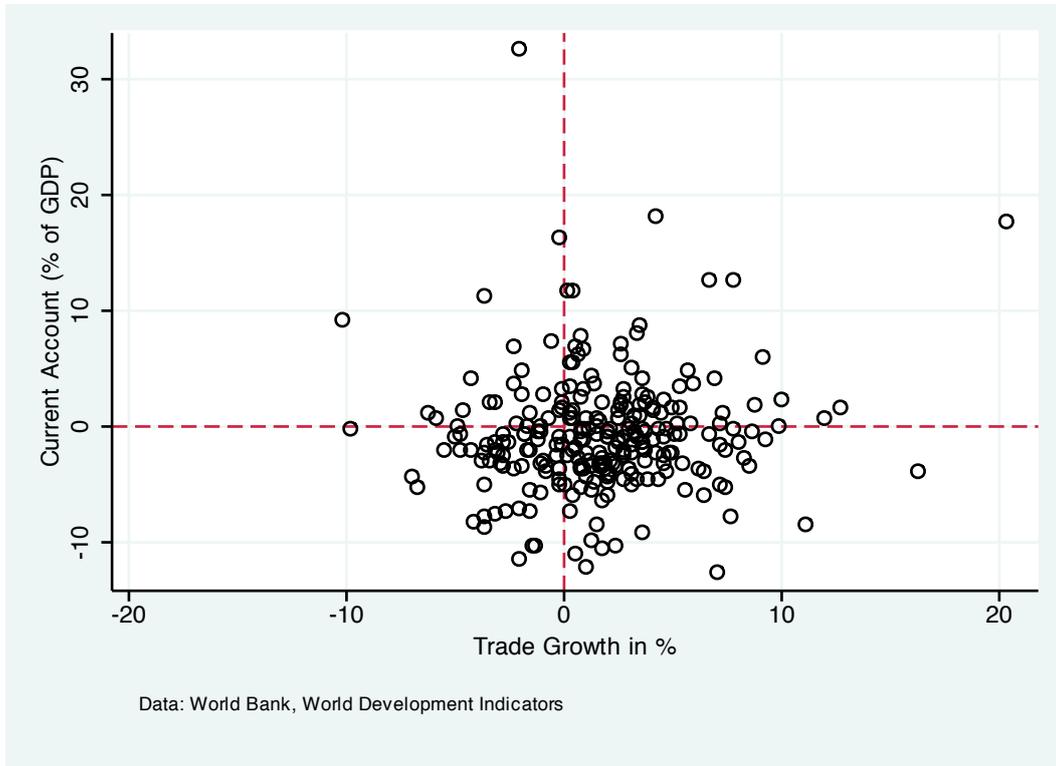
equal - we should observe high (low) emigration countries to have low (high) current account deficits (reflecting capital inflows), and high (low) immigration countries to have low (high) current account surpluses (reflecting capital outflows). Figure 4 therefore plots the same migration data against averages over the same 5-year periods of the current account, in absolute value and in percent of GDP. We restrict our plot to observations starting in 1975, which is when the episode of capital mobility has set in. Again, the figure does not tell us a convincing story of substitutability between migration and capital movements. Note that taking 5-year averages should avoid taking cyclical determinants of the current account for long-run capital movements.[7] As with figure 3, we have also run the same scatter plot, taking one and the other of the two series with a 5-period lag.

And finally, in figure 5 we plot the aforementioned current account data against the average increase of trade volumes used in figure 3. This checks for signs of substitutability between *trade and capital flows*, pretty much as figure 3 does for trade

---

[7]We are well aware of more elegant and better ways to remove cyclical elements in time series, but given the very rough nature of this exercise, more refined ways of calculating long-run trends would seem like an overkill.

and migration, and the outcome again is that convincing evidence of substitutability cannot be found.

The overall conclusion from this very rough inspection of the data is that the post World War II history of globalization is not characterized by a clear pattern of substitutability between internationalization of goods markets, international migration and international capital movements. Even without any formal test, the above figures appear to *reject substitutability*, while they do *not reject complementarity*. We must dig deeper in order to explore possible determinants and consequences of trade and factor movements. This is what we attempt to do in this survey. Our focus lies on international migration and its relationship to trade and capital movements as well as capital accumulation.

Before jumping into a focused theoretical and empirical analysis, section 2 first offers a brief tour through recent history and issues, essentially aiming at a theory-guided comparison between the migration that took place during the first wave of economic globalization towards the end of 19th century and until the Great War and the migration that took place a hundred years later.

One of the problems with the above quick inspection of the data is that lumping all types of labor into a single homogeneous factor is bound to cause aggregation error.Taking into account disaggregation, say along the skill dimension, a flow of immigration with a certain skill composition will be a complement to some types of native workers in the sending country, and a substitute to others. This is important particularly if migration takes place among countries that trade with each other, whence it is also important to analyze migration using models that feature international trade. A further problem in much of the discussion of complementarity versus substitutability between factor movements and trade is that factor movements are often discussed in the extreme form of complete factor mobility that wipes out all factor price differences. For migration at least, this is worryingly out of touch with reality where migration flows are governed by changes in policy-induced quantitative restrictions, moving economies from one distorted equilibrium to some other, without even getting close to equalization of wages.

Section 3 of our survey therefore pursues a general treatment of migration that

places emphasis on i) several types of labor with simultaneous movements in both directions between two countries, ii) the presence of pre-existing stocks of migrants in these countries, iii) migration flows governed by piecemeal changes in migration barriers (policy induced or otherwise), and iv) the trade effects as an important channel for the welfare and income convergence effects of migration flows. The section will take a neoclassical approach, with two features that separate it from subsequent sections. The first is that the model is static in nature, leaving consideration of accumulation issues to sections 4 and 5 below. The second is that it basically assumes free trade, leaving consideration of real trade costs that are endogenous to migration through network effects to section 6. Section 7 then concludes the paper.

2 A brief tour through history and issues

2.1 Mass migration of the 19th century

We have argued above that human history is, first and foremost, a history of migration. It was not until modern times, however, 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. Notably, however, at the beginning at least, receiving countries' governments did not restrict immigration flows. 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.[8]

---

[8]For extensive documentation of these aspects of the first era of mass migration, see Williamson (1997), Hatton and Williamson (1998), Aghion and Williamson (1998), O'Rourke and Williamson

The era of mass migration was also an era of large capital movements. Capital in the form of savings invested abroad was partly flowing in parallel to labor, i.e., from the European core to the Australasian and the American periphery, and partly from the European core to the Scandinavian periphery.[9] Moreover, migration and capital movements were paralleled by a dramatic fall in transport cost for goods, leading to a significant increase also in trade between both sending and receiving countries of capital and labor.[10] These observations beg two questions. The first is why capital was following labor in flowing to rich, rather than poor countries.[11] The second is whether this first wave of economic globalization testifies to complementarity between factor movements and trade also in a much shorter run than the above mentioned grand history of global human settlement. Both questions are still debated in the modern literature and will be taken up repeatedly in this chapter. A quick answer to the first question is that immigrants in receiving countries had created a huge demand for capital to work with, reflected in a high marginal return to capital.[12]

The question of complementarity versus substitutability of different forms of economic globalization does not lend itself to a quick answer. Obviously, co-movement of trade and factor flows does not proof complementarity, let alone causality. Indeed, the model of trade that most economists would invoke in order to explain the afore-mentioned income convergence and inequality trends observed during 1870-1910, both between and within different countries of the "Atlantic economy", would suggest substitutability, rather than complementarity between trade and factor flows. This is the

---

(1999) and Hatton and Williamson (2005). To give just a few numbers, the estimated labor force reduction of the two biggest sending countries over the period 1870-1910 was 45 percent for Ireland and 39 percent for Italy, with an estimated positive impact on the real wage for emigrant-competing workers equal to 32 percent and 28 percent, respectively. The labor force increase of the two biggest receiving countries over the same period was 86 percent for Argentina and 42 percent for Australia, with an estimated negative effect on the real wage rate for immigrant-competing workers equal to 21 and 15 percent, 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), chaper 9.

[9]See O'Rourke and Williamson (1999), chapter 12.

[10]See again O'Rourke and Williamson (1999), chapter 3.

[11]These capital flows were early examples of what later became known as the "Lucas paradox". Lucas (1990) was asking this question for the late 20th century, with a view on capital-scarce asian economies like India.

[12]See O'Rourke and Williamson (1999) p. 229f.

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.[13]

This view of trade implies a strong substitutability relationship with factor movements: Unless countries' endowments are too far apart, free trade leads to complete factor price equalization, with no incentive remaining for factors to move.[14] Conversely, even a small import barrier is enough to wipe out all trade, at least in a two-by-two model of the world economy, provided there is complete international mobility of either capital or labor.[15] Hence, not only are factor movements a perfect substitute for trade, but mobility of labor and capital are also substitutes for each other. In a similar vein, and more generally, international factor movements that correspond to the factor content of free trade would allow countries to achieve free trade (i.e., equal) factor prices without any trade in goods.[16]

These theoretical results imply strong policy conclusions. In particular, they imply that restricting immigration in order to avoid unwelcome income distribution effects will be frustrated unless trade is restricted as well. However, the policies observed during the era of mass migration between 1870 and 1910 do not square with this implication. Most of the labor receiving countries were pursuing protectionist trade policies early on, at a time when they were still vastly open to immigration. This certainly does

---

[13]See Feenstra (2004b) for a convenient survey of these proposition also for higher dimensions.

[14]"Too far apart" here means endowment points lying outside the so-called "factor price equalization region". This modern version of the factor price equalization theorem was first stated in Dixit and Norman (1980).

[15]This version of the substitutability result is due to Mundell (1957). The prices of factors and goods in this zero trade equilibrium are the same as under free trade.

[16]On this interpretation of the factor content of trade, see Deardorff and Staiger (1988). For a recent application of this idea, see Krugman (2008).

not reveal policy makers' belief in substitutability between trade and immigration. We shall see below that the same can be said for policies pursued with respect to modern migration in the late 20th century. Moreover, O'Rourke and Williamson (1999) argue against substitutability also on account of a statistical analysis of the relationship between trade and factor movements in the late 19th century.[17]

Given that the data reject the hypothesis of substitutability which is implied by the Heckscher-Ohlin model, may we still accept the Heckscher-Ohlin-type interpretation of the convergence and income distribution trends observed in the first wave of globalization? The answer is a qualified yes: Doing so involves no contradiction, but it does not not amount to an entirely convincing story. To see this, let us briefly turn to some theory.[18] Suppose that trade is due, not to factor endowment differences, but to some other country asymmetry. To be more precise, take a standard two-by-two model and suppose that a country has a Hicks-neutral technological advantage over its trading partner in the labor-intensive sector, and assume that this is the only asymmetry between countries. In any trading equilibrium that satisfies the law of one price on goods markets, this country must have a higher wage rate than its trading partner.[19] If we then allow for factor mobility, an inflow of labor into this country will add a Heckscher-Ohlin rationale for further trade, since the country now becomes a relatively labor abundant economy. The inflow of labor will increase its exports of the labor-intensive good. This is the essence of a result demonstrated by Markusen (1983), who provides the first rigorous treatment of complementarity between trade and factor movements.[20]

At this stage, our interest lies in how factor prices will evolve as this process of

---

[17]See O'Rourke and Williamson (1999), p. 264ff.

[18]We shall return to a theoretical treatment of complementarity in more detail in our theory section below.

[19]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*.

[20]Note the analogy of this complementarity effect to the trade implications of human settlement across the globe that we have mentioned in the introduction to this section. In both cases, migrants end up producing goods which are in excess supply (demand) in their "new (old) country".

complementarity unfolds. As indicated above, the answer is, indeed, factor price convergence. The inflow of foreign labor will eventually wipe out the high-wage-advantage afforded by superior technology in the labor-intensive sector. But complementarity between such migration and trade does make a difference. It has the important consequence that it takes more labor movements to achieve complete international convergence than would be the case without "complementary trade". Indeed, 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. Moreover, an interesting twist of this process is that internal redistribution now does not run against the factor that is scarce in the quantity sense, but scarce in the economic sense of commanding a relatively high price. Thus, observing complementarity between trade and factor movements, as in the literature on the "Atlantic economy", does not rule out a Heckscher-Ohlin interpretation of income convergence that is also observed for this economy.[21] But the implication that factor movements entail "slower convergence" seems somewhat at odds with the strong and quick convergence that the data seem to suggest. In this sense the Heckscher-Ohlin story of convergence coupled with complementarity of trade and factor movements is not entirely convincing.

To summarize, the era of mass migration during 1870-1910 features strong Heckscher-Ohlin-type price effects of migration, but it was not characterized by a substitutability relationship (in the quantity sense) between either migration and capital flows or between migration and trade.

2.2   Characteristics of modern migration

All forms of globalization were hit by a strong backlash in the first half of the 20th 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

---

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

decade of what may be called the second wave of economic globalization, again covering trade, migration and capital flows. Obviously, this second wave of globalization differs in many important respects from the first wave a hundred years ago. Of special interest in the present context, was it similar to the first in terms of the convergence effect and the complementarity of migration, goods trade and capital movements? We shall turn to these questions below, after briefly characterizing some important differences that sets migration of the past 5 decades apart from migration a hundred years ago.

Countries and magnitudes involved   As regards the regional pattern of migration flows, Hatton and Williamson (2005) speak 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 continents of migration, while the Persian Gulf has appeared as a new destination region. 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 lead to significant east-west migration within Europe. At the same time, the magnitudes and patterns of international migration were also importantly shaped by restrictive immigration policies; see below.

It is worth briefly comparing the magnitudes of migration flows. While in the era of mass migration yearly (net) immigration rates close to and above 1 percent of the population were quite common, net immigration rates exceeding 1 percent are rarely observed over the past five decades. Looking at 13 countries classified by the OECD as receiving countries over the period 1956 to 2009, we observe no more than 20 instances (out of 702) where net annual immigration rates were above 1 percent.[22]   However,

---

[22]The one country that has experienced an extended period of annual immigration rates above 1

the receiving countries of the second wave have typically had lower growth rates of the labor force, at least towards the end of the period considered. Hatton and Williamson (2005) calculate 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.

Selection effects The source country composition of migration has changed over time also within both the era of mass migration in the 19th century and the wave of globalization in the late 20th century. Any such change is bound to have implications for the skill level and initial wage earnings of migrants in the host country.[23] Hatton and Williamson (2006) demonstrate that in both waves of globalization this source country composition effect has worked towards a long-run deterioration 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 19th century and 20th century migration with respect to the relative "labor market quality of migrants", but the negative trend was more pronounced in the 20th century than in the 19th century.[24]

For a given source country with a specific distribution of skills and other labor market qualities among potential migrants, and with unrestricted immigration, the pattern of migration flows reflects selection effects in the emigration decision. A positive selection effect occurs, if the mean "labor market quality" leads to a mean wage of those who migrate which exceeds the mean wage of the entire population of potential migrants. This can be calculated either looking at wages of (would be) migrants in the sending country or the receiving country. There is a general presumption that migrants tend to be positively selected, at least if the return to skills is higher in the receiving than in

---

percent was Spain, which, oddly, is listed as an emigration country by the OECD. See the data set underlying OECD (2011), URL: http://dx.doi.org/10.1787/888932446759. Setting the threshold at 0.5 percent, a value almost always surpassed by immigration countries of the 19th century, the number of such cases in the late 20th century increases to 162, which is still less than a quarter. See also Felbermayr and Kohler (2006b).

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

[24]See Hatton and Williamson (2006) who argue that the a long-run trend of a falling "labor market quality" has contributed significantly to anti-immigrant sentiment already in the 19th century.

the sending country, which - arguably - often is the case.[25] This should be reinforced if the receiving country's immigration restrictions favors high-skilled immigrants, which is often true as well. However, empirical evidence on selection effects is mixed, both for the era of mass migration in the 19th century and the late 20th century.[26] Mixed evidence is not too surprising in view of the Roy model of selection, which does not lend itself to an unconditional prediction based only on expected earnings alone. As noted by Borjas (1987), the selection hypothesis always states positive or negative selection, conditional on the variance and covariance structure of the earnings distributions in the receiving and the sending countries, respectively. We shall briefly return to this in the next section.

Restrictive immigration policies   Modern migration is characterized, perhaps more than by anything else, by fears of unwelcome labor market effects in receiving countries. Such fears have developed towards the end of the 19th century as well, particularly in the US, but it was not until the 1920s that immigration restriction were imposed by many countries.[27] 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 (see next section) as well as econometric evidence questioning the widely held belief of wage pressure caused by immigration (see below), a generally restrictive policy stance by destination countries still marks the global migration landscape of today.

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 and to minimize the cost of integration. 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, quanti-

---

[25]See Borjas (1987) and Borjas (1999) for a an analysis based on this standard criterion. For a more elaborate analysis of the emigration decision, see Stark (1993).

[26]See Hatton and Williamson (2005), Hatton and Williamson (2006) for the 19th century and Borjas (1987), Chiswick (1999), Chiswick (2000), Chiquiar and Hanson (2005) and Grogger and Hanson (2011a) for evidence on the late 20th century.

[27]See Hatton and Williamson (2006).

tative restrictions have been binding. Hence, unlike the mass migration of the 19th century, international migration flows of the late 20th century have been influenced 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 19th century it was mostly seen as *em*igration.[28]

The restrictiveness of present day immigration policies is perhaps best realized by looking at the amount of illegal immigration and the cost incurred to overcome or circumvent immigration barriers. The two most important destinations for modern migration are the US and the EU. An estimated 4 percent of the US domestic population (and more than a third of its immigrants) are illegal residents, about 76 percent of them Hispanics.[29] 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 percent of the population.[30] With a total inflow of less than a percent, this is substantial. The cost of overcoming restriction is sizable, too. Between 1993 and 2007, around 7.000 people have died trying to get into the EU.[31] 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 the Ecuador, Mexico and the US have estimated this rent to be around $20 bn per year at the beginning of the past decade.[32] For the EU, the figure is estimated at €4 bn.[33]

---

[28]High and lasting restrictions on immigration stand in marked contrast to progressive liberalization of goods markets, starting soon after World War II through the GATT, and liberalization of capital markets starting after the breakdown of the Bretton woods system in 1973. Today, markets for goods and services as well as capital may truly be called global, whereas labor markets are still national, without much policy commitment to liberalization; see Freeman (2006). This asymmetry was nowhere near as characteristic of the 19th century, and it needs to be borne in mind when thinking about causality or complementarity relationships between migration, trade and factor movements.

[29]See Passel and Cohn (2009) and Hanson (2009). For an economic analysis, see Hanson (2006).

[30]See "The best of reasons", The Economist, Oct 31st 2002.

[31]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), p.29 and34.

[32]See "By a Back Door to the U.S.: A Migrant's Grim Sea Voyage", *The New York Times*, June 14, 2004.

[33]See "Decapitating the snakeheads", *The Economist*, Oct 5th, 2005.

Migration and development   If the significance of immigration restrictions sets the 20th century globalization apart from the era of mass migration, so does the income gap between potential sending and destination countries, respectively. This is mostly due to the fact that almost all countries in the world have now become potential sending countries. In particular, migration is now considered an important element in global development policy, which is a further characteristic of 20th century migration that was absent in the 19th century.

Income gaps 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 data set, Freeman (2006) compares wage gaps within occupations, taking the bottom and top 20 percentage point of the world-wide distribution for the period 1998-2002. Converting to PPP, he obtains bottom-to-top ratios range 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 percent 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 8.8% (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 percent.[34] This seems like an enormous gain, but judged from the above mentioned wage gaps they seem plausible.

However, income gaps may also 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-embodied) capital per worker, and it is these factors that the worker will immediately benefit from once moving. Hendricks (2002) decomposes observed 1990 wage gaps to the US for a large sample of countries. For low

---

[34]See World Bank (2006).

income countries, defined as countries with an observed gap larger than 60 percent, the average of observed country-specific income gaps is 82.3 percent. Differences in countries' physical capital endowments are able to explain a very small part of this gap, viz. 17.6 percentage points.[35] Adding observed worker skills increases the explained 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 $1/(1 - 0.823) = 5.65$).[36] 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 percent 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, then, do the factor price and convergence trends observed during the second wave of globalization compare to the first? We first turn to the effect of trade and migration on relative factor prices within countries. Research on the second wave of globalization has focused on wages of skilled relative to unskilled labor, the so-called skill premium, whereas literature on the 19th century trends in income distribution has focused more on labor income relative to non-labor income. 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,

---

[35]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.

[36]The numbers are from table 1 in Hendricks (2002). For the 5 lowest income countries, the reduction is from $1/0.058 = 17.24$ to $1/(1 - (0.406 + 0.196)) = 2.51$.

in the late 20th century the trend has been observed for almost all countries at the same time.[37] 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. Trade and migration have mostly been analyzed separately and with different methodological approaches.

Trade   There is a voluminous empirical literature that takes up this issue, aiming to quantify the explanatory power of trade and migration for the relative wage trends observed over the past three decades in industrial countries. Given the fundamental change in the nature of trade during the 20th 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.[38]

However, estimating the wage effects of (an increase in) trade is fraught with methodological problems.[39] 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

---

[37]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).

[38]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) and Berman et al. (1998).

[39]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).

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.[40] Moreover, a shift in the structure of vertical specialization, or off-shoring, is equivalent to a change in technology, leading to higher demand for skilled labor. After all, trade is an inherent part of a 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 *world-wide* nature of wage trends.[41] 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 is presently available.[42]

The modern literature proposes several new determinants of trade with different channels for trade to affect the skill premium. Do these channels enhance the role of trade as an explanatory factor for observed trends? We cannot go into details here, but two remarks are worth making.[43] First, the scale effects present in monopolistic competition models may explain the observed world-wide increase in the skilled premium, if the more skill intensive sectors feature a higher degree of such scale effects.[44] And secondly, the skill premium may also rise as a result of selection effects as emphasized by the recent literature emphasizing firm heterogeneity. This may come about as a result of a skill-biased productivity effect, meaning that more productive firms use a more skill-intensive technology. If this is true, then trade liberalization may entail a

---

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

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

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

[43]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).

[44]This has been demonstrated by Epifani and Gancia (2008). The scale effects alluded to in this argument are external scale effects that arise on the industry level, due a technology that has "love of variety" in the use of intermediate inputs.

more skill intensive average technology by virtue of factor reallocation between firms, which in turn derives from the usual selection effects that weed out the least productive firms.[45] As with the first effect, this may be expected to happen world-wide.

Migration This brings us to the potential for migration as an explanatory factor for the late 20th century trend in relative factor prices. Arguably, the wage effects of immigration should be easier to identify, since they do not require factor content calculations. The relationship between immigration and wages is also of more immediate importance for policy than the question of trade and wages, for the simple reason that immigration policy is much less restricted by international agreements than trade policy. Hence, if immigration is the proven culprit for unwelcome distributional trends, then restrictive immigration is the likely response. Moreover, the intuition for this is open to common sense and the theoretical prediction seems much more convincing and robust to start with than is the case for the factor content logic alluded to above. 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.

This may seem fairly straightforward for a single-sector economy, but 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 flat 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 theorem.[46] However, if there are many sectors differing in factor intensities, these flat 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

---

[45]In a recent paper Burstein and Vogel (2012) present an in-depth analysis of this channel, including a calibration exercise in order to quantify the effect. They also propose 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 world-wide trend rise in the skill premium. In their scenario, trade liberalization explains up to 80 percent of this trend.

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

aggregate labor demand function, it has the uneasy implication that the economies would be specialized in on disjoint sets of industries with non-overlapping ranges of factor intensities.[47]

The 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.[48]

Researchers have repeatedly 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, often 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) distinguishes 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 to about 10 percent of the population, has depressed wages paid to the "average" US worker by about 3.5 percent.

---

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

[48]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 which 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 (2009). All of these studies support the aforementioned consensus of a very low impact of immigration on wages earned by natives.

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 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.[49] 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.[50] It should be noted that this type of capital accumulation effect caused by migration may lead to capital imports, in which case migration and capital movements would appear as complements.

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.[51] 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

Did international trade and migration cause as much international income convergence in the second half of the 20th century as they did in the 19th century? We have mentioned above that the convergence effect of trade as well as migration during the 19th century was substantial. Taylor and Williamson (1997) have estimated the convergence effect by applying the measured cumulative migration flows for the various sending and receiving countries of the "Atlantic economy" between 1870 and 1910 to estimated elas-

---

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

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

[51]A recent paper by Aydemir and Borjas (2011) demonstrates 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.

ticities of the labor demand functions. 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 percent. The corresponding increase in Old World sending countries is 9.6 percent.

Suppose 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) point out that, these calculations would imply a migration-induced convergence that exceeds the convergence actually observed. Across all countries considered, measured inequality in 1910 was down to 72 percent of what it was in 1870. Undoing the migration effect through the aforementioned thought experiment takes inequality up to 7 percent *more* than it actually was in 1870. Thus, focusing entirely on migration and assuming a simple downward-sloping aggregate labor demand function thus "overexplains" income convergence.

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 estimates two things that are at the core of this chapter: The first is the immigration countries concerned were not just receiving labor, but capital as well; see above. Moreover, this capital was coming from the emigration countries. And the second is that they were trading economies. Ignoring such capital flows 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).[52]

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 tied down as sharply as in the two-factor-case.[53] 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.

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 (s. 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

---

[52]Taylor and Williamson (1997) provide 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).

[53]A detailed theoretical analysis follows in the subsequent section.

embodied in workers.

Ruling out technology and human capital differences (s. above), 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.[54] 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 which 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.[55]

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 19th century. Notice that in this scenario migration (or more generally factor flows) may, but need not, enhance trade in the sense of the Markusen theorem (s. above), since the reallocation is among disjoint sets of industries where the two countries are specialized. We 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 coun-

---

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

[55]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.

tries 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 19th century, O'Rourke and Williamson (1999) argue 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" 19th 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 unexploited arbitrage opportunities. The particular combination of trade and factor movements through time may reflect the sequence of events caused by 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.

Convergence through modern migration?   Mass migration in the 19th 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" 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 percent of what it had been in 1870. Hatton and Williamson (2005) describe 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 percent, rising to 53 percent by 1910.[56]

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) apply 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 19th 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 19th 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 20th 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

---

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

cut-off points for the bottom and the upper quartile 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* data base (s. above). For 1980 GDP per capita, the 25th percentile is a mere 0.138 of the 75th percentile, and for 2007 it is even less, viz. 0.125. The same calculations for private household expenditure per capita yield values of 0.179 and 0.140, respectively.[57]

2.5 Is modern migration complementary to trade?

We have pointed out above that existing literature on the late 19th century evolution trade and factor movements largely rejects substitutability and to some extent suggests complementarity between trade and factor movements as well as between migration and capital movements.[58] Indeed, the era of mass migration seems like a very short run and condensed form of the type of complementarity, alluded to in the opening paragraph, between migration and other forms of internationalization that has characterized the millennia of global human settlement. Moreover, the policies pursued do not give the impression that policy makers have believed in substitutability. This latter impression is also obtained by looking at policies in the late 20th century, although the policies pursued are markedly different. Then, policy makers were apparently aiming trade protection to avoid income effects that, according to the substitutability hypothesis, should also have been the outcome of the liberal policies pursued with resect to migration. Now, policy makers are pursuing protectionist policies with respect to immigration, while at the same time, at least in rhetoric, they agree on dismantling trade protection that should have the same distributional outcome. How about statistical evidence of complementarity or substitutability of trade and factor movements in the 20th century?

---

[57]We should hasten to add an important remark of qualification. Our focus here is not global inequality, which would require looking at internal distribution of income within countries. The point that we are trying to make here is that enormous income gaps still exist that will serve as powerful incentives for future migration, despite all migration flows that we have observed in the past. Needless to say that 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).

[58]The statistical evidence is reported in O'Rourke and Williamson (1999), chapter 13.

Somewhat surprisingly, statistical evidence on this question is scant; a survey is found in Gaston and Nelson (2011a). The only study explicitly looking at the relationship between migration and trade is Wong (1988). The study estimates a trade utility functions for the US.[59] A trade utility function gives the maximum utility that an economy may obtain through choice of a vector of net exports e, given that consumed quantities plus net exports (i.e., the country's outputs) are feasible, given the economies endowments v, and given the constraint that the value of net exports at world market prices is equal to (or larger) than b. Formally, we have $p \cdot e \geq b$, where p denotes the vector of world market prices and $\cdot$ denotes a scalar product of two vectors. This function may be written as $U(p, b, v)$. By extended versions of Roy's identity, and assuming perfect markets, the gradient of this trade utility function with respect to the price vector p, $U_p(p, b, v)$, may be interpreted as a vector-valued (sectoral) net export demand function, expressing sectoral net exports as functions of world market prices p, the economy's endowment v and the value constraint on net exports, b. By analogy the gradient with respect to the endowment vector v, $U_v(p, b, v)$, gives factor prices as functions of these same variables. Specifying $U(p, b, v)$ in translog form, Wong (1988) derives estimable forms for export share functions, which express net exports in value terms relative to domestic expenditure, and similarly for factor share functions which express factor incomes relative to GNP. His empirical model is highly stylized, containing only three types of goods, durable and non-durable domestic goods (consumed domestically and exported) and an aggregate imported good (not produced domestically). Among other things, empirical estimation yields values of $U_{pv}(p, b, v)$, the derivatives of net export functions with respect to both capital and labor. Applying these estimates to rough estimates of historical inflows of both capital and labor into the US after 1948, which are interpreted as $\mathrm{d}v$, Wong (1988) is then able to estimate the trade effects of these postwar factor movements as $U_{pv}(p, v, u) \cdot \mathrm{d}v$. The results are (i) that labor inflows have increased both exports and imports, while (ii) capital imports have increased exports but have had an ambiguous effect on imports.

We must be very cautious in interpreting this result as indicating complementarity

---

[59]These functions have been introduced by Woodland (1980). The subsequent description follows Neary and Schweinberger (1986).

between trade and migration as defined by Markusen (1983). Testing for this type of complementarity would require to investigate a) whether migration takes place due to trade where no migration would take place without trade, or b) whether more trade induces more migration. In such scenarios, trade would initially take place for reasons unrelated to migration and factor endowments. Alternatively, and in line with the type of complementarity defined in the opening paragraph, one could ask whether additional trade has been caused by migration, which itself is unrelated to trade.

Estimating the derivatives matrix $U_{pv}(p, b, v)$ and then calculating $U_{pv}(p, b, v) \cdot dv$, based on historical observations of dv, certainly does not allow us to answer question a), but it might potentially provide a clue regarding the answer to question b). More specifically, the vector product $U_{pv}(p, b, v) \cdot dv$ gives the change in excess supplies of goods upon a change in endowments dv. In a two-by-two case where the economy has a comparative advantage in the capital intensive good, an inflow of labor would lead to reallocation of both goods to the labor intensive import good, hence the excess supply of the export good should fall, and the excess supply of the import good should increase. In other words, and more generally, the Heckscher-Ohlin theory imposes structure on the matrix $U_{pv}(p, b, v)$.[60] And the structure is such that migration and trade should be revealed as substitutes. Within the model of Wong (1988), however, this argument does not go through, since it assumes perfect specialization on the exported good. Taking this assumption on face value, the implication is that an inflow of labor no longer causes the aforementioned reallocation, but will instead increase output and thus excess supply of the export good, and it will increase (excess) demand of the import good. Moreover, even for unchanged goods prices, the inflow of labor now causes a reduction in the wage rate, relative to the capital rental. The upshot is that labor inflow into a labor scarce economy that is completely specialized on the capital intensive good will enhance trade. Thus, migration and trade are complements, as long as the economy remains specialized. This is *one* possible interpretation of the complementarity finding in Wong (1988). However, once the economy becomes diversified because cheaper domestic labor makes the labor intensive import good domestically viable, Heckscher-

---

[60]See Dixit and Woodland (1982), Deardorff (1982) and Svensson (1984) for general formulations of the factor proportions theory that involve the equivalent of $U_{pv}(p, b, v) \cdot dv$.

Ohlin-type substitutability will start to prevail.

What, then, are we to conclude from this? Given the state of theoretical knowledge about the relationship between trade and migration, it seems futile to test these hypotheses in an unconditional way. We know from Markusen (1983) that trade caused by non-Heckscher-Ohlin determinants may cause (or increase) factor price differences, thus causing (or increasing) factor movements, given that factors are internationally mobile. On the other hand, we know from Mundell (1957) that trade barriers may give rise to lower trade and enhanced factor movements. These are the two traditional examples of specific hypotheses where trade and migration (or more generally factor movements) may be complements and substitutes, respectively. To be informative, empirical research on the relationship between trade and factor movements should address such specific hypotheses, as opposed to answering the questions about co-movements of the volume of trade and factor movements. We shall return to a specific modern hypothesis about complementarity between migration and trade in section 4 below.

3   A neoclassical view on migration, capital flows and trade

In this section we want to briefly summarize the key messages of neoclassical theory about migration and its relationship to capital flows and trade. For the present purpose, by neoclassical theory we essentially mean models of trade and factor movements that (i) assume a constant returns to scale technology and (ii) markets with "moderate degrees" of market imperfections, and (iii) describe full employment equilibria. We shall, however, subsequently relax the assumption of constant returns to scale, where it will also become evident what, exactly, we mean by "moderate degrees of market imperfections". Importantly, we assume that international movements of goods, labor and capital are responsive to prices, essentially arbitraging on price differences up to an international equilibrium where relevant price differences between countries are reduced to the costs of moving (goods, capital or labor) across borders, plus the price equivalent of policy induced barriers to such movements. In this section, we shall mainly focus on

stocks and flows of international migration, assuming free and costless trade between the countries involved. We assume that changes in stocks of migrants are driven by changes in political barriers to, or in the costs of, migration, but these changes are assumed exogenous and do not dependend on the amount of trade or movements of other factors. Conversely, assuming free and costless trade to start with, such changes in migrant stocks also do not affect trade barriers. This channel will be taken up separately in the subsequent section.

There are two important differences between international migration and international capital flows. First, migration necessarily involves movement of people. The same is not true for capital movements, which typically involve only a relocation of the factor input without movement of the factor owner. And secondly, capital movements are inherently related to capital accumulation. Capital very rarely moves in the sense of a relocation of *existing* physical capital. Instead, it moves in the sense of *new* capital being invested abroad. In contrast, with international migration it is always existing labor that is being relocated (in combination with movement of people). Hence it may and should be analyzed independently of accumulation of labor, i.e., population growth. In this section, we first discuss migration using static models. In section 4, we shall then analyze migration in dynamic models focusing on capital accumulation.

3.1   A normative view on migration

A central tenet of neoclassical theory is that factor movements driven by international differences in factor returns increases 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.

3.1.1 A simple, yet general model

We now use a very general, yet simple model that allows us to present the key normative messages of neoclassical theory on international migration. It is very general in that we allow for an (almost) arbitrary number of goods and factors and in that we allow for trade as well as factor movements. Moreover, it allows for two large countries, thus explicitly looking at the receiving and the sending country, and allowing for factor flows in both directions. Including several types of labor that may simultaneously move between two countries is important, since it allows us do address attempted selectiveness in migration policy. The model is simple in what we do in terms of comparative static analysis. We do not explicitly solve for the changes in any of the variables involved, but largely confine ourselves to results in terms of inequalities that may be interpreted as correlations of comparative static changes over goods and factors.

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. In addition to allowing for movements simultaneously of different types of labor and in both directions between the two countries, we emphasize the distinction between existing *stocks* of

migrants and *flows* of migration that add to these stocks.

In the following, we generally assume that the number of traded goods is larger than the number of primary factors. For simplicity, we first assume that both countries share the same technology which is described by a GDP-function, defined as

$$G(p, v) := \max_q \{p \cdot q, \text{ s.t. } (q, v) \in T(q, v)\} \tag{1}$$

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 economy's technological knowledge. Constant returns to scale imply a convex technology set. Moreover, we assume convex preferences, that may be 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 labor (i.e., 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, allowing for a meaningful discussion of migration, 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.[61] The implication then is that the two 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.

Let us assume that $v_A = V_A - m_{AB} + m_{BA}$, where V^A is the stock of factors owned by natives of country A, and the vector m_{AB} denotes 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. The vector V_A includes all factors, including non-labor factors, so that m_{BA} and 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 $Y_B = G(p, v_B) - w_B \cdot m_{AB} + w_A \cdot m_{BA}$, respectively. We assume V_A and V_B to be given, and we look at variations in the migrant stocks, i.e., migration flows $\mathrm{d}m_{AB}$ and $\mathrm{d}m_{BA}$.

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 6, 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 $\mathrm{d}m_{AB}$ and $\mathrm{d}m_{BA}$ 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

---

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

Figure 6. Wage arbitrage condition on migrant stocks and migration flows
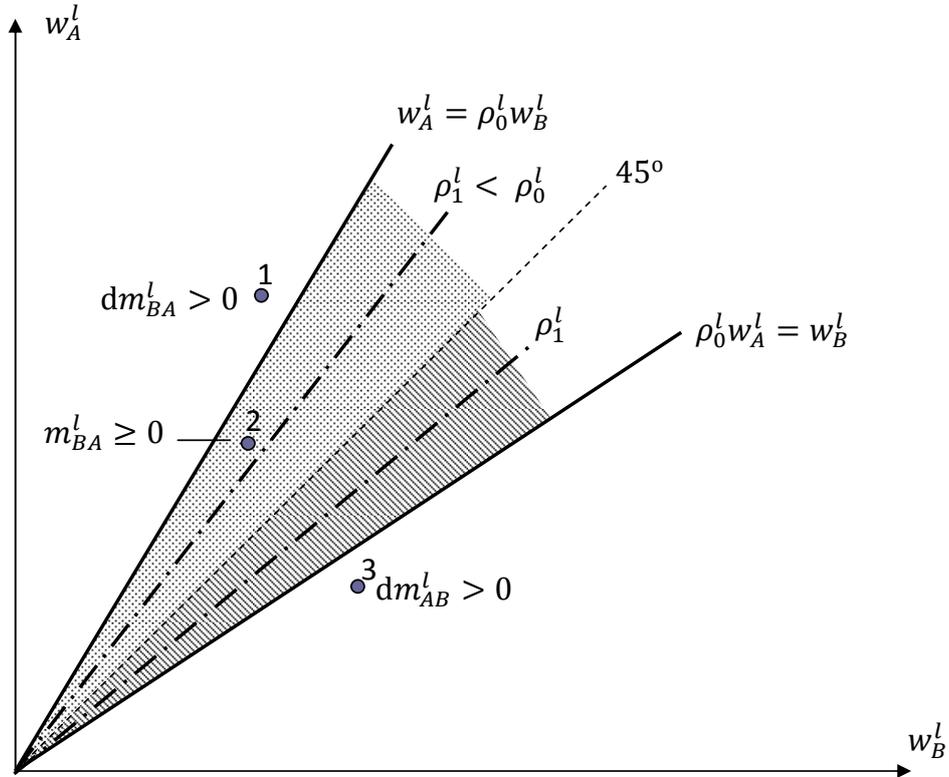
the ρ's represent the cost-equivalents of such restrictions, in addition to the costs of migration. Against the backdrop of this interpretation of ρ, we assume that migration flows for all types of labor are determined separately from each other as indicated in figure 6.

What can we say about the relationship between the ρ's 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 equilibrium, and assuming a stock $m^l_{BA} > 0$ to start with, a reduction from ρ^l_0 to $\rho^l_1 < \rho^l_0$ 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 $m^l_{AB} > 0$. 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 $m^l_{AB} > 0$, then a reduction from ρ^l_0 to ρ^l_1 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^l_A and w^l_B 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 6, 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 6. 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 $dU_A = H_p^A(p, Y_A) \cdot dp + H_Y^A dY_A$, with $dY_A = G_p(p, v_A) \cdot dp + G_v(p, v_A) \cdot dm_A - d(m_{BA} \cdot w_A) + d(m_{AB} \cdot w_B)$, and accordingly for country B. After some straightforward manipulation, we arrive at the following two equations describing welfare effects of *flow* migrations dm_{AB} and dm_{BA}:[62]

$$dU_A = G_v(p, v_A) \cdot dm_A - d(m_{BA} \cdot w_A) + d(m_{AB} \cdot w_B) + M_A \cdot dp \qquad (2a)$$

$$dU_B = -G_v(p, v_B) \cdot dm_B + d(m_{BA} \cdot w_A) - d(m_{AB} \cdot w_B) + M_B \cdot dp \qquad (2b)$$

In the first of these equations M_A denotes the vector of net commodity imports by natives of country A. Formally, $M_A := H_p^A(p, Y_A) - G_p(p, v_A)$, where $H_p^A(p, Y_A)$ 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 world-wide goods market equilibrium implies $M_A = -M_B$. The vector of goods price changes, dp, 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

$$dU_A = (w_B - w_A) \cdot dm_{AB} - m_{BA} \cdot dw_A + m_{AB} \cdot dw_B + M_A \cdot dp \qquad (3a)$$

$$dU_B = (w_A - w_B) \cdot dm_{BA} - m_{AB} \cdot dw_B + m_{BA} \cdot dw_A - M_A \cdot dp \qquad (3b)$$

These equations reveal three principle channels for welfare effects of migration. The first terms tell us that either country may derive a benefit from flow migration, if labor outflows are correlated, across different types of labor, with the initial differences in

---

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

factor rewards. This means that 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. Adding up the first terms in (3a) and (3b) we obtain a *world efficiency gain* equal to $(w_B - w_A) \cdot \mathrm{d}m_{AB} + (w_A - w_B) \cdot \mathrm{d}m_{BA}$. 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 (3a) and (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, $\mathrm{d}w_A$ and $\mathrm{d}w_B$. This effect essentially is a *"terms of trade effect"* that operates on existing *migrant stocks*. The effect is *second order* in nature, arising only for if the pre-existing migrant stocks are non-zero, $m_A \neq 0$.[63] Moreover, like all terms of trade effects, it cannot be positive for both countries at the same time, which is directly evident from (3a) and (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 world-wide 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, $\mathrm{d}p$, 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 goods market equilibrium and hence some goods price adjustment in any migration scenario. Generally, positive

---

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

a terms of trade effect in (3a) is likely to arise, if the migration flow $dm_A = -dm_B$ leads to a world-wide increase in supply of goods where *natives* of country A are net exporters.[64]

A somewhat less demanding way to obtain further insights is pursued in Felbermayr and Kohler (2007), where 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 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 dw_A and dw_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 no-traded goods, i.e., the net exchange between natives and resident migrants of non-traded goods; see again Felbermayr and Kohler (2007).[65] It is difficult to evaluate the relative importance of efficiency effects of migration that do not involve direct international conflict from the two types of terms of trade effects, which 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.[66]

3.1.3 The immigration surplus

Can we say more about the conditions under which a certain pattern of flow migration dm_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.

---

[64]Davis and Weinstein (2002) take a perspective on US immigration where the opposite is the case. They assume 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 calculate a close to 1 percent 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.

[65]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).

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

The wage effects from $\mathrm{d}m_A$ are given by

$$\mathrm{d}w_A = G_{vv}(p, v_A) \cdot \mathrm{d}m_A \quad \text{and} \quad \mathrm{d}w_B = -G_{vv}(p, v_B) \cdot \mathrm{d}m_A \quad (4)$$

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

$$\mathrm{d}m_A \cdot (\mathrm{d}w_A - \mathrm{d}w_B) \leq 0 \quad (5)$$

To fix ideas, let us focus on the third terms on the right-hand side of (3a) and (3b). They capture what is usually called the "immigration surplus" in the literature.[67] 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 (3a) and (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 (3a) and (3b) that both countries 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 $m_A \cdot (\mathrm{d}w_A - \mathrm{d}w_B) = -m_{BA} \cdot \mathrm{d}w_A + m_{AB} \cdot \mathrm{d}w_B$, 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 (4), a non-negative "immigration surplus" for country A will arise, if and only if

$$-m_{BA} \cdot G_{vv}(p, v_A) \cdot \mathrm{d}m_{BA} + m_{AB} \cdot G_{vv}(p, v_B) \cdot \mathrm{d}m_{AB} \geq 0 \quad (6)$$

Applying logic presented in Felbermayr and Kohler (2007) we may state a simple

---

[67]Borjas (1999) presents 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).

sufficient condition for this, which is that the migration flows $\mathrm{d}m_A$ satisfy $\mathrm{d}m_{BA} = \zeta m_{BA}$, 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 semi-definite. More specifically, under the aforementioned condition we then have $-m_{BA} \cdot G_{vv}(p, v_A) \cdot \mathrm{d}m_{BA} = -\zeta \left[m_{BA} \cdot G_{vv}(p, v_A) \cdot m_{BA} \right] \geq 0$ and $m_{AB} \cdot G_{vv}(p, v_B) \cdot \mathrm{d}m_{AB} = -\xi \left[m_{AB} \cdot G_{vv}(p, v_B) \cdot m_{AB} \right] \geq 0$, which leads to (6).

It follows from the above that a country may derive a strictly positive surplus from proportionally *reducing* its stock of *em*igrants (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 (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 GDP-function is "flat" in the direction of the relevant changes, $\mathrm{d}m_{AB}$ and $\mathrm{d}m_{BA}$. What is the economic interpretation? There are at least two interpretations. One has to do with the possibility of Rybczynski-type reallocations which allow the economy to absorb a change in factor supplies, such that marginal productivities of all factors remain unchanged. This possibility is well known from trade theory. Whether it becomes reality in a given migration scenario depends on two things. The first is which of the factors (if any) are mobile between sectors. And the second is the degree of diversification in

the economy's production. Notice that it is not mobility of labor alone that is at issue here, but mobility of all factors. Generally, the more diversified an economy's production, the more likely the possibility of a Rybczynski-type reallocation. However, under the present assumptions with endogenous determination of goods prices, and with different factor prices and cones of diversification to start with, a pure Rybczynski-type adjustment is ruled out, as we shall see in more detail when dealing with the question of convergence below.

The second interpretation is that a change in wage rates are simply negated by policy. An interesting case in point is analyzed in Davis (1998). He looks at two economies, one being the US the other the EU. He assumes only two factors, labor and capital. While the US has flexible labor markets, the EU is assumed to have a binding minimum wage rate. 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 by a third country, say Mexico. Normally, with the US being a large economy, any absorption of immigration at constant wage rates 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, 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. To put it very simply, the EU experiences a compensating "Rybczynski" reallocation which is associated with a move of EU labor into unemployment.

A variant of this result arises for a small country which is open to the world capital market. To see this, let us for the moment stick to the simple two-by-two model with capital and labor as the two factors of production. The economy now faces a given factor return to capital. This need not be the same as the world capital rental. There may be barriers to capital, introducing 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.[68] 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.[69]

3.1.4 Comparing migration policy to trade policy

In this subsection, we want to briefly discuss the relationship between this normative treatment (6) of migration and the gains from trade theorem. The gains from trade theorem states that under conditions of neoclassical trade theory, absent distortions, welfare of any country participating in free trade is at least as large as under autarky, assuming the availability of a costless compensation mechanism.[70] It thus looks at a *discrete* jump from autarky to free trade, whereby trade flows are governed by comparative advantage and the law of one price for all traded goods. In contrast, the above treatment of migration looks at *marginal* migration flows dm_{AB} and dm_{BA} which are not governed by the law of one price for all types of labor, but by variations on the costs of migration ρ^l (see above). Thus, the goods trade analogue to the calculus around inequality (6) is not found in the gains from trade theorem, but in the results on piecemeal trade policy reform. These results compare two equilibria, both distorted by trade barriers, and connected to each other by some change in trade barriers. Unlike the gains from trade theorem, and much like the above analysis, this involves second

---

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

[69]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.

[70]Importantly, this compensation mechanism need not bee of a strictly lump-sum nature. A suitable commodity/factor tax cum subsidy scheme will do the trick as well. See Dixit and Norman (1980) and Dixit and Norman (1986).

best equilibria and is therefore less straightforward.

This not the place to go into the theory of piecemeal trade policy reform, but the fundamental analogy is worth pointing out. Thus, Feenstra (2004b) presents a result on the welfare effect of trade policy reform which is formally analogous to (3) above, except that it is formulated in discrete changes. As with the above analysis, there is a terms of trade effect and a first-order efficiency effect. However, in the above analysis, the terms of trade effect comes in two forms, one relating to trade flows, the other relating to preexisting stocks of migrants, i.e., the immigration surplus. In the trade policy analysis, the efficiency effect comes in the form of a change in the revenues from trade policy that comes about through the reform-induced change in trade flows, evaluated at post-reform trade barriers. In this effect, post-reform barriers measure the distortion remaining after reform, and the effect is positive if on average imports of goods with high remaining barriers have increased. In our analysis, this type of effect appears in the first term of (3), where the vector $w_A - w_B$ similarly measures the degree of remaining distortions in an equilibrium where no *further* migration flows take place.[71]

The key question from a trade policy perspective is how to design the reform in trade barriers that ensures a welfare improvement, given that a jump to free trade is impossible. The literature has typically analyzed by looking at certain effects in isolation. For instance, for small economies, it is relatively straightforward to show that a proportional reduction in all barriers will do the trick. This is a classic result of welfare economics.[72] Another oft-quoted rule would be to first go for a reduction of the highest tariff, which is referred to as a "concertina tariff rule".[73] In some sense the above result that an equiproportional reduction of all immigrant and emigrant stocks will lead to a positive immigration surplus resembles the proportional tariff cut rule, but it looks at the terms of trade effect, not the first-order efficiency effect.

In general, a welfare enhancing piecemeal policy reform seems more difficult to identify for migration than for trade. The relevant perspective for modern migration

---

[71]In other words, the wages in $w_A - w_B$ are wages lying in the "cone of no migration flows" depicted in figure 6 above.

[72]See, for instance, Dixit and Norman (1980) and Vousden (1990).

[73]See for instance Neary (1998).

seems to be that of an immigration country being able to impose selective quantitative restrictions on immigration. On the one hand, this seems an advantage for the policy makers, as it affords a direct handle on migration flows. In contrast, trade policy is typically restricted to changes in tariff barriers, with a potentially complex reaction of demand and supply which seems impossible to predict. At the same time, however, these reactions are not arbitrary and the optimization of producers and households impose convenient restrictions on these reactions that, in turn, allow us to identify certain policy changes that should be welfare improving; see above. A policy maker intent on using selective immigration restrictions in pursuit of native welfare is facing the need of possessing reliable knowledge about the structural characteristics behind the wage and price reactions attendant upon the immigration flows determined by policy. This seems a much more arduous task than designing welfare-increasing trade policy reforms.

This is aggravated by two types of asymmetries. First, given quantitative restrictions on migration, trade policy does not face the need to worry about the migration effects of trade policy. In contrast, given that restrictions on trade, if present at all, are typically not of a quantitative nature, migration policy does face the challenge of estimating the trade effects of migration policy. And secondly, the distributional consequences of any change in migration policy are more difficult to deal with than distributional effects of trade policy. This is due to the fact, already emphasized several times above, that migration involves movement of people. If constitutional or civil rights constraints rule out discrimination in the application of a compensation mechanism based only on a person's immigrant status, any compensation of domestic workers who are hurt by immigration implies compensation also of immigrants, including newly arriving immigrants who have are not suffering a loss, but are indeed the main beneficiaries of migration. It seems questionable whether the efficiency gains from migration are large enough to compensate all losers if political constraints means that this implies subsidizing the winners. In Felbermayr and Kohler (2009) it is shown that this not possible using a tax-cum-subsidy mechanism of the type suggested by Dixit and Norman (1986) that allows for a Pareto improving compensation in efficiency enhancing trade liberalization scenarios. In this sense, trade appears as a politically more benign

form of globalization than international migration.

3.2 Distortions and policy

We now use this model to consider *distortions*. Let us thus assume that we have $w_A = \Omega_A \cdot G_v(p, v_A)$, where the diagonal matrix Ω_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. It is important to recognize that Ω captures any deviation from the envelope property of the GDP-function stating that the factor supply gradient of the GDP-function is equal to the equilibrium factor prices. The institutional interpretation of this 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 the quantity of good i produced has an effect on GDP which 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 $G_{pv}(p, v_A) \cdot dv$. Absent distortions, the effects of variations in inputs on the overall value of outputs is 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 spill-over effect of employment of labor in any one firm to the marginal productivity of the same type of labor employment in others. Positive spill-overs may arise, for instance, from human

capital embodied in certain types of labor, as suggested by Lucas (1988) and applied in traditional way to the immigration surplus in Wong (1995). Such spill-overs would be reflected in values below unity in the diagonal matrices Ω. In principle, there may also be negative external effects implying that wages are above the true marginal value productivities. This then leads toe values above unity in the Ω-matrices. Importantly, these are all 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 Ω. 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 flow migration now change to:

$$
\begin{aligned}
\mathrm{d}U_A &= (w_B - \Omega_A^{-1} \cdot w_A) \cdot \mathrm{d}m_{AB} + (\Omega_A^{-1} - \mathrm{I}) \cdot w_A \cdot \mathrm{d}m_{BA} \\
&\quad -m_{BA} \cdot \mathrm{d}w_A + m_{AB} \cdot \mathrm{d}w_B + M_A \cdot \mathrm{d}p \qquad (7a)\\
\mathrm{d}U_B &= (w_A - \Omega_B^{-1} \cdot w_B) \cdot \mathrm{d}m_{BA} + (\Omega_B^{-1} - \mathrm{I}) \cdot w_B \cdot \mathrm{d}m_{AB} \\
&\quad -m_{AB} \cdot \mathrm{d}w_B + m_{BA} \cdot \mathrm{d}w_A - M_A \cdot \mathrm{d}p \qquad (7b)
\end{aligned}
$$

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 of course be affected too.[74] Consider the first-order effects in turn. The first term in (7a) gives the net aggregate income effect accruing to country

---

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

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 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 $(\Omega_A^{-1} - \mathrm{I}) \cdot w_A$. The first two terms in (7b) are interpreted accordingly.

The new first-order efficiency terms appearing in (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 spill-over 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 fix ideas, for the time being let us assume 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:

$$\mathrm{d}U_A = (w_B^k - \omega_A^k w_A^k)\mathrm{d}m_{AB}^k + (\omega_A^l - 1)w_A^l\mathrm{d}m_{BA}^l \qquad (8a)$$

$$\mathrm{d}U_B = (w_A^l - \omega_B^l w_B^l)\mathrm{d}m_{BA}^l + (\omega_B^k - 1)w_B^k\mathrm{d}m_{AB}^k \qquad (8b)$$

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 $\left[\omega_B^k w_B^k - \omega_A^k w_A^k\right]\mathrm{d}m_{AB}^k + \left[\omega_A^l w_A^l - \omega_B^l w_B\right]\mathrm{d}m_{BA}^l$. More generally, this effect may be

written as

$$dU_A + dU_B = \left(w_B \cdot \Omega_B^{-1} - w_A \cdot \Omega_A^{-1} \right) \cdot dm_{AB} + \left(w_A \cdot \Omega_A^{-1} - w_B \cdot \Omega_B^{-1} \right) \cdot dm_{BA} \quad (9)$$

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 dm_{AB} and dm_{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 (7) capture in very general terms 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 spill-over effects, in addition to being less likely to draw on welfare state budgets. At the same time, however, if such spill-over 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.[75] In addition to selective migration policies, the pattern of migration flows regarding the importance of such spill-over effects will also be determined by the selection effects in the emigration decisions that we have briefly addressed in the previous section.

---

[75]Naturally, the above exposition cannot do justice to the many ideas that have been developed relating to the brain drain. Early papers on the brain drain are found in Bhawgwati (1983). For a treatment focusing on the economics of education, see Miyagiwa (1991). This literature typically departs from the present assumption that the migrants' well being is part of the sending country's welfare. It focuses on welfare of the sending country's remaining population. Furthermore, it is important to note in this context that endogenizing human capital formation opens up the possibility that externalities in human capital formation (as opposed to externalities in human capital inputs in production) generate a positive effect from anticipated emigration of high skilled labor. This has become known under "brain gain"; see Stark et al. (1997a) and Mountford (1997a). We shall return to the issue of brain drain in section 5 below.

3.3 Complementarity versus substitutability

In section 2 above, when comparing the two big waves of migration in recent history, we have discussed the issue of whether international factor movements are a substitute for international trade and whether migration is a substitute for international capital movements. We have emphasized that the factor endowments (Heckscher-Ohlin) view of comparative advantage strongly suggests substitutability between trade and factor movements, whereas a closer look at the data as well as the policies pursued suggests complementarity, rather than substitutability. It is now time, against the backdrop of the model developed above, to get a somewhat firmer theoretical handle on these concepts.

Broadly speaking, complementarity of substitutability may be defined with respect to quantities or prices.[76] We first turn to the quantity side. Suppose, for instance, that a reduction in the costs of migration for some type of labor, ρ^l, leads to enhanced migration from A to B. If, holding all other barriers to factor movements constant, this leads to enhanced migration in the same (opposite) direction of some other type of labor, say k, or of capital, then we may say that movements of these these two types of factors are complementary (substitutes for each other). In our general model, and with a view on policy, this definition does not seem to lead us very far. Of course the definition could be generalized to an average relationship across all types of labor, or all factors more generally. But still, knowing that starting out from some specific general equilibrium, some factors are complements and others are substitutes does not tell us a lot. The same holds for the quantity view on complementarity or substitutability between migration of any one type of factor and goods trade.

Economically, a more interesting and useful definition of complementarity or substitutability looks at the price side. Suppose, then, that a certain type of migration, say again of factor l, has a certain effect on factor prices in the sending and the receiving country, respectively, holding the cross-border stocks of a certain (sub)set of other type of labor (or, more generally, of factors) constant, say for reasons of a quantitative

---

[76]See Gaston and Nelson (2011b) for a similar discussion based on the two-dimensional Heckscher-Ohlin model. The present discussion reveals that, like almost all issues, complementarity and substitutability between factor movements and trade is much more difficult to pin down in models with many factors and goods.

restriction on factor flows. To make the argument a bit more specific, suppose that this factor price effect can in some sense be called an international factor price convergence, with the implication of an increase in the world-wide efficiency of factor use. We may then ask whether allowing for mobility of this (sub)set of other factors would reinforce this convergence, quite irrespective of the direction of factor flows. If it does, then we might call these factor movements complementary. If convergence is partly undone, then we might call them substitutes. It is relatively straightforward to apply this type of definition to the relationship between movements of any one factor and goods trade.

In the neoclassical model, the strongest form of substitutability in the quantity sense is represented by the factor price equalization theorem of the Heckscher-Ohlin model which states that free trade between countries sharing the same technology leads to complete factor price equalization, with no incentive remaining for factors to move, unless countries' endowments lie too far apart.[77] Conversely, even a small import barrier is enough to wipe out all trade, at least in a two-by-two model of the world economy, provided there is complete international mobility of capital, the underlying assumption again being an identical technology in both countries.[78] Since the same holds true also for labor mobility, this also illustrates a further implication of the factor price equalization theorem, viz. that mobility of labor and capital are substitutes for each other. Notice that this substitutability result relies on perfect mobility of factors, assuming zero costs of, and barriers to, factor movements. This implies equal factor returns in equilibrium, whereas our view of migration in this section assumes labor movements to be governed by migration costs.

In a similar vein, and more generally, international factor movements that correspond to the factor content of free trade would allow countries to achieve free trade (i.e., equal) factor prices without any trade in goods.[79] A weaker form of substitutability might be expected to hold in the sense of trade liberalization leading to factor price

---

[77]We may repeat that "too far apart" here means endowment points lying outside the so-called "factor price equalization region". This modern version of the factor price equalization theorem was first stated in Dixit and Norman (1980).

[78]This version of the substitutability result is due to Mundell (1957). The prices of factors and goods in this zero trade equilibrium are the same as under free trade.

[79]On this interpretation of the factor content of trade, see Deardorff and Staiger (1988). For a recent application of this idea, see Krugman (2008).

convergence, thus lowering the incentives for factor movements. And conversely, factor movements might similarly be expected to lead to factor price convergence and thus to reduce the case for trade based on differences in factor prices which in turn lead to differences in goods prices. As we shall see below, however, such convergence is questionable in the likely environment with many goods and factors.

In an environment like the above model where one considers small changes in costs of factor movements, we would also obtain substitutability in the price sense as defined above. To see this, take the simple two-by-two model with capital and labor. Suppose we have an arbitrage-free equilibrium with factor price differences across countries supported by costs of factor movements. Now, if lower migration costs lead to inward migration into the capital abundant economy, then - holding ruling out capital movements as in the above definition - this would lower the wage rate and increase the capital rental in the capital abundant economy. If we now allow for capital flows this higher capital rental would lead to capital inflows as well, and hence would partly undo the migration-induced capital movement.

It is probably fair to say that this view of substitutability is the dominating view in neoclassical theory of trade and factor movements. The pivotal assumption of this view is identical technology across countries. In the above analysis of migration, we followed this view. We have justified international differences in wage rates, a key precondition of migration, by invoking the modern version of the factor price equalization theorem, which states that a free trade equilibrium will feature different factor prices, if countries sharing the same technology have relative endowments that lie sufficiently far apart.

We do not take a firm stance on whether assuming identical technology for the present purpose is a useful assumption. But it is certainly worth exploring the implications of technological differences between countries for the question of complementarity versus substitutability between trade and factor movements. As we have seen in the previous section, such differences are an important possible source of complementarity between trade and factor movements. This has first been pointed out by Markusen (1983), who analyzes several alternative views of trade leading to complementarity.[80]

---

[80]Wong (1986) compares cases with complete mobility of goods only, mobility of capital only, and mobility of both goods and capital. He identifies conditions under which the all-encompassing mobility

Generally, complementarity arises whenever countries feature differences in structural characteristics other than different factor endowments, such that in a free trade equilibrium with strictly national factor markets countries will exhibit different factor prices. Under certain conditions, these factor price differences are larger than under autarky, whence we may say that trade has caused an incentive for factor movements. Allowing for factor mobility will then lead to factor movements which, in turn, adds an endowment-based rationale for trade. Hence, we may say that factor movements have caused further trade.

Arguably, among the possible sources of complementarity, international differences in technology are the most obvious and interesting case.[81] Thus, suppose that in the above model we have M factors and $N > M$ goods, and assume that α^i denotes a factor of Hicks-neutral technological superiority of country A, relative to country B, in producing good i. Denote by $c^i(w_B)$ the minimum unit-cost function of country B, which implies that the minimum unit-cost function for A is $c^i(w_A)/\alpha^i$. We first look at an equilibrium with free trade in goods, but without any factor trade, meaning in particular that there are no migrant stocks m_{AB} and m_{BA}. In such an equilibrium we have

$$p^i \alpha^i = c^i(w_A) \quad \text{and} \quad p^i = c^i(w_B) \tag{10}$$

for any good i which is produced in positive amounts in both countries. Due to free trade, all goods will have the same price in both countries. Equations (10) are standard zero profit equilibrium conditions. Let us assume that there are at least M goods produced jointly in both countries. We use $\bar{p}$ to denote a vector of prices for these goods. Without loss of generality, we may scale units such that $\bar{p}^i = 1$, whence a free

involves both, more trade as well as more capital movements than the corresponding limited mobility case. He calls this complementarity of goods trade and factor movements. The specific situations identified are variations of the cases identified by Markusen (1983).

[81]Svensson (1984) addresses complementarity in a different context where countries are assumed to share the same technology and where all trade is determined by endowment differences. But the notion of trade is extended, by complete analogy to trade in goods, to a subset of factors, while the remaining factors are non-traded. In such an environment, he then perturbates symmetric endowments of the two economies of the world, always making one country more abundantly endowed with the traded factors and the other more abundantly endowed with the non-traded factors. He then finds that in such a scenario an economy may have both more trade in goods as well as more trade in factors. However, in the present context this approach does not reveal any useful insights.

trade equilibrium is characterized by

$$\bar{\alpha} = \bar{c}(w_B) - \bar{c}(w_B) \tag{11}$$

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

We now write $c(w_A) - c(w_B)$ as a linear approximation

$$c(w_A) - c(w_B) \approx c_w(w_B) \cdot (w_A - w_B) \tag{12}$$

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 $\bar{\alpha} \approx c_w(w_B) \cdot (w_A - w_B)$. Observing that $\bar{\alpha} \cdot \bar{\alpha} > 0$, using (11) as well as (12), we may write[82]

$$\bar{\alpha} \cdot c_w(w_B) \cdot (w_A - w_B) > 0 \tag{13}$$

This has a straightforward meaning: In an equilibrium with free trade in goods but no factor trade, country A will on average have higher wages than country B for those types of labor which are intensively used in the production of goods where it has a strong technological advantage over country B. This is an alternative justification and interpretation of 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 $\bar{\alpha}$ through the factor intensity matrix $c_w(w_B)$.[83] Complemen-

---

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

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

tarity between trade and factor movements may now be examined by comparing the trade volumes of this free trade equilibrium with those of an equilibrium that features factor trade in addition to goods trade. Markusen and Svensson (1985) provide such a comparison, assuming mobility of a subset of factors. In their case 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 $\bar{\alpha}$ 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).

What about complementarity versus substitutability between movements of different factors? Neoclassical theory does *not* generally state that movements of different factors (say capital and labor, or different types of labor) are substitutes. There are two reasons for this. The first is that it depends on the underlying cause of the initial international gap in factor prices that gives rise to factor movements to start with. To see this, assume that there are only two factors, labor and capital. Under neoclassical conditions the inflow of labor weakly increases the marginal productivity of capital. "Weakly" simply means that under certain conditions internal factor reallocation may allow absorption of the inflow without any change in marginal productivities (factor prices) at all.[84] But suppose, for the sake of our argument, that the capital rental does rise as a consequence of immigration. What this means for capital movements depends on whether the positive wage gap that lies behind immigration is due to technological superiority, or to relative scarcity of labor (with identical technology in both countries) in the labor receiving country. In the former case, there will be inflows of both labor and capital, hence complementarity of labor and capital in production also means that labor movements are complementary to capital movements. In the latter case the two

---

[84]This was referred to as "factor price insensitivity" above.

types of factor movements are substitutes. Essentially, this distinction follows the same logic as the analysis of Markusen (1983) and Markusen and Svensson (1985) vis à vis the case of substitutability in models of endowment-based trade.

The second reason has to do with dimensionality. As is evident from above, the standard predictions of neoclassical trade theory do not hold in their usual forms, if the number of goods and/or factors exceeds 2. The predictions need to be weakened, and one way to do so is to resort to inequalities. For instance, in the general case inequalities (4) replaces the strong result, obtaining with two factors only, that an inflow of one factor with a constant stock of the other weakly increases the marginal productivity of the other factor.[85] Thus, with many factors it is perfectly possible that the return to some factor (say high-skilled native workers) is negatively affected by an inflow (and increased employment) of some other factor (say low-skilled labor).[86] Indeed, it is for this reason that in general setups all complementarity/substitutability results must be expressed in weak inequality form, as in Svensson (1984), Markusen and Svensson (1985) and Ethier and Svensson (1986). The usual interpretations of these inequalities is to state that relationships no longer hold for any pair of goods and/or factors, but only on average across goods and/or factors. In this sense, then, we may say that neoclassical theory predicts that substitutability obtains on average for factor movements caused by endowment ratios, while complementarity obtains on average for movements caused by technological superiority/inferiority. Notice, however, that looking at factors instead of goods the relevant concept is Harrod-neutral as opposed to Hicks-neutral technology gaps.

Essentially the same qualifications apply for a further fundamental message of neo-classical theory, viz. that a labor inflow should enhance accumulation of factors. It does so only to the extent that the factor accumulated is complementary to labor, so

---

[85]The "factor price insensitivity" result mentioned above is covered by the case of equality in the weak inequality (5). Insensitivity, $dw = 0$, obtains if the economy is sufficiently diversified, so as to absorb the change in factor endowments through Rybczynski reallocations without leaving the cone of diversification. Obviously, factor price insensitivity also obtains if $dv = \beta v$, where β is a scalar. This follows from homogeneity of the GDP function with respect to v; see Felbermayr and Kohler (2007).

[86]The well known Borjas (2003) setup is a case in point. The setup involves a nested CES production function with several types of labor, i.e., workers of certain levels of skills and experience levels. With many different types of labor involved, the cross-wage effects of immigration need not be positive for all conceivable skill-experience comparisons.

that an inflow of labor - other things equal - increases the marginal productivity of the accumulated factor. Whatever the details of accumulation, in conventional models of accumulation with diminishing marginal returns, this should enhance accumulation and raise the steady state level of the factor accumulated. But again, in a multi-factor setup, this complementarity relationship cannot be taken for granted. We shall return to a more detailed analysis of migration and accumulation of physical and human capital in section 5.

3.4 International convergence

3.4.1 Convex technology

Does neoclassical theory predict that factor movements lead to international factor price convergence? If we assume *complete* factor mobility, then this question essentially aims at the existence of a unique 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 intuitively, if the cause of factor movements lies in some form of technological superiority, then complete factor price convergence needs more factor movements than if the underlying cause of factor price gaps is different factor endowments with identical technology. This principle shows up in the case considered by Markusen (1983) in that international factor price equalization is reached only "at the far end" where factor movements have driven one of the two countries to complete specialization. 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.[87] 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

---

[87]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 the 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 1 good.

common technologies across all countries.[88] 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.

It seems more interesting, however, to look at convergence in a *local* sense, 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 $\mathrm{d}\left(w_A^k - w_B^k\right) < 0$. 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 L and K to denote two factors, say labor and capital, 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 counties through trade, whence differences in factor returns are always differences in *real* returns. The outcome of such movements, according to standard properties of the two-by-two model, is $\mathrm{d}w_A^L < 0$ and $\mathrm{d}w_A^K > 0$, 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

$$\mathrm{d}\left(w_A - w_B\right) \cdot \left(w_A - w_B\right) \leq 0 \tag{14}$$

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 that $\mathrm{d}v_A = S\left(w_A - w_B\right)$ captures the response of the factor inflow (outflow) in country A (B), where $w_A - w_B$ is a vector of factor price differences. The vector-valued function S may be interpreted as representing the outcome of all of country A's and country B's factor owners' decisions on where to supply their factors. The condition (5) then reads as

$$\mathrm{d}w \cdot S\left(w_A - w_B\right) \leq 0 \tag{15}$$

---

[88]This is, again, a reflection of the Le Chatelier-Samuelson principle.

One may list a few quite plausible properties of S, such as positive diagonal elements of the derivatives matrix S_w and the condition that $\mathrm{d}v_A^k > 0$ only if $w_A^k - w_B^k > 0$, where k indexes factors. However, this is not enough to guarantee convergence in the sense of (15). Conditions (14) and (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 (14). Importantly, however, theory does not suggest that there is any *force of divergence* either.

3.4.2 Increasing returns: new economic geography

We have so far assumed convex technology as a key element of neoclassical theory. Convexity of technology as such is conducive to convergence, since it implies diminishing marginal returns of individual factors. Allowing for non-convexity implies the existence in one form or another of increasing returns to scale, and 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.[89]

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 are no externalities, but internal to the firm, modeled through a fixed cost of production. Hence they require a departure also from the paradigm of perfect competition.[90] The 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. This is what we had in mind when describing our view of neoclassical theory by referring to "moderate degrees of market imperfections". 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 (1991a) and is now known as the "core-periphery model".[91] In some sense this model is the diametrical opposite of the models of trade and migration that we have used above.

---

[89]We treat the new economic geography as part of the neoclassical body of theory on the grounds that is general equilibrium in nature and that it features models with a zero profit equilibrium that do not involve market environments with strategic interactions.

[90]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).

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

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 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.[92] Importantly, manufacturing labor is assumed to be completely mobile between regions (countries), while agricultural labor is assumed immobile.[93] This is a further important departure from the above analysis where we have assumed factor-specific costs of cross-border movement.

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 are 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 to arise 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 world-wide. 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.

---

[92]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.

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

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*. Since the factor specific to manufacturing is immobile across countries, 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 forgone 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 (1991a) identifies 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 neo-classical 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 world-wide manufacturing factors. Notice that this effect only works if factor owners migrate with

the location of their factor use. Thus, it will typically be present for labor migration, but not for footloose capital. For labor migration, there is thus a destabilizing force both, from the perspective of serving markets as well as from the perspective of being served from markets.[94]

What determines the strength of these destabilizing forces? Obviously, the size of transport costs matter. To see this, first note that whenever instability obtains, by construction of our argument there will be two asymmetric stable equilibria. Moreover, if transport cost are zero, then 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 cost costs are infinite, then there is no trade. In this case there is no equilibrium other than a completely symmetric equilibrium.[95] 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 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

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

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

67

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 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.[96]

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.

---

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

4 Migration and the Formation of Physical Capital

As outlined in section 2, the first era of globalization in the 19th century was characterized by simultaneous capital and labor flows from Europe to the US (Solimano and Watts (2005)). By contrast, in the process of enlargement of the European Union (EU) to the East, labor was migrating from Eastern EU countries to the UK and to Western EU countries like Germany, whereas capital was accumulated faster in the East (Mora et al. (2004)). 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.[97] We ask how international (or interregional) labor market integration affects both private capital investment and labor migration over time. The empirical evidence discussed in this section focusses on foreign direct investment (FDI) rather than portfolio investment, as FDI seems more relevant in terms of our theoretical considerations which emphasize productive capital.

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 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) employs 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 capital-rich economy. In both economies there is a representative firm which produces a homogenous (numéraire) good according to a linearly

---

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

homogenous production function $Y = F(K, L) \equiv Lf(k)$, where Y is output, L is a homogenous labor input, K is the capital stock, and $k \equiv \frac{K}{L}$ is the capital-labor ratio; $f(\cdot) \equiv F(\cdot, 1)$ is strictly concave. The capital stock accumulates according to

$$\dot{K} = I - \delta K, \tag{16}$$

$K_0 > 0$, where I is gross investment in terms of the numéraire good and $\delta \geq 0$ is the depreciation rate. A dot on a variable denotes its derivative with respect to time. The time index, t, is omitted when obvious. We assume that installing investment I requires to incur capital adjustment costs $IG\left(\frac{I}{K}\right)$,[98] where G is an increasing and convex function.

Capital is mobile internationally. The interest rate, $r > 0$, is exogenously given from the world capital market. The representative firm maximizes the net present value of its future cash-flows, i.e., at time $t = 0$, it solves

$$\max_{\{L_t, I_t\}_{t=0}^{\infty}} \int_0^{\infty} e^{-rt} \left\{ F(K_t, L_t) - w_t L_t - I_t \left[1 + G\left(\frac{I_t}{K_t}\right) \right] \right\} dt \ \text{ s.t. (16)}, \tag{17}$$

and boundary conditions, where w is the wage rate. Thus, domestic savings have no effect on capital accumulation, which is determined by investment demand of the representative firm.[99]

Denote by q the multiplier to constraint (16), i.e., the shadow price of capital. The first-order condition of maximization problem (17) with respect to I then implies that $q = 1 + G\left(\frac{I}{K}\right) + \frac{I}{K}G'\left(\frac{I}{K}\right) \equiv \tilde{q}\left(\frac{I}{K}\right)$, where $\tilde{q}' > 0$.[100] Thus, $\frac{I}{K} = \tilde{q}^{-1}(q) \equiv \iota(q)$.

Writing (16) in per capita terms and using $I = \iota(q)K$, we have

$$\dot{k} = \left(\iota(q) - \delta - \frac{\dot{L}}{L} \right) k. \tag{18}$$

---

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

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

[100]The current-value Hamiltonian function associated with (17) reads $\mathcal{H} = F(K, L) - wL - I \cdot \left[1 + G\left(\frac{I}{K}\right) \right] + q \cdot (I - \delta K)$. First-order conditions with respect to control variables L and I are given by $F_1(K, L) = w$ and $\frac{\partial \mathcal{H}}{\partial I} = 0$. With respect to state variable K, we have $-\frac{\partial \mathcal{H}}{\partial K} = \dot{q} - rq$.

In a steady state with $\dot{K} = 0$, we have $\frac{I}{K} = \delta$, according to (16). Thus, the steady state shadow price of capital is given by $q = \tilde{q}(\delta) \equiv \bar{q}$. Combining $\frac{I}{K} = \iota(q)$ with the first-order condition with respect to the capital stock, q evolves according to

$$\dot{q} = (r + \delta)q - f'(k) - \iota(q)^2 G'(\iota(q)), \tag{19}$$

In the steady state, $\dot{q} = 0$ and $\frac{I}{K} = \iota(q) = \delta$ holds. Thus, (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 $w = \tilde{w}(\bar{k}) \equiv \bar{w}$.

Suppose the capital-rich economy is in steady state with capital-labor ratio $\bar{k}$ and wage rate $\bar{w}$ and that it is large in the sense that immigration has a negligible impact. To capture the notion of an initially capital-poor economy, we assume $k_0 = \frac{K_0}{L_0} < \bar{k}$, whence $w_0 < \bar{w}$.

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

$$\dot{\Omega} = r\Omega - \tilde{w}(k). \tag{20}$$

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; $B = b\left(\frac{\bar{w}}{r\Omega}\right)$, with $b' > 0$. Moreover, suppose that migration cost, C, increases proportionally with the emigration rate, $-\frac{\dot{L}}{L}$. Formally, $C = -\frac{1}{\mu}\frac{\dot{L}}{L}$, where parameter μ measures the degree of labor market integration. In equilibrium, the migration benefit equals migration costs, $B = C$. Thus,

$$-\frac{\dot{L}}{L} = \mu b\left(\frac{\bar{w}}{r\Omega}\right). \tag{21}$$

We are now ready to ask how the variables (k, q, Ω, L) evolve over time for $k_0 < \bar{k}$, according to dynamical system (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) examines the transitional dynamics of system (18)-(21) numerically, assuming that the production function F is of Cobb-Douglas type, adjustment cost G is linear and migration benefit function b is logarithmic. He finds monotonic convergence of both the wage rate and the shadow price of capital to the steady state. As the marginal return to capital, $f'(k)$, is high when $k_0 < \bar{k}$ but wages are low ($w_0 = \tilde{w}(k_0) < \bar{w}$), during the transition to the steady state, labor emigrates ($\dot{L} < 0$) and capital accumulates ($\dot{K} > 0$), i.e., $\frac{I}{K} = \iota(q) > \delta$. 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) demonstrates 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 (21) are typically justified by the notion that migration raises interregional differences of house prices. In fact, there is convincing evidence that immigration raises housing costs (e.g. Saiz (2003, 2007); Nygaard (2011)). Following Grossmann et al. (2012), we now incorporate the channel from migration to housing costs by introducing, in addition to a tradable goods sector, a non-tradable consumption goods sector. It uses land intensively and could be interpreted as a housing sector. To capture this assumption in sharp form, we assume that land is used in this sector only. Moreover, in order to model the migration decision of households appropriately, we use an overlapping-generations framework in discrete time.

Immigration drives up house prices. Consequently, there are supply responses through the accumulation of (residential) capital which are subject to capital adjustment costs. Again, there are no market imperfections and the foreign economy remains in steady state.

The tradable good is chosen as numéraire. Output levels of the tradable and non-tradable good, Y^T and Y^N, respectively, are given by neoclassical production technologies; these are represented by 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.

Analogously to (16), the capital stock in sector $j \in \{T, N\}$ evolves according to

$$K_{t+1}^j = I_t^j + (1 - \delta^j)K_t^j, \tag{22}$$

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

Individual live for two periods ("working-age" and "retirement") in overlapping generations. They draw utility from consumption of both goods in both periods of life and can save for retirement at the given interest r. Let p_t^N denote the (relative) price of the non-tradable good in period t. Indirect life-time utility of an individual born in t with wage income w_t but no other source of income can be written as $V(w_t, p_t^N, p_{t+1}^N)$, where V is increasing in w and decreasing in the non-tradable goods 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 $\tilde{w}^*$ and $\tilde{p}^*$ the steady state value of the wage rate and the price for non-tradables in the foreign economy, respectively. Thus, $V^* \equiv V(\tilde{w}^*, \tilde{p}^{N*}, \tilde{p}^{N*})$ is the steady state utility of a worker abroad with wage income $\tilde{w}^*$. Any

worker, wherever born, is indifferent between staying and migrating if $V(w_t, p_t^N, p_{t+1}^N) = V^*$.[101] This condition must hold in equilibrium with integrated labor markets. In this case, the number of workers, L, is endogenous. Prior to migration, the number of old natives, $L_{-1} > 0$, is given. With inelastic labor supply, in equilibrium, we have $L^N + L^T = L$.

In addition to possible productivity differences, bilateral migration flows depend on the (initial) difference in the population density. Intuitively, an increase in population density raises the price of non-tradables, because the non-tradable is land-intensive and an increased population size leads to a land dilution effect. In equilibrium, both the price of non-tradables 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 the price of non-tradables. This raises the shadow price of capital, inducing capital accumulation. Moreover, the price of land rises along with immigration during the entire transition path. Thus, even in the absence of negative wage effects of immigration,[102] native workers in the destination country may lose from labor market integration unless they own a sufficient amount of land. Conversely, outward migration slows down the capital accumulation process but may benefit natives 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 reverse migration during 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 and complementar-

---

[101]The implicit assumption is that there is a substantial fraction of landless individuals in the economy. As landless workers have a higher incentive to migrate than landowning workers, due to the declining marginal utility of income, this ensures that the marginal migrant is a landless worker.

[102]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 price of non-trabables and, for landowners, through the price of land.

ity between capital and labor is capable to explain factor movements in the same or in opposite directions, depending on initial conditions and the point in time of the transition. Moreover, substantial outward migration after labor markets are opened may be followed by immigration.

4.2 Increasing Returns and Agglomeration Effects

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

Burda and Wyplosz (1992) consider 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 apply their model to the case of the German unification. Allowing for bilateral factor movements as well, Faini (1996) and Reichlin and Rustichini (1998) capture "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. Final output (the numéraire good) of a domestic representative firm is produced according to

$$Y = AK^\alpha L^{1-\alpha}, \tag{23}$$

$\alpha \in (0,1)$, 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 in the economy, h,[103] and the average level of physical capital, $\bar{K}$, according to

$$A = h^\beta \bar{K}^\gamma, \tag{24}$$

---

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

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

We consider the case where capital is perfectly mobile, whereas after labor market integration there is still less than full labor mobility. Denote foreign variables by superscript *. As in section 3, 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 generate interior equilibria in a simple way, 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 factor input adjustment costs.

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

$$r \equiv \left. \frac{\partial Y}{\partial K} \right|_{K = \bar{K}} = \alpha h^\beta k^{\alpha + \gamma - 1} L^\gamma, \tag{25}$$

according to (23) and (24). Thus,

$$\frac{k}{k^*} = \left(\frac{h}{h^*} \right)^{\frac{\beta}{1 - \alpha - \gamma}} \left(\frac{L}{L^*} \right)^{\frac{\gamma}{1 - \alpha - \gamma}}. \tag{26}$$

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 (1990).

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

$$w = \left. \frac{\partial Y}{\partial L} \right|_{K = \bar{K}} = (1 - \alpha) h^\beta k^{\alpha + \gamma} L^\gamma. \tag{27}$$

Computing the relative wage $\frac{w}{w^*}$ from (27) and substituting (26), we find

$$\frac{w}{w^*} = \frac{k}{k^*} = \left(\frac{h}{h^*} \right)^{\frac{\beta}{1 - \alpha - \gamma}} \left(\frac{L}{L^*} \right)^{\frac{\gamma}{1 - \alpha - \gamma}}. \tag{28}$$

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, also $\frac{w_1}{w_1^*} < \rho$. 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

$$h_1 = \frac{h_0 + \lambda h_0^*}{1 + \lambda} < h_0, \tag{29}$$

whereas $h_1^* = h_0^*$. Using (28), (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

$$\frac{w_1}{w_1^*} = \left[\frac{\left(\frac{h_0}{h_0^*} + \lambda\right)^{\beta}(1+\lambda)^{\gamma-\beta}}{(1-\lambda)^{\gamma}}\right]^{\frac{1}{1-\alpha-\gamma}}. \tag{30}$$

If $\gamma \geq \beta$, $\frac{w_1}{w_1^*}$ is increasing in the fraction of immigrants in the total population, λ. On the one hand, migration depresses the average human capital level in the destination country, which reduces migration incentives (Burda and Wyplosz (1992)). On the other hand, immigration induces physical capital inflows, due to the complementarity between labor and physical capital. For $\gamma \geq \beta$, the latter effect on relative wages dominates the first one. Thus, if λ is high enough, $\frac{w_1}{w_1^*} > \rho$ such that maximum migration is a second equilibrium outcome, in addition to the equilibrium without migration. If $\frac{w_0}{w_0^*} > \rho$, an equilibrium without migration may not exist in any period, such that the low-income country may vanish in finite time. Capital flows in the direction of migration flows even enhance migration incentives over time.

Our simple model illustrates some general lessons from models with migration under increasing returns. First, as in the literature on new economy geography (see section 3.4.2),[104] they help to explain core-periphery patterns. More generally, our discussion

---

[104]The new economic geography literature typically abstracts from capital mobility. Ottaviano and Thisse (2004); Ottaviano et al. (2002) provide an excellent survey. Moreover, Baldwin and Martin

has highlighted the important insight that, in the presence of increasing returns and factor mobility, initial conditions ("history") matter for the range of equilibrium outcomes which can rationally be expected (see also Krugman (1991b)). We have seen that 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, both migration or no migration are 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 (2011) propose a dynamic multi-region setting with increasing returns to study the effects of a simultaneous integration of both capital and labor markets. In contrast to the presented model, they show that non-monotonic adjustments paths for the capital stock and the labor force may arise after an integration shock. Their contribution highlights the interaction of history and expectations for regional development.

4.3 Migration and Foreign Direct Investment: Empirical Evidence

Figure 7 exploits international data on the stock of immigrants (World Bank data) and the stock of inward FDI (IMF data) of OECD countries for the years 1990 and 2000 in order to show that, if anything, the relationship is positive and has become larger over time.

Buch et al. (2006) employ panel data for the time period 1991-2002 on both stocks of immigrants and stocks of inward FDI from foreign countries in the 16 German federal states. They find that an increase in the stock of immigrants significantly raises the stock of inward FDI, whereas a higher domestic labor force has no significant effect on inward FDI. 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 tend to be more highly skilled than immigrants

---

(2004) discuss international agglomeration effects from capital mobility in endogenous growth models, but abstract from labor mobility.

Figure 7. Relationship between the population-adjusted stock of immigrants and the stock of inward FDI in the year 1990 (red dots) and 2000 (blue dots) in OECD countries.
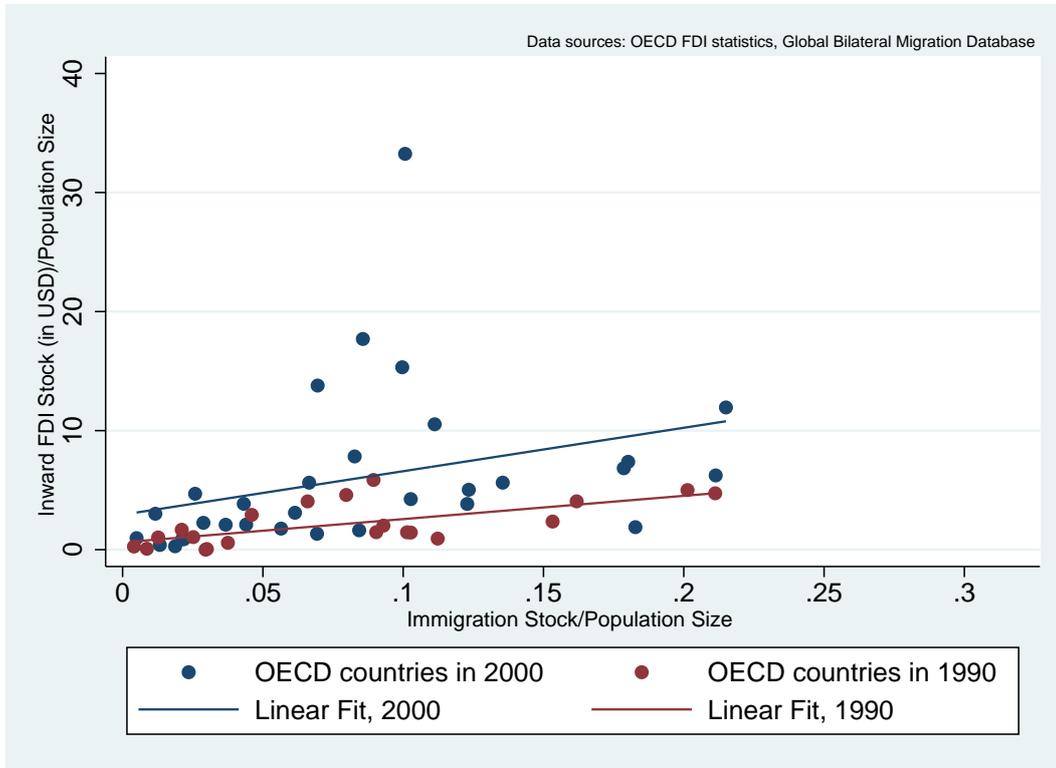
on average, this suggests that high-skilled immigrants are important to create ties to foreign companies investing in Germany. Conversely, the impact of an increase in inward FDI on immigration from the same source country is generally insignificant. It is positive in the east of Germany if and only if the source country has high-income. In sum, the evidence is consistent with agglomeration effects of high-skilled immigration.

Kugler and Rapoport (2007) investigate the impact of both, a higher stock of immigrants in the US and a higher immigration inflow on FDI financed by US firms in the immigrants' country of origin. They break down the immigrants by education level (low, medium, high) and FDI by sectors (manufacturing, services), finding 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. In the manufacturing

79

sector, higher labor inflows between 1990 and 2000 have a positive and significant effect on FDI flows from the US to the origin country of immigrants if and only if migrants enjoyed tertiary education. Tertiary educated immigrants also matter for the service sector. Whereas their stock in 1990 is positively associated with FDI outflows in the subsequent decade, a higher growth rate of high-skilled immigration reduces the growth rate of FDI outflows. This suggests that, at least from a dynamic perspective, high-skilled labor and FDI are complements.

Javorcik et al. (2011) study the effects of a higher stock of immigration to the US on the stock (rather than the change in the subsequent period) of outward FDI. By instrumenting immigration stocks in line with the literature on the determinants of migration (e.g. Beine et al. (2011); Grogger and Hanson (2011b)), their estimates avoid potential endogeneity problems of the earlier studies. These 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 one percent increase in the stock of skilled immigrants raises the stock of US outward FDI by about 0.5 percent.[105]

5 High-Skilled Migration and Productivity Growth

The relationship between high-skilled migration and productivity growth is potentially important beyond physical capital formation. For instance, the US still attracts talented people from abroad to places like Silicon valley and to elite universities. High-skilled immigrants often come from other advanced regions like Europe. As documented by Saint-Paul (2004) for the years 1990 and 2000, about 1.5 percent of the working-aged population born in the UK and Germany and about one percent born in Italy lived in the US. In 2000, the fraction of the expatriate population with tertiary education from France, UK, Spain and Germany was 56, 49, 44 and 42 percent, respectively, which is much higher than the fraction of skilled workers in the respective source country.

---

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

A significant proportion of the expatriates from Europe (e.g. 5 percent among the French) hold a PhD.[106] We would expect high-skilled immigrants to contribute to innovation and productivity growth in their host countries. Vice versa, 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 which 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 which lead to new capital goods.

The Romer-Jones Model Consider a large economy with population size $L_t = \bar{L}e^{nt}$ 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 $\bar{L}$. There is a representative household, who possesses the average amount of assets, and inelastically supplies one unit of labor to the production of a homogenous consumption good (the numéraire) or to a 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

$$U = \int_0^\infty \frac{(c_t)^{1-\sigma} - 1}{1 - \sigma} e^{-(\theta-n)t} \mathrm{d}t, \tag{31}$$

---

[106] Also Switzerland, known for its financial industry and health sector, attracts high-skilled migrants at a large scale. For instance, after immigration from 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 2007-2010 (www.bfs.admin.ch). Over 60 percent of immigrants from Germany hold a university degree.

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

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

$$Y = (L^Y)^{1-\alpha} \int_0^A x(i)^\alpha di, \tag{32}$$

$\alpha \in (0, 1)$, 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 forgone consumption can be transformed into one unit of a capital good; that is, marginal costs are given by the (endogenous) interest rate, r. The physical capital stock then is $K = \int_0^A x_i di$. In symmetric equilibrium where $x(i) = \frac{K}{A}$ for all i, we find that per capita income reads $y \equiv \frac{Y}{L} = k^\alpha (Al^Y)^{1-\alpha}$, where $k \equiv \frac{K}{L}$ and $l^Y \equiv \frac{L^Y}{L}$. 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 (32).

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

$$\dot{A} = \lambda A^\phi L^A, \tag{33}$$

$\lambda > 0$, $\phi \leq 1$. If $\phi > 0$ there is an "intertemporal knowledge spillover" from previous R&D. In his seminal paper on endogenous technical change, Romer (1990) assumes (implicitly) that $\phi = 1$ and $n = 0$, whereas Jones (1995) analyses 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 $g = \frac{\dot{A}}{A} = \frac{\dot{y}}{y} = \frac{\dot{k}}{k} = \frac{\dot{c}}{c}$.

In the Romer-model, with $\phi = 1$, $\frac{\dot{A}}{A} = \lambda L^A$ is increasing in the number of R&D workers, L^A. Romer (1990) shows that, in the steady state, L^A is increasing in population size, $\bar{L}$. In this sense, immigration of workers capable to perform 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 $g = \frac{\dot{A}}{A} = \lambda A^{\phi-1} L^A$ is time-invariant if and only if the growth rate of R&D employment, $n^A \equiv \frac{\dot{L}^A}{L^A}$, is time-invariant and $g = \frac{n^A}{1-\phi}$. 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 $\bar{L}$. 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, $\bar{L}$, raises the detrended level of the knowledge stock, $\tilde{A}_t \equiv \frac{A_t}{e^{gt}}$, thus raising the level of per capita income, $\tilde{y}_t \equiv \frac{y_t}{e^{gt}}$, in the long run (as $t \to \infty$). In other words, immigration of (skilled) labor raises the TFP level and thus makes an economy richer. This property is typically referred to as "weak scale effect".[107] If anything, international migration leads to divergence rather than convergence of per capita income across economies.

Directed Technical Change As emphasized in Acemoglu (1998) and Acemoglu (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, $\bar{H}$ and $\bar{L}$, respectively. Final output (the numéraire)

---

[107]An increase in $\bar{L}$ also raises welfare U. In the long run, the utility integral (31) is finite if $\rho > n + (1 - \sigma)g > 0$, which also ensures that the transversality condition for the problem of the representative consumer holds.

is produced under perfect competition according to the CES production function

$$Y = \left(\gamma(X_L)^{\frac{\varepsilon-1}{\varepsilon}} + (1-\gamma)(X_H)^{\frac{\varepsilon-1}{\varepsilon}}\right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{34}$$

$\gamma \in (0,1)$, $\varepsilon > 0$. 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

$$X_L = (L^X)^{1-\alpha} \int_0^{A_L} x_L(i)^\alpha di, \tag{35}$$

$$X_H = (H^X)^{1-\alpha} \int_0^{A_H} x_H(i)^\alpha di, \tag{36}$$

$\alpha \in (0,1)$, where $x_L(i)$ and $x_H(i)$ are inputs of machines which 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

$$\dot{A}_L = \lambda_L(A_L)^\phi H_L^A, \tag{37}$$

$$\dot{A}_H = \lambda_H(A_H)^\phi H_H^A, \tag{38}$$

$\lambda_L, \lambda_H > 0$, $\phi \leq 1$, where H_L^A and H_H^A are the number of scientists directed to innovations which are complementary to unskilled and 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, $P_L \equiv \frac{\partial Y}{\partial X_L}$, $P_H \equiv \frac{\partial Y}{\partial X_H}$. Using (34), this gives us relative inter-

mediate goods demand:

$$\frac{X_H}{X_L} = \left(\frac{1-\gamma}{\gamma}\right)^{\varepsilon} \left(\frac{P_H}{P_L}\right)^{-\varepsilon}. \tag{39}$$

According to (35), the inverse demand for machine i in the low-skilled intensive sector is $p_L(i) \equiv \alpha P_L \left(\frac{L^X}{x_L(i)}\right)^{\alpha-1}$. Thus, the optimal 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 (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

$$\frac{\pi_H}{\pi_L} = \left(\frac{P_H}{P_L}\right)^{\frac{1}{1-\alpha}} \frac{H^X}{L^X}. \tag{40}$$

Moreover, relative supply of composite inputs is

$$\frac{X_H}{X_L} = \frac{A_H H^X}{A_L L^X} \left(\frac{P_H}{P_L}\right)^{\frac{\alpha}{1-\alpha}}. \tag{41}$$

Equating the right-hand sides of (39) and (41) leads to a negative relationship between the relative price of the two composite inputs, $\frac{P_H}{P_L}$, and relative "efficiency units" of labor, $\frac{A_H H^X}{A_L L^X}$:

$$\left(\frac{P_H}{P_L}\right)^{\frac{1}{1-\alpha}} = \left(\frac{1-\gamma}{\gamma}\right)^{\frac{\varepsilon}{\alpha+\varepsilon(1-\alpha)}} \left(\frac{A_H H^X}{A_L L^X}\right)^{-\frac{1}{\alpha+\varepsilon(1-\alpha)}}. \tag{42}$$

Incentives to innovate in a certain direction depend on relative profits, $\frac{\pi_H}{\pi_L}$. Now consider a selective immigration policy towards skilled labor, such that $\frac{\bar{H}}{\bar{L}}$ rises. Consequently, the (steady state) labor allocation will change such that relative employment of skilled labor in production, $\frac{H^X}{L^X}$, rises. This has two counteracting effects on $\frac{\pi_H}{\pi_L}$. First, according to (40), for a given relative price $\frac{P_H}{P_L}$, relative profits in the high-skilled intensive sector rise ("market size effect"). Second, however, according to (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 $\frac{A_H}{A_L}$; 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 an in Young (1998).

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

$$Y = BZ^{1-\alpha-\beta}(L^Y)^\beta \int_0^N A(i)^{1-\alpha}x(i)^\alpha \mathrm{d}i, \tag{43}$$

$\alpha, \beta \in (0,1)$, $\alpha + \beta \leq 1$, 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 which we assume to be given exogenously from the world capital market. We allow for adverse congestion effects from increasing density of the (working-age) population, $D \equiv \frac{L}{Z}$, on the productivity parameter B:

$$B = D^{-\eta}, \tag{44}$$

$\eta \geq 0$; 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

$$A_{t+1}(i) = \bar{A}_t \Lambda(l_t(i)), \tag{45}$$

where $\bar{A}_t \equiv \frac{1}{N_t} \int_0^{N_t} A_t(i) \mathrm{d}i$ is the average product quality in period t. $\bar{A}$ 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 $\Lambda(0) = 1$. $\bar{A}_0 > 0$ and $L_0^Y \in (0, \bar{L})$ are historically given.

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

$$y \equiv \frac{Y}{L} = \left(\frac{\alpha^2}{r}\right)^{\frac{\alpha}{1-\alpha}} (l^Y)^{\frac{\beta}{1-\alpha}} D^{\frac{\beta-\eta}{1-\alpha}-1} N \bar{A}, \tag{46}$$

As shown in appendix, in equilibrium, R&D labor input per 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 (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, $\bar{A}$ evolves according to $\bar{A}_{t+1} = \bar{A}_t \Lambda(\tilde{l})$. It is not affected by population size, although aggregate R&D input, $N\tilde{l}$ is increasing in L. This result is implied by the assumption that the intertemporal spillover effect is driven by average product quality, $\bar{A}$, thus being independent of the number of firms in equilibrium. Consequently, there is no scale effect in growth rates.[108] 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 (46), these properties imply that the impact of immigration (increase in $\bar{L}$), on per capita income, y, is positive (negative) if $\beta > (<)\eta$. There are three kinds of scale effects. 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.[109] 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

---

[108]Lundborg and Segerstrom (1998) propose a quality-ladder endogenous growth model with North-South trade and migration from the South to the North. In their framework, immigration lowers wages and therefore spurs R&D employment in the North which is growth-enhancing. They find that, nevertheless, immigration lowers welfare in the North.

[109]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.

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 which only feature specialization gains.[110]

5.1.3 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 points 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 panel data for the period 1940-2000, Hunt and Gauthier-Loiselle (2010) argue that skilled immigration into the US could have raised US TFP growth. They find that an increase in the population share of immigrant college graduates by one percentage point raises patents per capita by 9-18 percent. Consistent with at least the weak form of positive scale effects from migration, by using international data on bilateral migration stocks, Grossmann and Stadelmann (2012b) find that high-skilled migration has a small but positive effect on the relative destination-to-source level of both income and TFP. Ortega and Peri (2012) argue that, in view of technology transfer through intermediate goods trade across countries (e.g. Coe and Helpman (1995)), effects of migration and trade have to be disentangled. Their instrumental variable estimations suggest that immigration has a positive effect on GDP per capita by raising the employment rate which, however, is offset by a negative effect on TFP. The latter result is consistent with congestion effects. However, their study does not distinguish between high-skilled and low-skilled immigration.

5.2 Brain Drain and Human Capital Formation

In an important early contribution, Bhagwati and Hamada (1974) analyze migration effects in a static context with rigid wages and endogenous education. They argue that,

---

[110]See Grossmann (2013) for further discussion. In view of (46), the facts that l^Y is time-invariant in the long run and $\bar{A}_{t+1} = \bar{A}_t \Lambda(\tilde{l})$, the steady state growth rate of income per capita is given by $g = (1+n)^{\frac{\eta-\beta}{1-\alpha}} \Lambda(\tilde{l}) - 1$. Thus, if $\eta > \beta$, g is decreasing in the population growth rate, n.

even in absence of externalities, 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 (1997b); Stark et al. (1997b); Beine et al. (2008, 2001). We illustrate the basic mechanism for this possibility in a simple dynamic framework before discussing empirical evidence.

5.2.1 A Simple Dynamic Model

Suppose there is a perfect competition in all markets. 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 better substitutable than human capital and raw labor. To capture this assumption in its sharpest relief, we assume that K and L are perfectly substitutable such that output depends on a simple composite of both factors, $X \equiv K + L$:

$$Y = AF(X, H) \equiv AHf(\kappa), \tag{47}$$

where $A > 0$ is a TFP parameter, H is the number of skilled workers remaining in the country after emigration, and $\kappa \equiv \frac{X}{H}$. F is a linearly homogenous 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 may 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 $\kappa = (f')^{-1}\left(\frac{r}{A}\right) \equiv \tilde{\kappa}(A)$,[111] where $\tilde{\kappa}' > 0$. Thus, the wage rate per unit of skilled labor reads as

$$w_H = A\left[f\left(\tilde{\kappa}(A)\right) - \tilde{\kappa}(A) f'\left(\tilde{\kappa}(A)\right)\right] \equiv \tilde{w}_H(A), \tag{48}$$

which 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,

---

[111]For notational simplicity, we suppress the interest rate, r, in functions we define from now on.

for a given level of TFP, 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 $t = 1, 2, ...$ are characterized by the intertemporal utility function

$$U_t(i) = u(c_{1t}(i)) + \theta u(c_{2t+1}(i)), \tag{49}$$

where $c_{1t}(i)$ and $c_{2t+1}(i)$ are consumption levels in the first and second period 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 standard condition $\theta(1 + r) = 1$. Thus, optimal savings of an individual with first-period (labor) income $y_t(i)$ imply $c_{1t}(i) = c_{2t+1}(i) = \frac{y_t(i)}{1+\theta}$. Consequently, intertemporal life-time utility is increasing in income; we have $U_t(i) = (1 + \theta)u\left(\frac{y_t(i)}{1+\theta}\right) \equiv V(y_t(i))$ 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 e are distributed according to 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. Provided that $w^* > w_H = \tilde{w}_H(A)$, an individual i acquires schooling if and only if

$$p \cdot V((1 - e(i))w^*) + (1 - p) \cdot V((1 - e(i))\tilde{w}_H(A)) \geq V(w_L) = V(r). \tag{50}$$

Thus, an individual i becomes skilled when individual time cost $e(i)$ is below some threshold level $\bar{e} = \bar{e}(p, A, w^*)$, which depends negatively on the level of TFP, A,

positively on the emigration quota, p, and positively on the foreign wage rate, w^*. Thus, the share of natives who acquire schooling,

$$s = \Phi(\bar{e}(p, A, w^*)), \tag{51}$$

is increasing in p and w^*, but decreasing in A.

Suppose that TFP is determined by the previous period's fraction of skilled workers after emigration took place. Formally, we assume $A_{t+1} = h_t$. The initial level of TFP, $A_0 > 0$, is given. As $H = (1-p)s$ is the fraction (and number) of skilled workers after migration took place and $1 - ps$ is the fraction (and size) of the population remaining in the country after migration, we have

$$A_{t+1} = h_t = \frac{(1-p)s_t}{1-ps_t} = \frac{(1-p)\Phi(\bar{e}(p, A_t, w^*))}{1-p\Phi(\bar{e}(p, A_t, w^*))} \equiv \tilde{h}(p, A_t, w^*). \tag{52}$$

The fraction of skilled workers after migration, h, is decreasing in the contemporaneous TFP level, A. This is because an increase in A raises the wage rate of skilled workers at home and therefore mitigates the wage gap towards a potential destination country. Consequently, a higher A is associated with a lower incentive to acquire schooling when $w^* > w_H$. Thus, (52) implies that the steady state level of TFP, which is implicitly given by $A = \tilde{h}(p, A, w^*)$, is unique.

May labor market integration for skilled workers raise the TFP level in the source country, therefore boosting the wage rate of skilled workers, $w_H = \tilde{w}_H(A)$? Both in the long run as well as during the transition to the steady state, the equilibrium level of A is increasing in the migration quota p, if $\tilde{h}(p, \cdot)$ is increasing in p. There are two counteracting effects. First, an increase in p lowers the fraction of skilled workers after migration, $h = \frac{(1-p)s}{1-p\cdot s}$, for a given share of skilled natives, s. This captures the standard "drain effect". Second, a higher migration quota raises $s = \Phi(\bar{e}(p, \cdot))$ due to better emigration prospects. 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, $K = H\tilde{\kappa}(A) - L$,[112] may increase in p for three reasons which 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 (52). Thus, if p increases, the emigration flow ps rises and physical capital inflows may rise as well.

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.

5.2.2 Empirical Evidence

Beine et al. (2008, 2001) provide empirical evidence which 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) report that the effect is higher for poor countries, Beine et al. (2008) do not find support for non-linearities. Beine et al. (2008) instrument 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 percent.

Based on this estimate, they then provide simulation results on the counterfactual share of skilled workers and compare it with the actual one. 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

---

[112]Recall that $\kappa = \frac{K+L}{H}$ by definition and $\kappa = \tilde{\kappa}(A)$.

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 or smaller than the drain effect.

6 Migration in the Gravity Equation of Trade

What is the causal relationship between international trade and international migration? There is a substantial empirical literature that studies this link, usually with the aim of 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.[113] Before we discuss the empirical evidence, a short review of the conceptual foundation of the gravity equation is warranted.

6.1 Conceptual foundation of the gravity equation

Tinbergen (1962) expresses 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}

$$M_{ij} = G \frac{Y_i Y_j}{D_{ij}^{\delta}}, \tag{53}$$

---

[113]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.

93

where G is a constant and δ is a parameter.[114] That formulation is formally akin to Newton's law of gravity in physics. In that context, X_{ij} is the force between two objects i and j, Y_i and Y_j would be their respective masses, G would be the gravitational constant, and δ would be equal to two. Researchers have long estimated (53) by adding a multiplicative error term and by applying OLS to the log-linearized model.

The size of markets–as captured by Y_i 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 of 2000 delivers coefficients on GDPs close to unity and an estimate of δ at -1.37.[115] 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.[116] Therefore, Anderson (2011) describes 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) have provided rigorous underpinnings 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) or Bergstrand and Egger (2011). 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 having at the same time 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 that delivery of goods uses resources in the same proportion as the production of those same goods. Iceberg trade costs, as

---

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

[115]Coefficients on log GDPs are numerically close to one, but statistically they are different from one at the 1% significance level. Results based on own estimations using data from the IMF.

[116]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.

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 when 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 Anderson and van Wincoop (2003) perfect competition model. A 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 which 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 goods amongst one class of goods (i.e., within the same sector). Finally, also the Melitz (2003) model, with Pareto distributed firm-level productivities, gives rise to a similar gravity equation. There the elasticity of trade flows with respect to trade costs is the negative of the Pareto shape parameter (Chaney (2008)). This result obtains despite of 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 also 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 drawing 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.[117]

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 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 (Dunlevy (2006); Combes et al. (2005); Herander and Saavedra (2005); Rauch and Trindade (2002)). Thus, migrant networks mitigate *imcomplete* 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 its 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 two first 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 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. That 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

---

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

above.

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

$$U_i = \sum_{j=1}^{C} \sum_{h=1}^{n_j} \left(a_{ij} m_{ijh}\right)^{\frac{\sigma-1}{\sigma}}, \sigma > 1, \tag{54}$$

where m_{ijh} denotes consumption of a generic variety h produced in country j. N is the number of countries, n_j is the mass of varieties produced in country j, and σ denotes the elasticity of substitution. The only modification relative to the standard specification is the inclusion of a bilateral affinity term a_{ij} which describes the preference of the representative consumer in country i for country j's products. Maximizing (54) subject to an appropriate budget constraint, utilizing the market clearing conditions, and writing iceberg trade costs $T_{ij} \geq 1$, the (c.i.f., i.e., inclusive of cost, insurance, and freight) value of bilateral imports M_{ij} of country i from j can be written as

$$M_{ij} = \frac{Y_i Y_j}{Y_w} \left(\frac{T_{ij}}{a_{ij}}\right)^{1-\sigma} \left(\tilde{P}_i \tilde{P}_j\right)^{\sigma-1}, \tag{55}$$

where the price indices $\tilde{P}$ solve $\left(\tilde{P}_j\right)^{1-\sigma} = \sum_{i=1}^{C} (Y_i/Y_w) (T_{ij}/a_{ij})^{1-\sigma} \left(\tilde{P}_i\right)^{\sigma-1}$; see Feenstra (2004a) for the details of the derivation. Anderson and van Wincoop (2003) call $\tilde{P}_i$ and $\tilde{P}_j$ indices of inward and outward *multilateral* resistance, respectively, because they depend on the trade costs of country i with all countries in the world, the number of which is given by C. The variables Y_i denote GDP of country i, the subindex w refers to the world. The elasticity of substitution in the underlying CES utility function is given by σ. We will be 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 the 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.[118]

---

[118]See Hogrefe, Jung and Kohler (2012).

The central insight of Anderson and van Wincoop (2003) is 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 $\tilde{P}_i$ terms.

Equation (55) 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 (55). This yields

$$M_{ij}^{1/2} M_{ji}^{1/2} = \frac{Y_i Y_j}{Y_w} \left(\frac{T_{ij}}{a_{ij}}\right)^{\frac{1-\sigma}{2}} \left(\frac{T_{ji}}{a_{ji}}\right)^{\frac{1-\sigma}{2}} \left(\tilde{P}_i \tilde{P}_j\right)^{\sigma-1}. \tag{56}$$

The simple arithmetic mean $M_{ij} + M_{ji}$ leads to an additive structure that does not lead 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 the 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} depend on costs related to informational frictions, I_{ij}, which, as described before, may be affected by migrant networks. 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,

$$I_{ij} = \iota \left(\{N_{ik}, N_{jk}\}_k\right) \text{ with } i, j, k = 1, \ldots, C,$$

where N_{ik} describes the stock of individuals from country (or ethnicity) k residing in i. Information about i in j or vice versa can, in principle, be conveyed by networks

of ethnicity k, where k can of course include i or j. Usually, the literature focuses on *direct* links (e.g., Combes et al. (2005), where $k \in \{i, j\}$, i.e, $I_{ij} = \iota(N_{ij}, N_{ji})$: here, 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. 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 on 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, $s_{ik} = N_{ki}/L_i$ 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

$$\ln I_{ij} = \ln \iota(N_{ij}, N_{ji}) = \bar{\mu}_1 \ln N_{ij} + \bar{\mu}_2 \ln N_{ji}, \tag{57}$$

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

$$\ln \iota(s_{ij}, s_{ji}) = \bar{\mu}_1 \ln s_{ij} + \bar{\mu}_2 \ln s_{ji} = \bar{\mu}_1 \ln N_{ij} + \bar{\mu}_2 \ln N_{ji} - \bar{\mu}_1 \ln L_i - \bar{\mu}_2 \ln L_j. \tag{58}$$

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.[119] It is in line with the idea that the pro-trade effects of migrants' networks are 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 $\bar{\mu}_1 \ln L_i$ and $\bar{\mu}_2 \ln L_j$ are often absorbed into 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

$$T_{ij} = \bar{T}_{ij} I_{ij} D_{ij}^{\bar{\delta}} e^{\bar{\lambda}(1-LANG_{ij})} e^{\bar{\gamma}(1-ADJ_{ij})} e^{\bar{\pi}(1-FTA_{ij})}, \tag{59}$$

where D_{ij} measures geographical distance, $LANG_{ij}$ is a dummy that takes value one 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 easy to include in a similar log-linear fashion. $\bar{T}_{ij}$ captures unobserved and hence omitted determinants of bilateral trade costs. Substituting for I_{ij} and taking logs, we obtain

$$\ln T_{ij} = \boldsymbol{\xi}'_T \mathbf{X}_{ij} - \nu_T^{im} \ln N_{ij} - \nu_T^{em} \ln N_{ji}, \tag{60}$$

where $\boldsymbol{\xi}$ is a vector of coefficients on controls $\mathbf{X}_{ij}$, 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. This is the *trade cost channel* of networks.

---

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

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 (54). We assume that a_{ij} is, amongst other things, a function of the 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:

$$\ln a_{ij} = \boldsymbol{\xi}_a' \mathbf{X}_{ij} + \nu_a N_{ij}, \tag{61}$$

where ν_a is expected to be positive.[120] 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 market bias that immigrants may have; it is also consistent with the idea that the presence of immigrants in some country may by 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.[121]

Egger, Ehrlich and Nelson (2012) argue that the functional forms implied by (60) and (61) 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. 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

---

[120] The network channel cannot operate against the direction of trade whence $\nu_a^{em} = 0$ and we can drop indices on ν_a.

[121] 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.

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 (55) poses a number of problems. First, while theoretical considerations imply the precise structure of (55), they tell us nothing about the correct specification of T_{ij}. The formulation (59) is common, but prone to specification error. The same holds true for (61). Also, one must make assumptions on the intercepts of the trade cost and preference functions, $\bar{T}_{ij}$ and $\bar{A}_{ij}$, the unmodeled parts of costs and preferences. These are usually assumed to be included by 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 $\tilde{P}_i$ are essentially unobserved since they do not correspond to official CPI deflators. Anderson and van Wincoop (2003) show how one can solve for the $\tilde{P}_i$ terms numerically and use them in an iterative estimation strategy. They demonstrate that the failure to control for multilateral resistance typically biases the absolute value of estimated trade cost variables upwards. Rauch and Trindade (2002) recognize 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 (2004a) and Anderson (2011) who argue that the use of importer and exporter specific fixed effects in a simple OLS model leads to very similar results than 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, yields identical results; see Baier and Bergstrand (2007) for

a similar strategy.[122]

Employing these specifications for T_{ij} and a_{ij} in (55), and using non-overlapping sets of country dummies μ_i and μ_j to control for the country-specific (multilateral) variables, we have

$$
\begin{aligned}
\ln M_{ij} &= \ln \left(\frac{Y_i Y_j}{Y^w} \right) + (\sigma - 1) \left(\boldsymbol{\xi}_a' - \boldsymbol{\xi}_T' \right) \mathbf{X}_{ij} + \\
&\quad (\sigma - 1) \left[\left(\nu_a + \nu_T^{im} \right) \ln N_{ij} + \nu_T^{em} \ln N_{ji} \right] + \mu_i + \mu_j + \epsilon_{ij}, \\
&= \ln \left(\frac{Y_i Y_j}{Y^w} \right) + \bar{\boldsymbol{\xi}}' \mathbf{X}_{ij} + \bar{\nu}^i \ln N_{ij} + \bar{\nu}^j \ln N_{ji} + \mu_i + \mu_j + \epsilon_{ij}.
\end{aligned} \tag{62}
$$

Alternatively, one can also normalize trade flows by $Y_i Y_j$ so that the term $\ln(Y_i Y_j)$ disappears from the right-hand-side of the equation. We will be interested by estimates of the parameters ν_a, ν_T^{im} and ν_T^{em} which are consistently and unbiased estimated under the assumption that $\{\mathbf{X}_{ij}, \ln N_{ij}, \ln N_{ji}\}$ 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 matrix (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 one 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 (55) requires knowledge of trade costs between *all* trading countries. Even if one is

---

[122]There are alternative ways to deal with multilateral resistance indices. Baier and Bergstrand (2009) have proposed to work 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 then multiplied. In the resulting equation, so that unobservable multilateral resistance terms drop out.

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 (62) 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 $i's$ 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 (62) 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 (62) 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., $\varepsilon_{ijt} - \varepsilon_{ijt-1}$ is white noise). Baier and Bergstrand (2007) recommend first-differencing in the context of gravity equations. They argue 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) warns that estimation of (62) 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 (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 (54). Since sectoral output and expenditure data is not easy to come by for many levels of aggregation and countries, it is preferable to control for these variables by writing $\ln\left(Y_i^s Y_j^s\right) = \ln Y_i^s + \ln Y_j^s$ and let the country-sector dummies take care of them.

For example, Rauch and Trindade (2002) run equation (62) on different dependent variables: First, they focus 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 pro-

ducers. Hence, σ is probably low for differentiated goods, higher for reference-priced ones, and highest for exchange-priced ones. Also, the informational needs 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 $\bar{\boldsymbol{\xi}}'$ and $\{\bar{\nu}^{im}, \bar{\nu}^{em}\}$ obtained from these different regressions. For example, even if the trade cost and the preference channel could be separated, for a given strength of the network effect ν_T^k, the estimated coefficient $(\sigma - 1)\nu_T^k$ would be large for homogeneous goods since the degree of substitutability is high and 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 present the reader with the key challenges found in the literature and how it has dealt with them. We focus on recent work which 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 with discussing conditions under which consistent estimates of a model such as (62) 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. It is thus necessary to leave this important question open.

6.2.1 A quick browse over different strands of thought

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 to ideally employ the full matrix of trade flows; this, in turn, also requires matching data on migrant stocks. While, quality data on 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 exist since 1950.

Almost all papers in the literature focus on the effects of immigration, i.e., they set $\bar{\mu}_2$ to zero in (57). The first paper in the modern empirical tradition is Gould (1994). He studies the effect of immigration on trade between the United States and 47 trading partners for the years of 1970-1986. He estimates 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 does not exploit the panel nature of the data, treating it as repeated cross-sections, but does include the lagged value of trade flows on the right-hand-side. Using non-linear least squares methods, and ignoring a potential Nickel bias, he distinguishes between producer and consumer goods and between imports and exports. He finds that immigrants increase trade, but the estimated effects are hard to compare to the subsequent literature that employs 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 finds 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 follow 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 country-level 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 exploit 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 (2007); Bandyopadhyay et al. (2008); Tadesse and White (2008). Herander and Saavedra (2005) analyzes the relative effects of state and nation-level migrant stocks for the US. Helliwell (1997) and Wagner et al. (2002) study 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 (55), 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 find for different countries. A third, more recent but much smaller host of papers exploits 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 investigate bilateral trade and migration between many source and destination countries. Hatzigeorgiou (2010a) examines a cross-section of 75 countries in 2000, while Egger et al. (2012) work with a set of 27 receiving OECD countries and 130 source

Figure 8. Summary of estimates in the literature
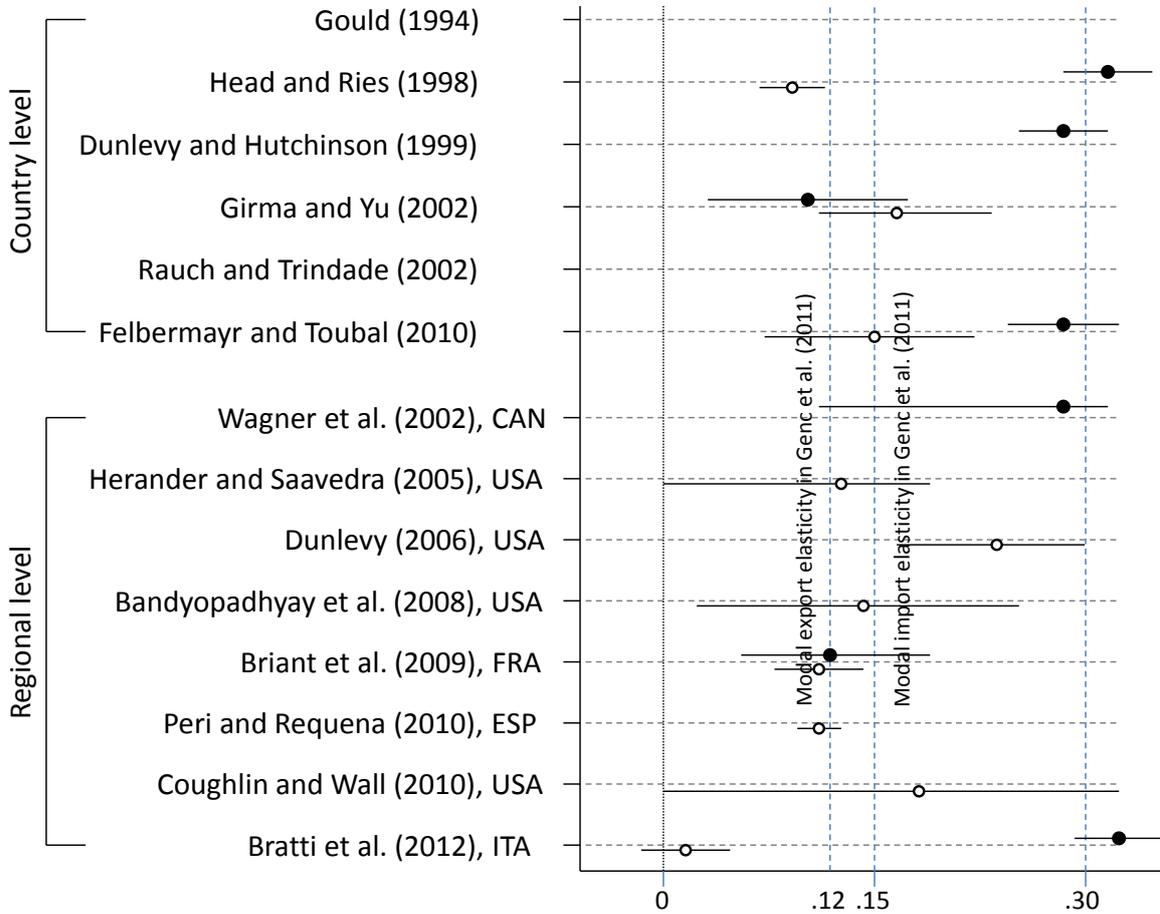
Note: Estimated elasticities of trade flows with respect to immigrant stocks in imports (black dots) or exports (white dots); 95% confidence intervals. Figure adapted from Bratti et al. (2012).

countries. Felbermayr and Jung (2009) make use of a panel of country-pairs, covering North-South trade and migration links for the years 1990 and 2000. Parsons (2012) employs a large panel of bilateral trade flows and stocks of migrants covering the years 1960, 1970, 1980, 1990, and 2000. Equation (55) 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 study the effect of the Chinese ethnic network on bilateral trade. They find 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 which we will revisit below.

Wrapping up, the literature finds positive, statistically significant effects of migrant networks on trade. Figure 8 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 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 (62) 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) criticizes 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) argues "*reservations persist as to the potential for other, unobserved phenomena to be stimulating both trade and*

110

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 experiment 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.

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) document 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 data bases that report bilateral migrant stocks for a number of countries over time. Ozden et al. (2012) have presented the most comprehensive data set 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) merges these data with trade data from Feenstra et al. (2005). The trade data is, however, spanning a smaller sample of countries than the migration data,

so that the author ends up working with 178 countries at most. Moreover, the trade data does not contain the year 1960. The author uses a model of the type described in equation (62) but augmented by country-pair fixed effects and exporter × year as well as importer × year dummies. However, his exercise leads to disappointing results. In the presence of pair-effects, 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 does 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 allocates their trade-enhancing effect to migrants.

One can have different 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 has many advantages since it has been carefully cleaned. However, its coverage is substantially smaller than the IMF's Direction of Trade (DoT) data set.

Table 2 shows regression results based on the Ozden et al. (2012) data merged with the DoT data. Pair-effects are removed by first-differencing the data. All regressions include full sets of exporter and importer effects interacted with year dummies. Columns (1) to (5) present cross-sections for the years 1960, 1970, 1980, 1990, 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 a doubling the number of immigrants in a country from some source country would increase imports

112

Table 2. Cross-country evidence: cross-sections versus panel estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Census years	1960	1970	1980	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 GDPi GDPj	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-squared	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.

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 deviations, 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 is 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). It implies that migrants have a larger effect on imports than on exports. This is well in line with results documented in Figure 8.

Columns (6) and (7) in Table 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% $(2^{0.059} = 1.042)$. This is still no trivial effect. The differences relative to Parsons (2012) have two origins: first, a larger data set 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 than, 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 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 concerning reverse causality. However, Wooldridge (2002) recommends 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) apply the test in a two-period model of North-South trade and migration and find 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 2.

Instrumental variables strategies. Recently, some authors have used instrumental variables techniques in gravity models of the type given by (62). 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 regress 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 foreign countries residing in that Spanish province at 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 data set 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.

Peri and Requena-Silvente (2010) instrument the changes in immigrants in a particular province by imputing net inflows of immigrants based on historical data. More precisely, they use the distribution of immigrants by nationality and across provinces from 1993, i.e., before the strong rise in immigration to Spain, and attribute 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 wit 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 8, but certainly on the low side.

115

The instrumental variables strategy has been used in subsequent research, for example by Bratti et al. (2012), who use province-level data for Italy, covering the period 2002-2009 during which Italy experience 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 8. 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 to become statistically indistinguishable from zero, while the effect of immigrants on imports goes up 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 (55), 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, potential informational asymmetries 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 (62) above, and writing square brackets to denote a functional dependence, the empirical estimate of the network $\bar{\nu}^i := (\sigma - 1)(\nu_a + \nu_T^{im})$. In light of the above discussion, it is very reasonable to assume that ν_T^{im} depends negatively on σ. Hence, the estimate of $\bar{\nu}^i$ 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 good. 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) classify goods in a way directly related to estimates of σ taken from Broda and Weinstein (2006).

Cross-country evidence. Table 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 immigrants in the destination country and the share of emigrants 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 8, since those refer to log levels of migrants obtained in models with only one migrant variable on the right hand side. The standardized beta coefficients to the estimates (0.278 for immigrants and

117

Table 3. The effect of emigrants and immigrants on imports across different classes of goods

Dependent variable: log 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)
R2	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).

0.154 for emigrants) are 0.08 and 0.05, respectively.[123] 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 emigrants may help overcome lack of information or informational asymmetries. However, different to emigrants, immigrants may have a special preference for goods from their home countries, so that the difference between the estimated coefficients

---

[123]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 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.

may be a 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 find 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 3 imply trade costs savings that are at least twice as big for differentiated goods than for homogeneous goods (in ad valorem tariff equivalents).

Region-level evidence. Peri and Requena-Silvente (2010) run separate regressions for trade flow aggregates made up by 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 two and a half across the high and the low differentiated goods.[124]

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.

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 overcom-

---

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

ing informational frictions between their host and source countries they must at least be able to secure themselves jobs in their countries of residence which allow them to capitalize their knowledge. Many unskilled immigrants occupy jobs that do not provide them with this option, for example, 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 migrans: 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 micro-data is still missing. More work in this area is very much welcome.

Immigrant education. Columns (4) to (6) in Table 3 presents 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 more than double in column (4). Hence, a one-point increase in the share of high-skilled migrants has twice as strong a trade creating effect that 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 one 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 by Herander and Saavedra (2005). The trade-promoting effects of migrant networks is larger the better the ability of that group to receive and process information on trading opportunities.

Immigrant occupation. Aleksynska and Peri (2011) provide the first study that differentiates immigrants according to their occupations in the host country. They use a new data set provided by the OECD which covers 89 destination countries and up to 233 countries of origin. The data refers to the years 2000-2002 and provide information on immigrants' age, gender, schooling level, labor market status and occupation at the 1-digit ISCO classification. These variables are merged with standard trade data and gravity covariates.

Their gravity equation follows (62). 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 enterprizes, as well as on groups 5 and 9, which contain salespersons 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) document the fol-

lowing 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 8. 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 and beyond that of the total migrants. Every else 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 include the shares of immigrants with primary, secondary and tertiary education into their regression along with the share of immigrants in business network occupations. They find a significant trade-enhancing effect of the highest education group, but no effect of other skill classes. They continue to report a strong effect of the business network shares. Differentiating between educational categories within their occupation classes, they argue 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 find it worthwhile to export to foreign markets. The paper by Chaney (2008) derives a gravity equation for the Melitz (2003) model where firms are assumed to sample their time-invariant productivities from a Pareto distribution. He shows 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 by Melitz (2003) can also help understand the occurrence of country pairs which 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 propose 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.

Evidence based on aggregate region-level trade flows. Coughlin and Wall (2011) use data on manufacturing exports of 48 US states (Hawai 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 have 47,776 observations. 7296 of these involve 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 have panel data, they include country-pair effects into their regressions to control for unobserved sources of heterogeneity; see below for more discussion. They run 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 run a log-linear fixed effects model combining the two margins. For the log of exports to be defined for zero trade flows, they add one to each export value in the data. They report 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 find 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 by Melitz (2003), such costs would take the form of fixed market access costs rather than of variable costs.

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 which 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 (62), 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 find that immigrants effect 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.[125]

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 ex-

---

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

ist, 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 straight-forward 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) uses 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 estimates Tobit models augmented by state and country fixed effects on a cross-section of around the year 1990.

Dunlevy (2006) finds 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 if 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 one of the source country is Ireland, Canada, Australia, New Zealand or the United Kingdom and zero else, 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 finds 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 current 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 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

The Chinese network.   The most prominent evidence for market-replacing networks comes from the study by Rauch and Trindade (2002) who investigate how ethnic Chinese minorities residing in different countries promote trade between these countries. They characterize 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 aim 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 (1960)-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 overcoming 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 do 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 exploit the fact that the ethnic Chinese network also spans countries different from China. They consider 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 (62) in several respects: first, they estimate a threshold Tobit model to account for zero trade flows. Second, they use the log sum of imports and exports as the dependent variable. Third, they do not control for multilateral resistance terms and they do not include country dummies into 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 draw 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 find that CHINSHARE has a strong positive effect on bilateral trade in both 1980 and 1990 and across goods classes made up 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 as big than that for goods traded on exchanges. Next, the authors differentiate 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 trade creating effect of the Chinese network is by 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) show that this strong trade creation effect translates into an ad valorem tariff reduction equivalent of about 6%.

Indirect network effects. The empirical strategy of Rauch and Trindade (2002) reflects the methodological state of art of 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) revisit the evidence using such a revamped gravity framework. Rather than focussing on strong versus weak

networks, they make a clearer distinction between direct and indirect links.

Table 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, however, never run regressions of aggregate trade. The estimate amount 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 five, 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 one if the bilateral relationship involves China and zero otherwise. Evaluated at the respective sample means, the effect of direct links is associated to substantial trade creation (4.7%), while trade creation due to indirect links is very minor (less than 0.1%). Columns (4) 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 4 cannot be directly compared to Rauch and Trindade (2002), since they do not further differentiate between strong and weak networks and between different commodity groups. Felbermayr et al. (2010b) provide estimates that incorporate these features. They find that tariff equivalents are larger for differentiated than for homogeneous goods. The tariff equivalent of a strong network with differentiated goods is between 2 to 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 focussing on strong direct links. Here, in 1990 and for differentiated goods, the 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

Table 4. The direct and indirect trade effects of the Chinese 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
R2	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: $N = 2,520$ in 1980, $N = 2,795$ in 1990, and $N = 3,259$ 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).

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) find evidence for several other co-ethnic networks. Average tariff equivalents as measures of the value of those networks are, however, never exceed 0.1%. Interestingly, the highest value is associated to the Maroccan network, followed by the Ghanaese and Danish ones.

Vézina (2012) also investigate the interaction between the immigrant network and source country institutions in a cross-section of Swiss exports in homogeneous, reference-priced and differentiated goods to 177 foreign countries. Defining the extensive margin as the number of SITC 4-digit product lines per partner and the intensive margin as the average value of exports, he employs an instrumental variables strategy based on immigration patterns to France an on visa restrictions used by Switzerland. His measure of institutions is the corruption index from Kaufmann and Mastruzzi (2010). He finds that the elasticity of exports with respect to immigrants is highest in the sector of differentiated goods and if the export partner country is highly corrupt. That effect, however, is operative only on the extensive margin. The intensive margin is unaffected by the migrant network, regardless of the degree of product differentiation and of export partner country institutions.

7 Summary

Migration is the archetype form of globalization and the oldest action against poverty. Modern history has witnessed two big waves of migration: The era of mass migration towards the end of the 19th century until the outbreak of the Great War, and the recent wave of economic globalization more or less exactly a hundred years later. In both episodes, migration was accompanied by a significant increase also of trade as well as capital flows between the sending and receiving countries of migration.

In this chapter, we have reviewed both empirical and theoretical research on the possible relationships between these three forms of globalization. We started out by means of a brief tour through history, comparing the key characteristics of the two waves of economic globalization, thereby also identifying the key issues that modern

economic research on migration has been addressing.

Economic historians exploring trade and factor movements between countries of the 19th century "Atlantic economy" tend to conclude a complementarity relationship between migration and other forms of globalization, trade and capital movements. Moreover, they conclude that migration has caused significant convergence between incomes per capita and wages paid in sending and receiving countries of migration. Migration in the second half of the 20th century was much different. The direction of migration flows had changed, with Central Europe becoming a major destination of migration flows, and with countries in Europe as well as in North America running restrictive immigration policies. As regards the issues addressed, modern economics has focused on the effects of migration on welfare effects, on international convergence as well as internal distribution the income, and the role of migration as an ingredient of policies to fight world poverty. Modern economics has also questioned the conventional view, mainly inherited from trade theory developed in the interwar period by Heckscher and Ohlin, that in almost all of these key issues factor movements and trade are substitutes.

In the third section of this chapter, we have reviewed the neoclassical part of this modern theory. It is separated from subsequent sections by focusing on static models, leaving the relationship between trade and capital formation to the two subsequent sections. Moreover, it assumes that barriers to migration and trade are exogenous, leaving the analysis of network effects through which migration may enhance trade to the final section. Our model focuses on the effects of migration flows between two countries, allowing for a disaggregated view of migration that distinguishes between different types of labor, and emphasizing the existence of migration stocks reflecting previous migration flows. The equilibria considered are determined by (changing) costs of international migration. We have drawn a major line of division between a world characterized by constant returns to scale and perfect competition, and a world of internal economies of scale and monopolistic competition.

We have developed a simple, yet general model of migration under constant returns in order to take a normative view on migration. We have identified three principal welfare channels which help us understand not just the effects of migration, but also the

forces behind migration policies and the potential conflict between objectives that may be pursued by sending and receiving countries, respectively. The first channel involves a first-order effect that derives from discrepancies across countries in the marginal productivities of different types of labor. The direction of this effect depends on the correlation between these discrepancies and the amounts of migration flows of different types of labor. The second channel involves a second-order effect that derives from the pre-existing stocks of immigrants present in the two countries and how these are correlated with the wage effects of migration. This is often referred to as the immigration surplus. Finally, the third channel derives from terms of trade changes that are brought about by international migration.

Looking at these welfare channels from a policy perspective, we find that there is a potential for conflict in that the sending and receiving country cannot simultaneously fine-tune their policies in order to achieve a maximum welfare gain from migration. This potential is aggravated if we introduce distortions, i.e., wages that deviate from the respective marginal productivities. As regards internal distribution within a country, we have emphasized a principal difference between migration and trade: Where redistributive policies are available for governments to achieve a Pareto improvement from gains from trade, such policies are impossible for migration gains, unless governments in immigration countries are willing to discriminate against immigrants. Moreover, although there is principal parallel between migration policies and policies of piecemeal trade liberalization, a welfare-increasing policy reform seems significantly more difficult to attain for migration than for trade.

As regards the question of complementarity or substitutability between trade and migration, the pivotal question is whether or not countries share a common technology as typically assumed by Heckscher-Ohlin trade models. If trade is driven by non-endowment based pattern of comparative advantage, say technological differences, and if traded goods feature different factor intensities, then trade leads to factor price divergence, which in turn may cause migration, adding a factor-endowment basis to trade, thus leading to a complementarity relationship. While migration and trade still both lead to convergence, with complementarity full convergence needs a lot more factor migration than under substitutability.

A pivotal question for whether or not migration induces convergence is whether or not technology is convex. Conventional trade theory typically assumes that it is, whence migration causes convergence. In contrast, new models of trade allow for non-convex technology through fixed cost. In such a world, migration no longer causes convergence, but may unleash forces of divergence, as emphasized in the literature on the new economic geography. Readdressing the question of substitutability versus complementarity in models of new economic geography, one may follow two distinct approaches. Focusing on migrant stocks and trade flows in the potential stable equilibria, one observes ambiguity. But focusing on migration flows and trade in the adjustment process, one finds strong complementarity.

The dynamic interaction between migration and capital formation is all but trivial. In the fourth section, we focused the discussion on 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 causally higher immigration (emigration) leads to increased (reduced) capital accumulation. This is driven by the complementarity between labor and capital in the production function. Under constant returns, if the capital stock and therefore wages are initially low, labor market integration leads first to emigration and decreased (but still positive) capital accumulation. We may thus observe emigration and capital inflows at the same time. Later in the transition, there is immigration (i.e. reversed migration) while the capital stock approaches its steady state level. Distinguishing tradable and non-tradable goods allows us to derive novel welfare effects. If the non-tradable goods sector is land-intensive, for instance, capturing the housing sector, immigration may reduce individual welfare for individuals with a low land-endowment. The reason is that immigration drives up the relative price for non-tradables, as land is a fixed factor which 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 shifts the focus from wage effects of migration, which are typically in the center of the immigration debate albeit typically found to be rather low empirically, 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 strong or weak scale effects with respect to the growth rate or the level of GDP per capita. As a result of such scale effects, selective migration towards high-skilled labor may affect the direction of technical change towards innovations which raise the productivity of skilled relative to unskilled labor. However, if there are congestion effects from higher population density on total factor productivity, immigration may mitigate technological progress despite positive scale effects which run through an increased specialization of labor. Finally, we related recent research on possibly positive effects of brain drain on the domestic human capital stock of source economies to the direction of physical capital flows. We argued that in the case where a less restrictive immigration policy in advanced countries significantly raises human capital formation in source countries by improving emigration prospects of skilled workers, higher emigration may go along with capital inflows. However, empirical evidence on a positive effect of higher emigration prospects on the domestic human capital level suggests that such a gain from brain drain is observed for a minority of developing countries under specific circumstances only.

In the sixth section we have reviewed the empirical literature exploring the link between migration and trade. It 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.

While immigrant networks have long played an important role in gravity models, it is only over the last 10 years that the theoretical foundations have been made clear and implications for consistent estimation have been derived. Against this backdrop, section 6 revisits the identification of the overall trade-creating effect of migration and its

break-down 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 clarifies the role of product differentiation for the size of estimated effects, discussing the role of immigrants' education and occupation, and highlighting the distinction between the intensive and the extensive margin, respectively.

Finally, section 6 also reviews the role of trading partner characteristics, and, finally, also distinguishes 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. On the basis of received evidence we conclude that both channels are present in the data and that the preference channel may account for as much as half of the total trade creating effect of immigration.

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 $p(i) = \frac{r}{\alpha}$ with

$$p(i) = \alpha B \left(\frac{A(i)}{x(i)} \right)^{1-\alpha} (L^Y)^\beta Z^{1-\alpha-\beta} \left[= \frac{\partial Y}{\partial x(i)} \right] \tag{63}$$

and solving for $x(i)$ implies

$$x(i) = A(i) \left(B \frac{\alpha^2}{r} (L^Y)^\beta Z^{1-\alpha-\beta} \right)^{\frac{1}{1-\alpha}}. \tag{64}$$

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

$$\Pi_{t+1}(i) \equiv \frac{\pi_{t+1}(i)}{1+r} - w_t l_t(i) - w_t f, \tag{65}$$

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

$$\pi_{t+1}(i) = (1-\alpha) r^{-\frac{\alpha}{1-\alpha}} \alpha^{\frac{1+\alpha}{1-\alpha}} B^{\frac{1}{1-\alpha}} \bar{A}_t \Lambda(l_t(i)) (L^Y)^{\frac{\beta}{1-\alpha}} Z^{\frac{1-\alpha-\beta}{1-\alpha}}, \tag{66}$$

according to $p(i) = \frac{r}{\alpha}$, (64) and (45). Using (66) in (65), the first-order condition for the optimal choice of R&D input $l_t(i)$ is

$$\chi \Lambda'(l(i)) B^{\frac{1}{1-\alpha}} (L^Y)^{\frac{\beta}{1-\alpha}} Z^{\frac{1-\alpha-\beta}{1-\alpha}} = \omega, \tag{67}$$

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

$$\chi \Lambda(l(i)) B^{\frac{1}{1-\alpha}} (L^Y)^{\frac{\beta}{1-\alpha}} Z^{\frac{1-\alpha-\beta}{1-\alpha}} = (l(i) + f)\omega. \tag{68}$$

Combining (67) and (68), the equilibrium R&D labor input of each firm, $\tilde{l}$, is time-

invariant and uniquely given by

$$0 = \frac{\Lambda(\tilde{l})}{\Lambda'(\tilde{l})} - \tilde{l} - f. \tag{69}$$

Moreover, the wage rate is given by $w = \frac{\partial Y}{\partial L^Y} = (1 - \alpha)\frac{y}{l^Y}$. Using (44) and (46), we find

$$\omega = (1 - \alpha)\left(\frac{\alpha^2}{r}\right)^{\frac{\alpha}{1-\alpha}} B^{\frac{1}{1-\alpha}} \left(\frac{Z}{L^Y}\right)^{\frac{1-\alpha-\beta}{1-\alpha}} N. \tag{70}$$

Combining (67) and (70) then implies

$$N_{t+1} = \frac{\alpha \Lambda'(\tilde{l})}{1+r} L^Y_{t+1}. \tag{71}$$

In labor market equilibrium,

$$L^Y_t + \int_0^{N_{t+1}} (l_t(i) + f)\mathrm{d}i = L_t. \tag{72}$$

Using $l(i) = \tilde{l}$ for all i, (69), (71) and $L_{t+1} = (1 + n)L_t$, we see that the fraction of labor devoted to manufacturing evolves according to

$$l^Y_{t+1} = \xi \cdot (1 - l^Y_t). \tag{73}$$

$\xi \equiv \frac{1+r}{\alpha \Lambda(\tilde{l})(1+n)}$, where $l^Y_0 = \frac{L^Y_0}{L}$ 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 (71) and (73). Denote the steady state value of l^Y by $\tilde{l}^Y$. According to (73), we have $\tilde{l}^Y = \frac{\xi}{1+\kappa}$. The steady state is globally stable if $\xi < 1$, which is well possible (recall that $\Lambda(0) = 1$ and $\Lambda' > 0$).

References

Abel, Andrew B., "Dynamic Effects of Permanent and Temporary Tax Policies in a q Model of Investment," *Journal of Monetary Economics*, 1982, *9* (3), 353–373.

Acemoglu, Daron, "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics*, 1998, *113*, 1055–1089.

_ , "Directed Technical Change," *Review of Economic Studies*, October 2002, *69* (4), 781–809.

_ **and Joshua Angrist**, "How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws," *NBER Macroeconomics Annual*, April 2001, *15*, 9–74.

_ **and Simon Johnson**, "Disease and Development: The Effect of Life Expectancy on Economic Growth," *Journal of Political Economy*, December 2007, *115* (6), 925–985.

Aghion, Philippe and Jeffrey G. Williamson, *Growth, inequality, and globalization: theory, history, and policy*, Cambridge: Cambridge University Press, 1998.

Aleksynska, Mariya and Giovanni Peri, "Isolating the Network Effect of Immigrants on Trade," Working Papers 2011-26, CEPII research center December 2011.

Anderson, James E., "The Gravity Model," *Annual Review of Economics*, 2011, *3*, 133–160.

_ **and Eric van Wincoop**, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, March 2003, *93* (1), 170–192.

_ **and** _ , "Trade Costs," *Journal of Economic Literature*, September 2004, *42* (3), 691–751.

Arrow, Kenneth J., "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, June 1962, *29* (3), 155–173.

Aydemir, Abdurrahman and George J. Borjas, "Cross-country Variation in the Impact of International Migration: Canada, Mexico, and the United States," *Journal of the European Economic Association*, 2007, *5* (4), 663–708.

_ **and** _ , "Attenuation Bias in Measuring the Wage Impact of Immigration," *Journal of Labor Economics*, 2011, *29* (1), 69–112.

Baier, Scott L. and Jeffrey H. Bergstrand, "Do free trade agreements actually increase members' international trade?," *Journal of International Economics*, March 2007, *71* (1), 72–95.

Baldwin, Richard E. and Philippe Martin, "Agglomeration and Regional Growth," in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Elsevier, 2004, chapter 60, pp. 2671–2711.

Bandyopadhyay, Subhayu, Cletus C. Coughlin, and Howard J. Wall, "Ethnic Networks and US Exports," *Review of International Economics*, 02 2008, *16* (1), 199–213.

Bardhan, Ashok Deo and Subhrajit Guhathakurta, "Global Linkages of Subnational Regions: Coastal Exports and International Networks," *Contemporary Economic Policy*, 04 2004, *22* (2), 225–236.

Beine, Michel, Frédéric Docquier, and Caglar Özden, "Diasporas," *Journal of Development Economics*, May 2011, *95* (1), 30–41.

_ , _ , **and Hillel Rapoport**, "Brain Drain and Economic Growth: Theory and Evidence," *Journal of Development Economics*, February 2001, *64* (1), 275–289.

_ , _ , **and** _ , "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers," *Economic Journal*, 04 2008, *118* (528), 631–652.

Bergstrand, Jeffrey H. and Peter Egger, "Gravity Equations and Economic Frictions in the World Economy," in David Greenaway Daniel Bernhofen, Rodney Falvey and Udo Krieckemeier, eds., *Palgrave Handbook of International Trade*, Palgrave-Macmillan Press, 2011.

Berman, Eli, John Bound, and Stephen Machin, "Implications of Skill-Biased Technological Change: International Evidence," *The Quarterly Journal of Economics*, 1998, *113* (4), 1245–1279.

_ , _ , **and Zvi Griliches**, "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers," *The Quarterly Journal of Economics*, 1994, *109* (2), 367–397.

Berry, R. Albert and Ronald Soligo, "Some Welfare Aspects of International Migration," *Journal of Political Economy*, 1969, *77* (5), 778–794.

Bhagwati, Jagdish and Koichi Hamada, "The Brain Drain, International Integration of Markets for Professionals and Unemployment : A Theoretical Analysis," *Journal of Development Economics*, April 1974, *1* (1), 19–42.

Bhawgwati, Jagdish N., *International Factor Mobility: Essays in International Economic Theory*, Cambridge, Mass.: MIT Press, 1983.

Borjas, George, "The Economic Analysis of Immigration," in O. Ashenfelter and D. Card, eds., *Handbook of Population Economics, vol. 3A*, Elsevier, 1999, pp. 1697–1760.

_ , "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *The Quarterly Journal of Economics*, 2003, *118* (4), 1335–1374.

Borjas, George J., "Self-Selection and the Earnings of Immigrants.," *American Economic Review*, 1987, *77* (4), 531–553.

___ , **Jeffrey Grogger, and Gordon H. Hanson**, "Comment: On Estimating Elasticities of Substitution," *Journal of the European Economic Association*, 2012, *10* (1), 198–210.

___ , **Richard B. Freeman, and Lawrence F. Katz**, "How Much Do Immigration and Trade Affect Labor Market Outcomes?," *Brookings Papers on Economic Activity*, 1997, *1997* (1), 1–90.

Bratti, Massimiliano, Luca De Benedictis, and Gianluca Santoni, "On the Pro-Trade Effects of Immigrants," IZA Discussion Papers 6628, Institute for the Study of Labor (IZA) June 2012.

Braun, Juan, "Essays on Economic Growth and Migration." PhD dissertation, Harvard University, Cambridge MA 1993.

Briant, Anthony, Pierre-Philippe Combes, and Miren Lafourcade, "Product Complexity, Quality of Institutions and the Pro-Trade Effect of Immigrants," CEPR Discussion Papers 7192 March 2009.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, May 2006, *121* (2), 541–585.

Bruder, J., "Are Trade and Migration Substitutes or Complements? The Case of Germany," Working Paper, University of Rostock April 2004.

Bryant, John, Murat Genc, and David Law, "Trade and Migration to New Zealand," Treasury Working Paper Series 04/18, New Zealand Treasury September 2004.

Buch, Claudia M., Jörn Kleinert, and Farid Toubal, "Where Enterprises Lead, People Follow?: Links between Migration and FDI in Germany," *European Economic Review*, November 2006, *50* (8), 2017–2036.

Burda, Michael, "Factor Reallocation in Eastern Germany after Reunification," *American Economic Review*, May 2006, *96* (2), 368–374.

___ **and Charles Wyplosz**, "Human Capital, Investment and Migration in an Integrated Europe," *European Economic Review*, April 1992, *36* (2-3), 677–684.

Burstein, Ariel and Jonathan Vogel, "International Trade, Technology, and the Skill Premium," *Working Paper*, 2012.

Card, David, "The Impact of the Mariel Boatlift on the Miami Labor-Market," *Industrial and Labor Relations Review*, 1990, *43* (2), 245–257.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, September 2008, *98* (4), 1707–21.

Chiquiar, Daniel and Gordon H. Hanson, "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 2005, *113* (2), pp. 239–281.

Chiswick, Barry P., "Are Immigrants Favorably Self-Selected?," *American Economic Review*, 1999, *89* (2), 181–185.

_ , "Are Immigrants Favorably Self-Selected? An Economic Analysis," in Caroline D. Brettell and James F. Hollifield, eds., *Migration Theory: Talking Across Disciplines*, Routledge, 2000.

Ciccone, Antonio and Giovanni Peri, "Identifying Human-Capital Externalities: Theory with Applications," *Review of Economic Studies*, 04 2006, *73* (2), 381–412.

Cline, William, *Trade and Income Distribution*, Washington: Institute for International Economics, 1997.

Co, Catherine, Patricia Euzent, and Thomas Martin, "The export effect of immigration into the USA," *Applied Economics*, 2004, *36* (6), 573–583.

Coe, David T. and Elhanan Helpman, "International R&D Spillovers," *European Economic Review*, May 1995, *39* (5), 859–887.

Combes, Pierre-Philippe, Miren Lafourcade, and Thierry Mayer, "The trade-creating effects of business and social networks: evidence from France," *Journal of International Economics*, May 2005, *66* (1), 1–29.

Coughlin, Cletus C. and Howard J. Wall, "Ethnic networks and trade: Intensive versus extensive margins," *Economics Letters*, October 2011, *113* (1), 73–75.

Daniel, Aart Kraay Kaufmann and Massimo Mastruzzi, "The Worldwide Governance Indicators: Methodology and Analytical Issues," World Bank Policy Research Working Paper 54, World Bank 2010.

Davis, Donald and David E. Weinstein, "Technological Superiority and the Losses from Migration," *NBER Working Paper 8971*, 2002.

Davis, Donald R., "Does European Unemployment Prop up American Wages? National Labor Markets and Global Trade," *The American Economic Review*, 1998, *88* (3), 478–494.

Deardorff, Alan V., "Weak links in the chain of comparative advantage," *Journal of International Economics*, 1979, *9* (2), 197 – 209.

_ , "The General Validity of the Heckscher-Ohlin Theorem," *The American Economic Review*, 1982, *72* (4), 683–694.

_ , "Factor Prices and the Factor Content of Trade Revisited: What is the Use?," *Journal of International Economics*, 2000, *50* (1).

_ **and Robert W. Staiger**, "An interpretation of the factor content of trade," *Journal of International Economics*, 1988, *24* (1Ũ2), 93–107.

Diamond, Jared M., *Guns, germs and steel: the fates of human societies*, London: Cape, 1997.

Dixit, Avinash and Alan Woodland, "The relationship between factor endowments and commodity trade," *Journal of International Economics*, 1982, *13* (3Ũ4), 201 – 214.

_ **and Victor Norman**, "Gains from trade without lump-sum compensation," *Journal of International Economics*, 1986, *21* (1Ũ2), 111–122.

Dixit, Avinash K. and Joseph E. Stiglitz, "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 1977, *67* (3), 297–308.

_ **and Victor D. Norman**, *Theory of international trade : a dual, general equilibrium approach*, Cambridge: Cambridge University Press, 1980.

Dornbusch, Rudiger., Stanley Fischer, and Paul. A. Samuelson, "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods," *The American Economic Review*, 1977, *67* (5), 823–839.

Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson, "Heckscher-Ohlin Trade Theory with a Continuum of Goods," *The Quarterly Journal of Economics*, 1980, *95* (2), 203–224.

Dunlevy, James A., "The Influence of Corruption and Language on the Protrade Effect of Immigrants: Evidence from the American States," *The Review of Economics and Statistics*, February 2006, *88* (1), 182–186.

_ **and William K. Hutchinson**, "The Impact of Immigration on American Import Trade in the Late Nineteenth and Early Twentieth Centuries," *The Journal of Economic History*, December 1999, *59* (04), 1043–1062.

Dustmann, Christian, Francesca Fabbri, and Ian Preston, "The Impact of Immigration on the British Labour Market," *The Economic Journal*, 2005, *115* (507), F324–F341.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, September 2002, *70* (5), 1741–1779.

Egger, Peter, Maximilian von Ehrlich, and Douglas Nelson, "Migration and Trade," *The World Economy*, 2012, *35* (2), 216–241.

Epifani, Paolo and Gino Gancia, "The Skill Bias of World Trade," *The Economic Journal*, 2008, *118*, 927–960.

Ethier, Wilfred J., "The general role of factor intensity in the theorems of international trade," *Economics Letters*, 1982, *10* (3Ũ4), 337–342.

_ **and Lars E.O. Svensson**, "The theoremes of international trade with factor mobility," *Journal of International Economics*, 1986, *20* (1Ũ2), 21–42.

Faini, Riccardo, "Increasing Returns, Migrations and Convergence," *Journal of Development Economics*, April 1996, *49* (1), 121–136.

Feenstra, Rober, *Advanced International Trade: Theory and Evidence*, Princeton University Press, 2004.

Feenstra, Robert C., *Advanced international trade : theory and evidence*, Princeton, NJ: Princeton University Press, 2004.

_ **and Gordon H. Hanson**, "Foreign Investment, Outsourcing and Relative Wages," in "The Political Economy of Trade. Policy: Papers in Honor of Jagdish Bhagwati," The MIT Press, 1996.

_ **and** _ , "Foreign direct investment and relative wages: evidence from Mexico's Maquiladoras," *Journal of International Economics*, 1997, *42*, 371–393.

_ **and** _ , "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979-1990," *Quarterly Journal of Economics*, 1999, *114* (3), 907–940.

_ , **Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo**, "World Trade Flows: 1962-2000," NBER Working Papers 11040, National Bureau of Economic Research, Inc January 2005.

Felbermayr, Gabirel J., Wido Geis, and Wilhlem Kohler, "Restrictive Immigration Policy In Germany: Pains and Gains Foregone?," *Review of World Economics*, 2010, *146* (1), 1–21.

Felbermayr, Gabriel J. and Benjamin Jung, "The pro-trade effect of the brain drain: Sorting out confounding factors," *Economics Letters*, August 2009, *104* (2), 72–75.

_ **and Farid Toubal**, "Revisiting the Trade-Migration Nexus: Evidence from New OECD Data," *World Development*, 2012, *40* (5), 928–937.

_ **and Wilhelm Kohler**, "Exploring the Intensive and Extensive Margins of World Trade," *Review of World Economics*, December 2006, *142* (4), 642–674.

_ **and** _ , "Immigration and Wages in General Equilibrium: A Theoretical Perspective," in F. Foders and R.J. Langhammer, eds., *Labor Mobility and the World Economy*, Springer-Verlag, 2006, pp. 51–80.

_ **and** _ , "Immigration and Native Welfare," *International Economic Review*, 2007, *48* (3), 731–760.

_ **and** _ , "Can International Migration Ever Be Made a Pareto Improvement?," in Mooslechner P. Nowotny, E. and D. Ritzberger-Grünwald, eds., *The Integration of European Labour Markets*, Edward Elgar, 2009, pp. 32–49.

_ , **Benjamin Jung, and Farid Toubal**, "Ethnic Networks, Information, and International Trade: Revisiting the Evidence," *Annales d'Economie et de Statistique*, 2010, (97-98), 41–70.

Freeman, Richard B., "People Flows in Globalization," *Journal of Economic Perspectives*, June 2006, *20* (2), 145–170.

Friedberg, Rachel M., "The Impact of Mass Migration on the Israeli Labor Market," *The Quarterly Journal of Economics*, 2001, *116* (4), 1373–1408.

_ **and Jenifer Hunt**, "The Impact of Immigrants on Host Country Wages, Employment and Growth," *Journal of Economic Perspectives*, June 1995, *9* (2), 23–44.

Fujita, Masahisa and Jacques-François Thisse, *Economics of agglomeration : cities, industrial location, and regional growth*, Cambridge: Cambridge University Press, 2002.

_ , **Paul Krugman, and Anthony J. Venables**, *The Spatial Economy: Cities, Regions, and International Trade*, Vol. 1 of *MIT Press Books*, The MIT Press, 2001.

_ , **Paul R. Krugman, and Anthony J. Venables**, *The spatial economy : cities, regions, and international trade*, Cambridge, Mass.: MIT Press, 2001.

Galbraith, John Kenneth, *The nature of mass poverty*, Cambridge, Mass.: Harvard University Press, 1979.

Gao, Ting, "Ethnic Chinese Networks and International Investment: Evidence from Inward FDI in China," *Journal of Asian Economics*, August 2003, *14* (4), 611–629.

Gaston, Noel and Douglas R. Nelson, "Bridging Trade Theory and Labor Econometrics: The Effects of International Migration," *Journal of Economic Surveys*, 2011, *Early view, doi: 10.1111/j.1467-6419.2011.00696.x.*

_ **and** _ , "International Migration," in Falvey Rod Greenaway David Bernhofen, Daniel M. and Udo Kreickemeier, eds., *Palgrave handbook of international trade*, Palgrave Macmillan, 2011, pp. 660–697.

Genc, Murat, Masood Gheasi, Peter Nijkamp, and Jacques Poot, "The Impact of Immigration on International Trade: A Meta-Analysis," IZA Discussion Papers 6145, Institute for the Study of Labor (IZA) November 2011.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "Human Capital and Regional Development," NBER Working Papers, National Bureau of Economic Research, Inc June 2011.

Ghatak, Subrata, Monica Ioana Pop-Silaghi, and Vince Daly, "Trade and migration flows between some CEE countries and the UK," *Journal of International Trade & Economic Development*, 2009, *18* (1), 61–78.

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

Gould, David M, "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows," *The Review of Economics and Statistics*, May 1994, *76* (2), 302–16.

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

_ **and** _ , "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, May 2011, *95* (1), 42–57.

Grossman, Gene M. and Esteban Rossi-Hansberg, "External Economies and International Trade Redux," *Quarterly Journal of Economics*, 2010, *125*, 829–858.

Grossmann, Volker, "Entrepreneurial Innovation and Economic Growth," *Journal of Macroeconomics*, December 2009, *31* (4), 602–613.

_ , "Structural Change, Urban Congestion, and the End of Growth," *Review of Development Economics*, 2013, *forthcoming*.

_ **and David Stadelmann**, "Does International Mobility of High-Skilled Workers Aggravate Between-Country Inequality?," *Journal of Development Economics*, May 2011, *95* (1), 88–94.

_ **and** _ , "Does High-Skilled Migration Affect Publicly Financed Investments?," *Review of International Economics*, 2012, *forthcoming*.

_ **and** _ , "Wage Effects of High-Skilled Migration: International Evidence," *World Bank Economic Review*, 2012, *forthcoming*.

_ **, Andreas Schäfer, and Thomas M.Steger**, "Immigration and Capital Formation: The Role of Non-Tradable Goods and Landownership," Technical Report, University of Fribourg and University of Leipzig (mimeo) 2012.

Hanson, Gordon, "International Migration and the Developing World," in "Handbook of Development Economics" 2010, chapter 66, pp. 4363–4414.

Hanson, Gordon H., "Illegal Migration from Mexico to the United States," *Journal of Economic Literature*, 2006, *44* (4), 869–924.

_ , *The Economics and Policy of Illegal Immigration in the United States*, Washington, D.C.: Migration Policy Institute, 2009.

Harrison, Ann, John McLaren, and Margaret McMillan, "Recent Perspectives on Trade and Inequality," *Annual Review of Economics*, 2011, *3* (1), 261–289.

Hatton, Timothy J., "Should We Have a WTO for International Migration?," *Economic Policy*, 2007, *22* (50), 339–383.

_ **and Jeffrey G. Williamson**, *The age of mass migration: causes and economic impact*, New York: Oxford University Press, 1998.

145

_ **and** _ , *Global migration and the world economy: two centuries of policy and performance*, Cambridge, Mass.: MIT Press, 2005.

_ **and** _ , "International MIgration in the Long Run: Positive Selection, Negative Selection, and Policy," in F. Foders and R.J. Langhammer, eds., *Labor Mobility and the World Economy*, Springer-Verlag, 2006.

Hatzigeorgiou, Andreas, "Does Immigration Stimulate Foreign Trade? Evidence from Sweden," *Journal of Economic Integration*, 2010, *25*, 376–402.

_ , "Migration as Trade Facilitation: Assessing the Links between International Trade and Migration," *The B.E. Journal of Economic Analysis & Policy*, 2010, *10* (1), 24.

Hayashi, Fumio, "Tobin's Marginal q and Average q: A Neoclassical Interpretation," *Econometrica*, January 1982, *50* (1), 213–24.

Head, Keith and John Ries, "Immigration and Trade Creation: Econometric Evidence from Canada," *Canadian Journal of Economics*, February 1998, *31* (1), 47–62.

Helliwell, John F., "National Borders, Trade and Migration," *Pacific Economic Review*, 1997, *2* (3), 165–185.

Helpman, Elhanan and Paul R. Krugman, *Market structure and foreign trade : increasing returns, imperfect competition, and the international economy*, Cambridge, Mass.: MIT Press, 1985.

_ , **Marc Melitz, and Yona Rubinstein**, "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 05 2008, *123* (2), 441–487.

Hendricks, Lutz, "How Important Is Human Capital for Development? Evidence from Immigrant Earnings," *American Economic Review*, 1 2002, *92* (1), 198–219.

Herander, Mark G. and Luz A. Saavedra, "Exports and the Structure of Immigrant-Based Networks: The Role of Geographic Proximity," *The Review of Economics and Statistics*, May 2005, *87* (2), 323–335.

Hiller, Sanne, "The Export Promoting Effect of Emigration: Evidence from Denmark," Working Papers 11-08, Aarhus University, Aarhus School of Business 2011.

Hunt, Jennifer and Marjolaine Gauthier-Loiselle, "How Much Does Immigration Boost Innovation?," *American Economic Journal: Macroeconomics*, April 2010, *2* (2), 31–56.

Iranzo, Susana and Giovanni Peri, "Schooling Externalities, Technology, and Productivity: Theory and Evidence from U.S. States," *The Review of Economics and Statistics*, May 2009, *91* (2), 420–431.

Jansen, Marion and Roberta Piermartini, "Temporary Migration and Bilateral Trade Flows," *World Economy*, 2009, *32* (5), 735–753.

Javorcik, Beata S., Caglar Özden, Mariana Spatareanu, and Cristina Neagu, "Migrant Networks and Foreign Direct Investment," *Journal of Development Economics*, March 2011, *94* (2), 231–241.

Jones, Charles I, "R&D-Based Models of Economic Growth," *Journal of Political Economy*, August 1995, *103* (4), 759–84.

Jones, Charles I., "Growth: With or Without Scale Effects?," *American Economic Review*, May 1999, *89* (2), 139–144.

_ , "Growth and Ideas," in Philippe Aghion and Steven Durlauf, eds., *Handbook of Economic Growth*, Vol. 1 of *Handbook of Economic Growth*, Elsevier, 2005, chapter 16, pp. 1063–1111.

Krugman, Paul R., "Scale Economies, Product Differentiation, and the Pattern of Trade," *The American Economic Review*, 1980, *70* (5), 950–959.

_ , "Increasing Returns and Economic Geography," *Journal of Political Economy*, 1991a, *99* (3), 483–499.

_ , "History versus Expectations," *Quarterly Journal of Economics*, 1991b, *106*, 651–677.

_ , "Growing World Trade: Causes and Consequences," *Brookings Papers on Economic Activity*, 1995, *1995* (1), 327–377.

_ , "Technology, trade and factor prices," *Journal of International Economics*, 2000, *50*, 51–71.

_ , "Trade and Wages, Reconsidered," *Brookings Papers on Economic Activity*, 2008, *2008* (1), 103–137.

Kugler, Maurice and Hillel Rapoport, "International Labor and Capital Flows: Complements or Substitutes?," *Economics Letters*, February 2007, *94* (2), 155–162.

Lalonde, Robert J. and Robert H. Topel, "Economic Impact of International Migration and the Economic Performance of Migrants," in M.R. Rosenzweig and O. Stark, eds., *Handbook of Population and Family Economics*, Elsevier, 1997, pp. 799–850.

Leamer, Edward E., "In Search of Stolper-Samuelson Linkages between International Trade and Lower Wages," in S. Collins, ed., *Imports, Exports and the American Worker*, Brookings Institution, 1997, pp. 141–214.

_ , "What's the use of factor-contents?," *Journal of International Economics*, 2000, *50(1)*, 17–49.

_ **and James A. Levinsohn**, "International Trade Theory: The Evidence," in G.M. Grossman and K. Rogoff, eds., *Handbook of International Economics, vol. III*, Elsevier, 1995, pp. 1339–1394.

Legrain, Philippe, *Immigrants: Your Country Needs Them*, London: Little, Brown, 2009.

Lucas, Robert E. Jr., "On the mechanics of economic development," *Journal of Monetary Economics*, 1988, *22* (1), 3–42.

_ , "Why doesn't Capital Flow from Rich to Poor Countries?," *American Economic Review*, 1990, *80* (2), 92–96.

_ , "Migration and Economic Development in Africa: A Review of Evidence," *Journal of African Economies*, December 2006, *15* (2), 337–395.

Lundborg, Per and Paul S. Segerstrom, "The Growth and Welfare Effects of International Mass Migration," Working Paper Series 146, Trade Union Institute for Economic Research September 1998.

Markusen, James R., "Factor movements and commodity trade as complements," *Journal of International Economics*, 1983, *14* (3-4), 341–356.

_ **and Lars E. O. Svensson**, "Trade in Goods and Factors with International Differences in Technology," *International Economic Review*, 1985, *26* (1), 175–192.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, November 2003, *71* (6), 1695–1725.

Millimet, Daniel L. and Thomas Osang, "Do state borders matter for U.S. intra-national trade? The role of history and internal migration," *Canadian Journal of Economics*, February 2007, *40* (1), 93–126.

Mishra, Prachi, "Emigration and wages in source countries: Evidence from Mexico," *Journal of Development Economics*, 2007, *82* (1), 180–199.

Miyagiwa, Kaz, "Scale Economies in Education and the Brain Drain Problem," *International Economic Review*, 1991, *32* (3), 743–759.

Mora, Toni, Esther Vayá, and Jordi Surinch, "The Enlargement of the European Union and the Spatial Distribution of Economic Activity," *Eastern European Economics*, September 2004, *42* (5), 6–35.

Mountford, Andrew, "Can a brain drain be good for growth in the source economy?," *Journal of Development Economics*, 1997, *53* (2), 287–303.

_ , "Can a Brain Drain Be Good for Growth in the Source Economy?," *Journal of Development Economics*, August 1997, *53* (2), 287–303.

Mundell, Robert A., "International Trade and Factor Mobility," *The American Economic Review*, 1957, *47* (3), 321–335.

Mundra, Kusum, "Immigration and International Trade: A Semiparametric Empirical Investigation," *Journal of International Trade & Economic Development*, 2005, *14* (1), 65–91.

Neary, J. Peter, "Pitfalls in the Theory of International Trade Policy: Concertina Reforms of Tariffs, and Subsidies to High-Technology Industries," *The Scandinavian Journal of Economics*, 1998, *100* (1), 187–206.

__ **and Albert G. Schweinberger**, "Factor Content Functions and the Theory of International Trade," *The Review of Economic Studies*, 1986, *53* (3), 421–432.

Nygaard, Christian, "International Migration, Housing Demand and Access to Homeownership in the UK," *Urban Studies*, 2011, *48* (11), 2211–2229.

OECD, *International Migration Outlook 2011*, Paris: OECD, 2011.

Ohlin, Bertil Gotthard, *Interregional and international trade*, Cambridge, Mass.: Harvard Univesity Press, 1933.

Olivero, MarŠa PŠa and Yoto V. Yotov, "Dynamic gravity: endogenous country size and asset accumulation," *Canadian Journal of Economics*, February 2012, *45* (1), 64–92.

O'Rourke, Kevin H. and Jeffrey G. Williamson, *Globalization and history: the evolution of a nineteenth century Atlantic economy*, Cambridge, Mass.: MIT Press, 1999.

Ortega, Francesc and Giovanni Peri, "The Role of Income and Immigration Policies in Attracting International Migrants," IZA Discussion Papers 6655, Institute for the Study of Labor (IZA) June 2012.

Ottaviano, Gianmarco and Jacques-Francois Thisse, "Agglomeration and Economic Geography," in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Elsevier, 2004, chapter 58, pp. 2563–2608.

Ottaviano, Gianmarco I. P. and Giovanni Peri, "Rethinking the Effect of Immigration on Wages," *Journal of the European Economic Association*, 2012, *10* (1), 152–197.

Ottaviano, Gianmarco, Takatoshi Tabuchi, and Jacques-Franáois Thisse, "Agglomeration and Trade Revisited," *International Economic Review*, May 2002, *43* (2), 409–436.

Ozden, Caglar, Christopher R. Parsons, Maurice Schiff, and Terrie L. Walmsley, "Where on earth is everybody ? the evolution of global bilateral migration 1960-2000," *World Bank Economic Review, forthcoming*, 2012.

Panagariya, Arvind, "Evaluating the factor-content approach to measuring the effect of trade on wage inequality," *Journal of International Economics*, 2000, *50*, 91–116.

Parsons, Christopher R., "Do migrants really foster trade ? the trade-migration nexus, a panel approach 1960-2000," Policy Research Working Paper Series 6034, The World Bank April 2012.

Passel, Jeffrey S. and D'Vera Cohn, "A Portrait of Unauthorized Immigrants in the United States," *Pew Research Center Publications*, 2009.

Peri, Giovanni and Francisco Requena-Silvente, "The trade creation effect of immigrants: evidence from the remarkable case of Spain," *Canadian Journal of Economics*, November 2010, *43* (4), 1433–1459.

Rappaport, Jordan, "How Does Labor Mobility Affect Income Convergence?," *Journal of Economic Dynamics and Control*, March 2005, *29* (3), 567–581.

Rauch, James E., "Networks versus markets in international trade," *Journal of International Economics*, June 1999, *48* (1), 7–35.

__ **and Vitor Trindade**, "Ethnic Chinese Networks In International Trade," *The Review of Economics and Statistics*, February 2002, *84* (1), 116–130.

Reichlin, Pietro and Aldo Rustichini, "Diverging Patterns with Endogenous Labor Migration," *Journal of Economic Dynamics and Control*, May 1998, *22* (5), 703–728.

Richardson, J. David, "Income Inequality and Trade: How to Think, What to Conclude," *The Journal of Economic Perspectives*, 1995, *9* (3), 33–55.

Romer, Paul M, "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, October 1986, *94* (5), 1002–37.

__ , "Endogenous Technological Change," *Journal of Political Economy*, October 1990, *98* (5), S71–102.

Saiz, Albert, "Room in the Kitchen for the Melting Pot: Immigration and Rental Prices," *The Review of Economics and Statistics*, August 2003, *85* (3), 502–521.

__ , "Immigration and housing rents in American cities," Technical Report 2 March 2007.

Sala-i-Martin, Xavier, "The World Distribution of Income: Falling Poverty and Ě Convergence, Period," *The Quarterly Journal of Economics*, 2006, *121* (2), 351–397.

Samuelson, Paul A., "The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent," *The Economic Journal*, 1952, *62* (246), 278–304.

Schäfer, Andreas and Thomas Steger, "Journey into the Unknown? Economic Consequences of Factor Market Integration under Increasing Returns to Scale," CESifo Working Paper Series 3676, CESifo Group Munich 2011.

Sinn, Hans-Werner, "Germany's Economic Unification: An Assessment after Ten Years," *Review of International Economics*, February 2002, *10* (1), 113–28.

Solimano, Andres and Nathalie Watts, "International Migration, Capital Flows and the Global Economy: A Long Run View," Technical Report, United Nations Economic Development Division, Santiago, Chile 2005.

Stark, Oded, *The migration of labor*, Cambridge, Mass.: Blackwell, 1993.

_ **and J. Edward Taylor**, "Migration Incentives, Migration Types: The Role of Relative Deprivation," *The Economic Journal*, 1991, *101* (408), 1163–1178.

_ , **Christian Helmenstein, and Alexia Prskawetz**, "A brain gain with a brain drain," *Economics Letters*, 1997, *55* (2), 227–234.

_ , _ , **and** _ , "A Brain Gain with a Brain Drain," *Economics Letters*, August 1997, *55* (2), 227–234.

Svensson, Lars E.O., "Factor trade and goods trade," *Journal of International Economics*, 1984, *16* (3Ű4), 365–378.

Tadesse, Bedassa and Roger White, "Cultural Distance and the US Immigrant-Trade Link," *The World Economy*, 2008, *31* (8), 1078–1096.

Tai, Silvio, "Market structure and the link between migration and trade," *Review of World Economics (Weltwirtschaftliches Archiv)*, July 2009, *145* (2), 225–249.

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

Tinbergen, J., *Shaping the World Economy; Suggestions for an International Economic Policy*, Twentieth Century Fund, New York, January 1962.

Tong, Sarah Y., "Ethnic Networks in FDI and the Impact of Institutional Development," *Review of Development Economics*, November 2005, *9* (4), 563–580.

Topel, Robert H., "Factor Proportions and Relative Wages: The Supply-Side Determinants of Wage Inequality," *Journal of Economic Perspectives*, June 1997, *11* (2), 55–74.

Vézina, Pierre-Louis, "How migrant networks facilitate trade: Evidence from Swiss exports," *Swiss Journal of Economics and Statistics, forthcoming*, 2012.

Vousden, Neil, *The economics of trade protection*, Cambridge: Cambridge University Press, 1990.

Wagner, Don, Keith Head, and John Ries, "Immigration and the Trade of Provinces," *Scottish Journal of Political Economy*, December 2002, *49* (5), 507–25.

White, Roger and Bedassa Tadesse, "Immigration Policy, Cultural Pluralism And Trade: Evidence From The White Australia Policy," *Pacific Economic Review*, October 2007, *12* (4), 489–509.

Williamson, Jeffrey G., "Globalization and Inequality: Past and Present," *The World Bank Research Observer*, 1997, *12* (2).

Wong, Kar-yiu, "Are international trade and factor mobility substitutes?," *Journal of International Economics*, 1986, *21* (1-2), 25–43.

—, "International Factor Mobility and the Volume of Trade: An Empirical Study," in R.C. Feenstra, ed., *Empirical Methods for International Trade*, MIT Press, 1988, pp. 231–250.

—, *International trade in goods and factor mobility*, Cambridge, Mass.: MIT Press, 1995.

Woodland, A. D., "Direct and Indirect Trade Utility Functions," *The Review of Economic Studies*, 1980, *47* (5), pp. 907–926.

Wooldridge, Jeffrey M., *Econometric analysis of cross section and panel data*, MIT Press, 2002.

World Bank, *Global Economic Prospects: Economic Implications of Remittances and Migration*, Washington, D.C.: The World Bank, 2006.

Young, Alwyn, "Growth without Scale Effects," *Journal of political economy*, February 1998, *106* (1), 41–63.