CHAPTER 1

Migration Theory*

Örn B. Bodvarsson, Nicole B. Simpson†, Chad Sparber†**
**California State University, Department of Economics, TAH 328, 6000 J Street, Sacramento, CA 95819, USA
†Colgate University, Department of Economics, 13 Oak Dr., Hamilton, NY 13346, USA

Contents

* The authors would like to thank Pia Orrenius, Barry Chiswick, and participants at the 2012 Southern Economic Association Meetings in New Orleans for detailed comments. All errors are our own.

Handbook of the Economics of International Migration, Volume 1A
ISSN 2212-0092, http://dx.doi.org/10.1016/B978-0-444-53764-5.00001-3

1. OVERVIEW

This chapter provides a comprehensive expository survey and synthesis of the theoretical literature on the determinants of migration. There are four themes to the chapter: (1) Most importantly, migration is an act of human capital investment, thus the core of migration theory is based upon the human capital investment model. People migrate if the returns to doing so outweigh the costs. (2) There is no theoretical distinction between internal (domestic) and external (international) migration, though the types and magnitudes of entry barriers vary across the two forms of migration. (3) While most of the theoretical literature focuses on migration as a static decision, more recent models have incorporated dynamic features of migration. (4) Much remains to be done to further refine the theory of international migration.

The chapter is divided into three sections. Section 2 provides a narrative highlighting the migration theory literature starting with Adam Smith (1776) and ending roughly in 2000. The section is especially important for readers wanting a historical perspective into economists' understanding of migration. It establishes the broad themes that continue to influence immigration models today, including the recognition of human capital investment as the key motive for migration. Section 3 discusses much more recent literature. Though it covers a relatively short period in the development of economic thought, this literature is quite large. The papers surveyed in this section delve more deeply into the themes identified by their predecessors by introducing more formal analytical rigor and new insights to the existing immigration framework. Two important innovations are the inclusion of endogenous and dynamic elements into the decision-making process.[1] In the concluding section, we suggest further refinements of the theory.

2. FROM ADAM SMITH TO THE NEW MILLENNIUM

In this section, we survey literature beginning with Adam Smith and continuing through roughly the year 2000. The narrative is divided into three subsections. The first details pre-1960 literature on the determinants of migration.While this literature is small, it laid important groundwork for later analysis. It was not until the advent of the human capital

[1] In both Sections 2 and 3 of this chapter, most of the literature that we discuss highlights US immigration, but it is not our intention to focus on the US case only. In fact, any of the models discussed can be easily applied to any source–destination pair.

model in the early 1960s and the pioneering work on migration by Sjaastad (1962) that the theoretical literature on the determinants of migration really began to form. Our second subsection surveys this literature, highlighting what became the core of migration theory—that migration is an act of human capital investment and that migrants respond to spatial differences in labor market opportunities net of costs. Finally, the third subsection discusses alternative motives for migration that economic researchers identified fairly early in the literature's development.

2.1 Pre-1960 literature

Adam Smith (1776) was the first economist to write on migration. In his *An Inquiry into the Nature and Causes of the Wealth of Nations*, Smith wrote:

> *. . . the wages of labour vary more from place to place than the price of provisions. The prices of bread and butcher's meat are generally the same or very nearly the same through the greater part of the United Kingdom. These and most other things which are sold by retail, the way in which the labouring poor buy all things, are generally fully as cheap or cheaper in great towns than in the remoter parts of the country . . . But the wages of labour in a great town and its neighbourhood are frequently a fourth or a fifth part, twenty or five-and-twenty per cent higher than at a few miles distance. Eighteen pence a day may be reckoned the common price of labour in London and its neighbourhood. At a few miles distance it falls to eight pence, the usual price of common labour through the greater part of the low country of Scotland, where it varies a good deal less than in England. Such a difference of prices, which it seems is not always sufficient to transport a man from one parish to another, would necessarily occasion so great a transportation of the most bulky commodities, not only from one parish to another, but from one end of the kingdom, almost from one end of the world to the other, as would soon reduce them more nearly to a level. After all that has been said of the levity and inconstancy of human nature, it appears evidently from experience that a man is of all sorts of luggage the most difficult to be transported.*

This quote foreshadows research in the migration field that was to take place two centuries later. Smith observed that the rural/urban wage differential substantially exceeded the differential for commodity prices. Regional differences in commodity or factor prices provide opportunities for arbitrage. Intercity movements of goods (trade) arbitrages away commodity price differentials, whereas intercity movements of people (migration) arbitrages away wage differentials. Smith thus suggested that migration is a response to spatial disequilibrium in labor markets and that a key determinant of migration is spatial differences in the returns to labor supply. Hicks (1932, p. 76) made the same point one and a half centuries later when he stated that ". . . differences in net economic advantages, chiefly *differences in wages*, are the main causes of migration" (our italics).

Smith's example would seem to imply that if regional differences in wages exceed regional differences in prices, migration would be plentiful and trade would be modest. Yet Smith found trade to be more intense than migration. The reason is that migration is hampered by certain barriers that trade is not. This is apparent when he states "man is of all sorts of luggage the most difficult to be transported." Smith did not identify these

migration barriers but they would surely include both direct and indirect costs such as relocation expenses, the abandonment of firm-specific assets, the sacrifice of pension rights, and the psychological costs of leaving family and friends behind. Smith astutely observed a regularity still present today—that migration flows are often small despite very substantial international wage differences. The reason is simple: international migration costs are simply too high relative to the gains. By the same reasoning, the large and persistent spatial dispersion of wages Smith saw in the UK likely resulted from low levels of migration.[2]

Ravenstein (1889) provided perhaps the first important analysis after Smith (1776) that significantly contributed to understanding the determinants of migration and helped lay the groundwork for much research that was to follow. He used British census data on nationality and residence along with vital statistics and immigration records to establish seven "laws" of migration. Greenwood (1997) summarizes the laws as the following: (1) most migrants move only a short distance and usually to large cities; (2) cities that grow rapidly tend to be populated by migrants from proximate rural areas and gaps arising in the rural population generate migration from more distant areas; (3) out-migration is inversely related to in-migration; (4) a major migration wave will generate a compensating counter-wave; (5) those migrating a long distance tend to move to large cities; (6) rural persons are more likely to migrate than urban persons; and (7) women are more likely to migrate than men.

Zipf (1946) hypothesized that the volume of migration between two places will be in direct proportion to the product of the populations of the two locations, and inversely proportional to distance. This "$P(1)P(2)/D$" hypothesis—where $P(1)$ is origin population, $P(2)$ is destination population, and D is distance between origin and destination—came to be known as the *gravity model* of migration.[3]

[2] Shields and Shields (1989) suggested that Smith's observation can be captured by the expression for wage convergence below. Labor moves from place i to place j if the wage is higher in j than i, with the amount of migration related to the wage difference as follows: $M_{ij}=\beta_{ij}(W_j-W_i)$, where W is the wage, M is the number of migrants, and β reflects impediments to migration (e.g., distance, imperfect information, and any artificial restrictions that have the effect of blocking the adjustment of wages to migration). The greater is β the lower are the impediments to migration.

[3] The gravity model of migration is an application of Newton's law of gravity, which is used to predict the level of interaction between two bodies. Newton's law states that "Any two bodies attract one another with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between them." In applying Newton's law, Ravenstein's (1889) insight was to treat "mass" as the population of a place and "distance" as referring to miles between two places. The intuition behind the gravity model of migration is that since larger places attract people, ideas, and goods more than smaller places, and more proximate places have a greater attraction, there should be more migration between two places that are more populated and/or are more proximate. The gravity model has also been widely used to describe international trade flows. See Tinbergen (1962), Linneman (1966), Anderson (1979), and Deardorff (1998) for important developments.

The gravity model has very important economic implications. First, by proposing that the volume of migration will be inversely related to the distance between origin and destination locations, the model allows for distance to serve as a proxy for the costs of migration. This is intuitively appealing since longer distances traveled usually imply greater explicit and implicit migration costs, and hence act as a deterrent to migration. Second, it predicts that the volume of migration will be higher for origin and destination communities with large populations. This too has intuitive appeal. At any point in time, some fraction of people in the origin will face opportunities in alternative locations that exceed their opportunities at home, net of migration costs, and those persons will migrate. Assuming that this fraction is constant, then the number of people choosing to emigrate will increase with home population size. Similarly, as the population of the destination community rises, the number of employment opportunities will rise, which will induce more immigration.[4]

Upon reflection on the writings of Smith (1776), Hicks (1932), Ravenstein (1889), and Zipf (1946), one can see the emergence of the core of migration theory that would develop more formally beginning in the 1960s: migration is primarily driven by a desire to maximize one's return to human capital investment. People respond to spatial differences in labor market opportunities by migrating if those opportunities dominate the costs of relocation. Opportunities to migrate are proportional to population size, while migration costs are related to distance. Though these insights laid the foundation for the core of formal theoretical models to follow, it is important to note that empirical studies in this period began to identify more specific determinants of migration that theorists would later incorporate into their models as well. Two papers, discussed below, are particularly noteworthy.

First, Jerome's (1926) empirical analysis linked American business cycle fluctuations in employment with fluctuations in net migration from Europe.[5] Though his paper focused on the business cycle and did not include any theoretical model of migration, his findings provided clear motivation for important work in the migration field that came years later. In particular, he found that: (i) Immigration appeared to respond to changes in domestic employment conditions. That is, his results support the prediction of the traditional economic model of migration—which came four decades later—

[4] There is an additional implication: since Zipf's (1946) original formulation includes the product of the origin and destination populations, this implies that the marginal effect of a change in the origin population on migration will depend on the size of the destination population and vice versa.

[5] We should also mention an important study by Douglas (1919) on the skill distribution of immigrants between 1871 and 1909. Foreshadowing work decades later on the determinants of the composition of migration flows, Douglas used basic statistical analysis to dispute a widely held assertion at the time that more recent migration flows (coming primarily from Southeastern Europe) had on average lower skill levels than less recent flows (coming primarily from Northwestern Europe). Douglas showed that the newer flows comprised proportionately 50% more skilled worker than the older flows.

that the most important determinant of relative migration flows is the spatial differences in real earnings.[6] (ii) Immigration was more sensitive to labor market conditions in the destination region than in the source region. Thus, he was the first to argue for asymmetric effects between "demand pull" and "supply push" factors, finding that destination region conditions are more important in pulling immigrants into a region than origin region conditions are in pushing emigrants out. (iii) There is a two-way causality between immigration and domestic employment. This finding foreshadowed the huge literature on the labor market impact of immigration that began to develop real momentum in the 1990s.

Second, Kuznets and Rubin (1954) analyzed the role of policy and macroeconomic conditions in affecting the costs and benefits of immigration. In particular, they assessed how war and restrictive immigration policies dampened population growth in the USA between 1870 and 1940. Among many important results, Kuznets and Rubin established that the revolutionary change in immigration policy following World War I, coupled with the Great Depression, were likely responsible for a 29.3% reduction in the number of foreign-born persons in the USA during the 1930s, a decade when total population grew by nearly 9 million. In contrast, immigration contributed about one-seventh of the total growth in population, and about 20% of the growth in the labor force between 1870 and 1910. Confirming some of the findings of Jerome (1926), Kuznets and Rubin found that long-term swings in immigration tended to follow long-term swings in GDP per worker. Furthermore, they presented evidence indicating that cyclical changes in in- and out-migration helped to moderate the business cycle, or at least the cycle's effect on native unemployment.

The studies discussed above were collectively important in recognizing early that migration decisions are driven by potential costs and benefits. It was not until the development of the human capital investment model in the 1960s, however, that theoretical models of the determinants of migration were formalized.

2.2 Forming the core of migration theory: migration as human capital investment

The notion of migration as human capital investment is a unifying theme that serves as the most fundamental idea underlying most current economic theories of migration. Unfortunately, it is easy to lose sight of this theme since language in the literature has tended to treat internal (within-country) and external (international) migration as distinct phenomena. Various oft-cited expository surveys (e.g., Massey et al., 1993; Greenwood, 1997;

[6] We also want to highlight Hansen's (1940) comprehensive analysis of flows of American- and Canadian-born persons in each other's countries. His study traced flows as early as the seventeenth century and found that Americans and Canadians frequently "intermingled," moving to places where there were the highest returns to land, labor or capital, and regardless of whether these places were south or north of the border between the two countries.

Lucas, 1997) may have reinforced this tendency, while much of the literature through the mid-1980s—which focused on understanding internal migration flows—may have inadvertently created the misleading perception that internal and international migration are theoretically distinct.

We contend that there is a single theory of migration derived from an application of the traditional human capital model that does not depend on whether migration is internal or international. Migration across "regions" or "countries" can be viewed more or less synonymously, though it is clear that institutional aspects may differ across internal and external migration decisions due to regulations and legal restrictions, language, culture, and other institutional factors. That potential international migrants face different entry barriers than people considering domestic moves is important for understanding an individual's decision set, but it does not change the underlying motive to move—the possibility of realizing better opportunities far from home. In fact, the severity of such barriers has varied across time and space. International barriers are relatively recent in world history, whereas internal migration restrictions have long existed, including prohibitions on serfs from freely moving to cities in medieval Europe,the substantial legal barriers to internal migration in Czarist Russia and the USSR, and China's internal passport system ("Hukou") today. We imagine the modern United States as a country without restrictions on interstate mobility, yet state-specific occupational licensing laws act as barriers to internal migration much in the same way as immigrant quotas are barriers to external migration. One does not need different models to account for these phenomena. Rather, one needs a single model that accounts for different costs.

The human capital model implies that the migrant's goal is to maximize utility by choosing the location that offers the highest net return to human capital, hence labor supply.[7] Accordingly, we will call this view the "labor supply" view of migration. Sjaastad (1962), who pioneered the application of human capital theory to understanding migration, argued that migration is the act of locating one's skills in that market that offers the highest return. While migration can be skill-augmenting,[8] the important point

[7] Most economists presume that the primary reason for migrating is to maximize one's income, specifically income earned in the labor market. Hence, it is taken for granted that utility maximization is achieved through the maximization of income. However, people can migrate for reasons other than income maximization, e.g., family reunification, seeking refuge or political asylum, etc. Those reasons are compatible with utility maximization, but not necessarily with income maximization. We will assume for now that when the migrant seeks out the location that provides the highest utility, he is simultaneously seeking the market that provides him with the highest possible income.

[8] Migration can *indirectly* lead to augmentation of one's skills, however. For example, migration could result in a person locating to an employer that makes more efficient use of her skills and thus allows her human capital to grow faster through on-the-job training than would be the case in the origin. As another example, while we usually think of migration as a phenomenon among people already out of school and working, we can think of migration as the act of moving to an educational institution (e.g. a foreign graduate program) that adds the most to one's human capital. Migration as an educational investment has received much less attention in the literature.

is that by relocating, a person avails herself of the market that provides the highest return to her skills. In fact, if we think of migration as the act of moving to that market which maximizes one's income, then migration need not always be a significant change in geographic location. For example, a person can "migrate" by changing occupations (e.g. he may switch from being a schoolteacher to a real estate agent, but stay in the same general area).[9] By rationalizing migration to be a form of human capital investment, Sjaastad's model added a dynamic element, time, to the migration problem.

While Sjaastad did not provide a formal mathematical model, he asserted that the prospective migrant calculates the value of the opportunity available in the market at each alternative destination relative to the value of the opportunity available in the market at the point of origin, subtracts away the costs of moving (assumed to be proportional to migration distance), and chooses the destination that maximizes the present value of lifetime earnings. Nearly all modern neoclassical economic analyses of the migration decision proceed from this basic framework.[10] Within this framework, migration is treated generally as a once-and-for-all decision involving a change in the location of one's employment, and is no more than an inter-temporal version of the labor market model applied to the case of migration. The model can easily be used to show that prospective migrants respond to differences in real earnings differences across labor markets in different geographic locations.

Non-monetary gains enjoyed from moving (e.g., amenities such as better climate and recreational opportunities, a desirable social, political or religious environment, or more desirable quantities of public goods available at the destination) are not counted among migration returns in the classical Sjaastad model. This does not mean that the influences of amenities and consumption goods are irrelevant. Sjaastad reasoned that spatial differences in such influences on migration will already be accounted for by geographic differences in living costs (his model includes spatial differences in real pecuniary returns to migration). For example, a more pleasant climate in California versus Minnesota should already be reflected in higher prices for California real estate.[11]

[9] Polachek and Horvath (1977) pointed out that "Migration is defined as the flow of people from one given point or set of points in space to another point or set over some finite time interval. Often a minimum distance criterion, that need not be uniform for all inhabitants, is implicitly assumed ... Yet under less restrictive definitions that, for example, may include occupational changes, such a person would be classified as a migrant" (p. 105). Their study did not deal explicitly with occupational changes, though.

[10] We refer the reader here to expository surveys by Greenwood (1975, 1985, 1997), Molho (1986), Shields and Shields (1989), Bauer and Zimmermann (1995), Ghatak et al. (1996), and Gorter et al. (1998).

[11] Shields and Shields (1989) suggest an interesting implication of this. If differences in real estate prices are the primary source of differences in the costs of living between origin and destination, then the returns to human capital investment forthcoming from migration would be reflected in *nominal* income differences. Consequently, we would argue that if one adjusts nominal income differences for differences in the costs of living, one would be mixing together the investment and consumption returns to migration, a strategy that would likely be at odds with Sjaastad's suggestions for modeling migration as human capital investment.

Costs of migration might include losses from selling one's home, car or appliances prior to the move, or additional expenses incurred to replace certain assets left behind at the destination. Also, a move will sometimes necessitate a loss of job seniority, employer contributions to pension plans, and other types of employment benefits. Importantly, however, Sjaastad's framework includes features of a gravity model by viewing distance as a proxy for migration costs. The greater the distance traveled, the greater are the monetary costs of migration such as transportation expenses, food and lodging costs for oneself and one's family during the move, and interruptions in income while between jobs. The cost of acquiring information about job vacancies will rise with distance, whether the information is acquired formally (advertisements in publications and employment agencies) or informally (provided by friends and relatives, for example). Distance can also raise psychological costs of migration since it leads to uncertainty about the new community and its quality of life, displeasure from breaking ties with family and friends, and other stresses of relocation. Furthermore, potential migrants might consider negative spillover effects on friends and family left behind.[12]

Altogether, Sjaastad's model accounts for four aspects of the migration investment decision: (i) the imperfect synchronization of migration's benefits and costs over time; (ii) earnings differences between origin and destination locales; (iii) cost of living differences between the origin and destination; and (iv) the migrant's rate of time preference. Suppose that there are geographic differences in earnings and that a person will retire in T periods. Let W_t^A represent earnings per period in the origin, W_t^B earnings per period available in the destination, CL_t^A an index measuring the cost of living in the origin, CL_t^B an index measuring the cost of living at the destination, i the discount rate, and C the cost of migration. In discrete time, the present value of the net gain to migration π is

$$\pi = \sum_{t=1}^{T} \frac{\left(W_t^B - W_t^A\right)}{(1+i)^t} - \sum_{t=1}^{T} \frac{\left(CL_t^B - CL_t^A\right)}{(1+i)^t} - C(D, X) \tag{1.1}$$

where D is distance between origin and destination locales and X is a vector of any other determinants of migration costs.[13] In continuous time, the present value is

[12] Schwartz (1973) suggested that psychological costs will indeed vary with distance, arguing that migrants often make visits and phone calls to their previous locations to reduce the psychological displeasures of moving, and these costs typically rise with distance. See also Lundborg (1991) for further development of this concept.

[13] Some costs are unlikely to be related to distance. For example, the costs of occupational licensure vary across states, as do vehicle registration costs. In the case of international migration, the costs of obtaining visas will vary according to the country of entry.

$$\pi = \int_{t=0}^{T} \left[W_t^B - W_t^A - CL_t^B + CL_t^A \right] e^{-rt} dt - C(D, X) \tag{1.2}$$

In both the discrete- and continuous-time versions of the model[14] the decision-maker moves only if $\pi > 0$. If multiple destinations are possible, then equations (1.1) and (1.2) are computed for all alternatives and the individual chooses the option yielding the highest value of π. All theoretical applications of the human capital model to migration behavior use some permutation of equations (1.1) or (1.2).

Sjaastad's work has substantially influenced most modern studies of migration. However, it is a very simplified framework and it is important to emphasize its limitations. We emphasize these limitations not necessarily to criticize the framework offered by Sjaastad, but rather to stress that researchers' interests in overcoming these limitations served as the impetus for much of the subsequent literature. These limitations include:

1. *It is a single period model.* The model does not explain why some people migrate on multiple occasions during their lifetimes and it provides no implications relating the likelihood of migration to a person's position in her life-cycle. Some researchers have argued that migration is a life-cycle decision problem, meaning that it is a decision that depends upon the person's age, what point he is at in his career, and how much he values income relative to leisure. For example, it is well known that many older persons tend to have a lower likelihood of migration than younger persons. Overcoming this limitation would require a dynamic model examining the migration decision in the context of a person's life-cycle.
2. *The unit of analysis is the individual.* Some researchers have argued that the preferences and goals of a potential migrant's friends and family members must be taken into account when analyzing the migration decision. For example, if a husband and wife both work, then the husband's decision to migrate is likely to depend upon his wife's career prospects at the destination and vice versa. Overcoming this limitation would require a model in which the decision-making unit is the family, not just one person in isolation.
3. *"Push" or "pull" effects are assumed to be symmetrical.* Migration can be induced by enhanced income opportunities in the destination (the most common "pull" factor) or deteriorating income opportunities in the origin (a common "push" factor). It is well known in the empirical migration literature that pull factors tend to be stronger than push factors, but the Sjaastad model does not offer a clear rationale for such

[14] Sjaastad did not specify an equation for the net present value of migration gains. Therefore, equations (1.1) and (1.2) should be viewed as general formulations of the investment problem verbalized by Sjaastad. Note also that all theoretical and empirical studies involving the human capital investment approach to the study of migration flows utilize some behavioral model that is equivalent to, or is some permutation of, equations (1.1) or (1.2).

differential effects. Overcoming this limitation would require more complex versions of Sjaastad's framework.[15]

4. *Migrants are assumed to be perfectly informed*. It is very likely that a prospective migrant will be uncertain about the size and path of his lifetime earnings stream at the destination. This uncertainty is expected to be particularly acute for those contemplating international migration involving very long distances and changes in language and customs. Migration uncertainty may be less important for someone who must choose between long-term contracts with known compensation both at home and away (e.g., a tenured college professor who is contemplating a move to another university). It can be significant, though, for persons in piece pay occupations (salespeople, agriculture, self-employed persons, etc.), those in occupations where there is significant risk of disruptions in employment, or for those who are contemplating an occupational change. The greater is the degree of uncertainty, the greater will be both informational and psychological costs. If there is uncertainty, the individual's attitudes toward risk will influence his choice to migrate. If a person is sufficiently risk averse and the perceived level of risk at the destination is sufficiently high, he may choose not to migrate. The Sjaastad model does not address how migrants can cope with uncertainty, nor does it ask whether the degree of uncertainty is dependent on the amount of past migration. For example, one would expect that if there are greater ties between the origin and destination due, say, to greater past migration, then there will be lower information costs. Uncertainty could be endogenous to the amount of past migration and is likely to be connected with the amount of psychological costs. Sjaastad did not consider past migration and assumed constant psychological costs. However, psychological costs could depend upon the size of the migrant community and are likely to increase with distance migrated. Overcoming this limitation would require accounting for uncertainty in the decision-making process.
5. *Remittances are ignored*. Many international migrants remit some of their destination country earnings back home. If a prospective migrant plans to do this, then the benefits to migration will include the benefits of remitting. Assume that the benefits

[15] A more sophisticated analysis was later provided by Chiswick (1999), who portrays the migrant as calculating the net rate of return to migration. In Chiswick's model, there are two fundamental components—the (gross) rate of return from migration and the interest cost of funds to finance it. Suppose the spatial income difference rises. Because of the wealth effect, the effect on the prospective migrant's behavior depends upon whether the greater income difference resulted from a higher destination wage or a lower origin wage. A higher destination wage means the migrant's wealth would be higher than before if he moved, whereas a lower origin wage means that the dollar wealth gain from migrating would not be any different. Thus, a higher destination wage tends to enhance migration's lure more than a lower origin wage. Furthermore, Chiswick's analysis suggests that for the same rate of return, the interest cost of funds depends upon whether wages increase in the destination or decrease in the origin. If wages increase in the destination, interest costs tend to be relatively lower compared to the case where wages in the origin fall.

of migration rise with an increase in the home country value of remittances received. Then, the benefits of migrating will be positively related to the real exchange rate between the destination and source countries. For example, if the price of the destination country's currency in the source country rises, adjusted for inflation, remittances received by family and friends back home will be more valuable. The appreciation of the destination country's currency will thus boost the benefits of migration. Furthermore, the returns to migration will be augmented the lower are the costs of remitting earnings back home. For example, the recent growth in the international money transfer industry is due mostly to the growth of both legal and illegal migration. Technological advances in that industry, combined with greater competition, should result in lower costs for transferring funds, which should boost the amount of migration. Overcoming this limitation would require incorporating remittances, exchange rates, and financial sector innovations into the model.

Though we credit Sjaastad for his contributions conceptualizing migration as human capital investment, he was not the only economist in the 1960s to do so. Gary Becker's first writings on human capital theory, which came only a few years after Sjaastad published his pioneering work, provide another important example. In his book, which builds upon the work of Schultz (1961) and was the impetus for what is now the very extensive literature on human capital theory, Becker (1964, p. 7) states "The many forms of such [human capital] investments include schooling, on-the-job training, medical care, *migration*, and searching for information about prices and income" (our italics). Echoing Sjaastad, Becker argues that the decision to relocate is, first and foremost, an investment decision because it involves the incurring of direct and indirect costs up-front in order to realize an (uncertain) payoff in the future.[16]

The pioneering work of Sjaastad and Becker is enormously helpful in understanding the international migration decision. An international migrant incurs transportation, visa, opportunity, and non-monetary costs now in hopes of a payoff in the destination in the future. The expected payoff could be a superior return to one's human capital in the destination region (the case of labor market migration), but there could be other payoffs such

[16] Becker makes repeated references to the causes, as well as consequences, of migration throughout his book. In a discussion of specific training, he mentions the likelihood of imperfect transferability of human capital, especially in high-skill areas, across borders (pp. 27–28). He observes that there is often relatively little worker mobility across borders despite huge real international earnings differences (p. 28). He observes that immigrants, when doing international job searches, often have to pay their own way to the destination country (pp. 32–33). He observes that younger persons are more likely to migrate than older persons, all other things being equal (p. 50). He predicts that temporary migrants to urban areas will have less incentive to invest in urban skills than permanent residents (p. 51), and he observes that abler persons have a higher proclivity to migrate (p. 63). All of these observations and predictions are, in one way or another, implications of a model based on the view that migration is a form of human capital investment.

as greater political freedom, reduced risk of loss and limb in the destination country (the case of refugee migration), or the emotional/psychological benefits of being with relatives in the destination (the case of family migration). Or the expected payoff could be the acquisition of skills in the destination region through study at educational institutions, training opportunities, or unique employment opportunities only available in the destination.

There are clearly a wide variety of motives for international migration. While the primary interest in the literature has been economic migration, one cannot discount the importance of other motives. In recent decades, US immigration has been substantially driven by the desire for family reunification. Many countries have absorbed large numbers of refugee migrants. Migration is frequently driven as much by consumption motives (e.g., retirees' search for warmer climates) as by work interests. It is not unusual to see families move because they seek a higher quality of life elsewhere, not necessarily higher income. Nevertheless, *all* forms of voluntary migration, including those that may seem to be motivated by non-economic factors, are in the main determined or influenced by conventional economic (human capital) forces.

2.3 Early extensions of the Sjaastad model

It is convenient to think of the Sjaastad model as taking a "labor supply view" of migration in which an individual's principal motivation for migration is to improve the rate of return to his/her human capital, net of migration costs. Most alternative models of migration are simply extensions of this model, adding, for example, additional motives for migration and/or explicit sources of migration costs. In this subsection, we survey the early literature extending Sjaastad's model.

2.3.1 The migrant as a consumer

Sjaastad's human capital investment model of migration does not account for amenities in the utility function, but rather counts them as geographic differences in the costs of living. By focusing on income differences as the lone motivation for migration, however, the human capital model implicitly assumes that all goods are tradable, thereby effectively treating utility and income synonymously. In contrast, many regional economists emphasize that migration is primarily a *consumption* decision.[17] That is, the "consumer model" of migration offers a richer specification of the human capital hypothesis that separates tradable and non-tradable goods (i.e., amenities) in the utility function directly. By placing greater weight on amenities, the model can generate stable

[17] This literature was partly motivated by disappointing results from tests of the Sjaastad model. By the 1980s, these tests frequently failed to confirm spatial differences in earnings as a determinant of migration. For a sampling of important early papers emphasizing the consumer model of migration, see Rosen (1974), Graves (1979, 1983), Greenwood (1997), Glaeser and Shapiro (2003), and Green et al. (2006).

equilibria with equal utility across regions and unequal levels of income even in the absence of migration costs.[18]

Typical models based on the consumption view proceed from the assumption that households and firms are always in equilibrium at their different locations. Regional income disparities do not necessarily mean that there will be utility gains from migrating. If gains are available, they will be arbitraged away quickly because mobility is high, information costs are low, and markets are generally quite efficient. Consequently, persistent regional differences in wages, rents, and prices represent a compensating differential for regional differences in amenities. Some areas have features that are more attractive than other areas. Thus, there will be a spatial equilibrium that reflects differences in amenities across localities.

According to the consumer model of migration, migration flows are triggered by changes in life-cycle factors such as the onset of retirement or changes in real incomes that will alter the demand for amenities. For example, long-term technological advances will raise peoples' real incomes and, assuming that amenities are normal goods, boost the demand for those amenities. Because amenities tend to be distributed unevenly, migration will occur. Consequently, amenity-rich areas will experience in-migration, driving down wages and driving up land prices, whereas in amenity-poor areas, wages will rise and rents will fall. Technological advances could have the same effects on producer demand for amenities. An important point made by practitioners of the consumption view is that migration cannot spatially equilibrate wages. People will only migrate if the value they attach to amenities exceeds the lower purchasing power that results from decreased wages and increased rents.

The notion that people migrate in response to spatial differences in amenities also extends to public goods. Long before regional economists were constructing models relating spatial equilibrium to amenities, Tiebout (1956) argued that an important factor explaining why people move from one locality to another is differences in the quality of public goods such as police and fire protection, education, hospitals, courts, beaches, parks, roads, and parking facilities. According to the Tiebout hypothesis, consumers "vote with their feet," picking communities that best satisfy their preferences for public goods. Hence, location decisions of households will depend on the local fiscal policies of various localities. Furthermore, the greater the number of localities and the more diffuse is the distribution of the quality of public goods among them, the closer the consumer/voter will come to satisfying his preference pattern.

[18] A related, but alternative, methodology comes from Shields and Shields (1989). Motivated by the *new household economics* literature of Becker (1965), Lancaster (1966), and Willis (1973), they model migration as a household decision contingent upon household production costs that are influenced by locational amenities.

Country differences in amenities may help explain international migration flows to some extent. For example, a country's political system, the risk of persecution, the likelihood of cultural acceptance, how permissive the environment is to creative expression, the crime rate, and weather are all potentially strong determinants of migration. Furthermore, international migrants could be searching for higher levels of public goods such as quality health care, educational systems, and judicial systems. However, the consumption view cannot deny that relative migration flows and destination choice are still going to be influenced by conventional economic (human capital) forces. Moreover, the costs of international migration may be prohibitively high. Thus, while the consumer model of migration adds insight into the migration decision, it is not unreasonable to prefer the more simplified human capital models to explain interregional migration flows.

2.3.2 The influence of kinship and migrant networks

In Sjaastad's model, pecuniary migration costs depend only on distance traveled, nonpecuniary psychological costs are constant, and there are zero information costs. Later work has argued that psychological and information costs are likely to fall when there is greater access to family, friends, and other previous migrants in the destination. In the sociology-based migration literature, the community of family and friends at the destination is often referred to as a *kinship network* and the community of earlier migrants is referred to as a *migrant network*. Access to these networks can greatly improve the efficiency of migration. For example, Yap (1977) has suggested that migration to an area will be enhanced if migrants have relatives and friends already there, languages in the origin and destination are similar, or if a relatively large stock of migrants from the origin previously migrated to the destination earlier in time. A similar point has been made by Hugo (1981), Taylor (1986), and Massey and Garcia Espana (1987). If a person is moving to an area where there is a network of migrants already there, that could result in lower employment search costs, lower costs of securing housing and child care, and more protection from crime. Having family and friends already at the destination could reduce the stress associated with an interregional move and significantly reduce language barriers in the destination.

If kinship and migrant networks are effective in reducing information and psychological costs, then migration costs will be endogenous to the volume of past migration. Specifically, moving costs should decrease with the number of migrants already settled in the destination. Carrington et al. (1996) tested a discrete-time model of internal migration based on the assumption of endogenous mobility costs. They hypothesize that with endogenous mobility costs, migration builds over time and the volume of migration may increase even as destination–origin wage differentials narrow. Furthermore, their model demonstrates that migration tends to follow specific geographic channels, with the first cohort of migrants being those with the lowest migration costs.

The incorporation of kinship and migrant networks into the migrant's objective function is straightforward. For example, rather than just assuming some general migration cost function with no underlying determinants, the model could relate migration costs to a risk variable such as the probability of earnings losses during some initial period at the destination, but this risk variable will be lower the greater is the size of kinship and/or migrant networks. This was done by Taylor (1986),[19] who developed a model of household[20] labor allocation that incorporates the influence of kinship networks at the destination in reducing the risks of migrating. He argues that kinship networks influence household labor decisions by serving as "migration insurance," i.e. insurance against income loss at the destination. The benefits component of the migrant's objective function could also include the assumption that kinship and migrant networks augment utility because of the psychological gain that comes from having familiar faces and contacts in a new place.

Though the examples above identify the lessened psychological costs of migration associated with larger migrant networks at the destination, a less-cited literature highlights the increased costs associated with kinship networks. That is, individuals with well-established kinship networks at home will experience deep psychological costs from leaving behind family and friends should they decide to migrate. Various researchers have suggested that these costs are likely to increase with distance traveled (see Beals et al., 1967; and Schwartz, 1973).

Ultimately, the desire to increase one's income remains the primary motivating force behind the migration decision in both the classical human capital investment model and models emphasizing kinship networks. The only fundamental difference between the two is that the latter provides a richer explanation for non-pecuniary migration costs. Such costs are no longer constant, and are reduced by the presence of extensive kinship and migrant networks in the destination. Networks can therefore stimulate migration. In Section 3.2.1, we provide a longer discussion of more recent work in this area.

2.3.3 Migration decisions in a life-cycle context

The Sjaastad model treats migration as a single-period problem. It does not explicitly address return migration (the decision to return to the home region) or circular migration (the decision to frequently cross a border). Instead, the Sjaastad model implicitly assumes that such actions arise only when the particular costs and benefits of migration change to alter the migration decision. Empirically, many people have a high periodicity of migration, whereas others have a low periodicity. The Sjaastad model is silent on the causes of this heterogeneity other than to assume that repeat migrants are those whose migration

[19] See also Goodman (1981).

[20] The decision-making unit in Taylor's model is the family, but his model is easily applicable to the individual.

costs and benefits change more frequently than others. That is, explanations for repeat and return migration are not predictions formally implied by the Sjaastad model, but instead are simply assumptions.

Polachek and Horvath (1977) argue that if migration is viewed as an investment process undertaken at each stage of the life-cycle, then refutable explanations emerge for the periodicity of migration. They proceed from the assumption that what matters to people are locational characteristics, and that relocation involves choosing the set of characteristics that maximizes utility. Locations can be thought of as composites of various locational characteristics such as unemployment rates, prices, industrial composition, occupational opportunities, or per-capita governmental expenditures on education. As a person moves through his/her life-cycle, his/her demand for locational characteristics changes.

A second essential point in the Polachek and Horvath (1977) model is their recognition of and emphasis on the costs of acquiring information about alternative locations. Such costs entail a joint investment with mobility. Mobility is a response to changing demand for locational attributes, but location choice can only be facilitated by the acquisition of information that reveals where one can find desirable attributes.[21] Mobility is thus the outcome of changes in individual demand for known locational characteristics.

Because there are multiple stages to the life-cycle, it is very likely that there will be multiple episodes of mobility during a person's life. Migrants with high periodicities are those whose preferences, benefits, and costs change more rapidly than others. Furthermore, older persons will migrate less frequently than younger persons. These and other implications are derived formally using an optimal control theory framework that treats the choice of locational attributes, information, goods, and human capital acquisition as simultaneously determined.

Importantly, Polachek and Horvath (1977) also demonstrate that without information search and the acquisition of human capital, their model reduces to the simple Sjaastad model of dichotomous choice. It simply treats migration as a continual process of revising one's demand for locational attributes over the life-cycle, concurrent with human capital acquisition. More recently, dynamic macroeconomic models have been used to better understand circular migration. We discuss those developments in Section 3.3.3.

2.3.4 The effects of uncertainty on migration

The Sjaastad model assumes that a migrant can be certain to secure employment in the destination. If migrants find work at the destination instantly upon arrival and migration

[21] The problem, as Polachek and Horvath (1977) point out, is that mobility causes information to depreciate, resulting in losses. The losses in information value must be weighed against the gains that come from moving to a place with more desirable attributes.

costs are non-existent, a pure disequilibrium model would thus imply complete wage convergence between origin and destination locales. This assumption is clearly problematic. A number of development economists, beginning with Todaro (1969, 1976) and Harris and Todaro (1970), have argued that this assumption is very unrealistic for cases involving rural-to-urban migration in developing countries.[22] They point out that when rural migrants come to the city, there can be a long waiting period before a job is secured in the urban "modern" sector. When the migrant arrives in the urban area, he typically joins a large pool of unemployed and underemployed workers, all spending time in the urban "traditional" sector, where they are either fully or partly unemployed and often performing menial tasks for low pay. Todaro's work may be viewed as a modification of the Sjaastad model, where expected income at the destination is substituted for actual income:

$$\pi = \int_{t=0}^{T} \left[p(t) W_t^B - W_t{}^A - CL_t^B + CL_t^A\right] \mathrm{e}^{-rt} \mathrm{d}t - C(D, X) \tag{1.3}$$

where $p(t)$ is the probability a migrant will be employed in the modern urban sector in period t, B is the destination (i.e., the urban sector), and A is the origin (i.e., the rural sector). Recall that D is distance between the origin and destination, and X is a vector of any other determinants of migration costs. In this context, rural-to-urban migration will continue as long as the expected wage in the urban sector, net of migration costs, equals the wage in the rural sector (where the probability of finding employment is assumed to be 100%). This basic model was subsequently extended by a number of researchers to take into account other important and unique features of developing countries.[23] Unemployment risk has also been captured in modern models of international migration, which we discuss in Section 3.1.7.

Wage uncertainty in the destination can also be modeled as a sequential search process in which the migrant maximizes expected net income and faces a stationary probability

[22] For a very thorough review of the literature on internal migration in LDCs, see Lucas (1997). For a very recent and thorough review specifically of the literature on rural to urban internal migration in LDCs, see Lall et al. (2006).

[23] Harris and Todaro (1970) extended the Todaro (1969) model to the case where there is a wage subsidy policy in the urban sector and restrictions on rural-to-urban migration. They demonstrated that: (a) under certain conditions, either policy can lead to welfare improvement; and (b) maximum welfare improvement will result from implementing the policies concurrently. Bhagwati and Srinivasan (1974) demonstrate, however, that once the migrant's utility function is explicitly considered, either policy can be shown to be welfare-maximizing. Corden and Findlay (1975) extended the Harris and Todaro model to allow for capital mobility, whereas Fields (1979) extended it to allow for job search by migrants, preferential hiring by educational level, and labor turnover. Calvo (1978) extended the Harris and Todaro model to include a trade union in the urban sector, whose objective is to maximize the difference between its members' incomes and what they would earn in the rural sector.

distribution of wages at the destination (Pickles and Rogerson, 1984; McCall and McCall, 1987). In each period, an observation from that distribution is revealed in the form of a wage offer and an individual compares the offer with his reservation wage (e.g., the wage at the origin). When a wage draw exceeds the reservation wage net of migration costs, the individual migrates. This process also affects the length of time before a move is made, if a move is made at all. One important implication of the model is that the more (less) favorable labor market conditions are at the origin, the longer (shorter) it will take for a sufficiently attractive wage offer to arrive at the destination, hence the longer (shorter) it will take on average for an individual to decide to migrate. This may help explain a key stylized fact that regional and international migration are "slow bleed" phenomena in which migration responds only sluggishly to real income differences.[24]

Burda (1993, 1995) offered a somewhat different explanation for how labor market uncertainty in the destination can lead to this slow bleed nature of migration. He argues that it often pays to delay making a relocation decision until more information is received. In Burda's models, the migrant has to choose between two strategies: (i) migrate immediately and take the risk that migration will be more profitable than staying at home; or (ii) wait until he knows for sure whether migration is profitable or not. Procrastination has value—it will be more profitable to delay a migration decision if the benefits of waiting for information exceed the costs. In equilibrium the probability of receiving good news about the destination's labor market is equivalent to the probability of migrating. Thus, the expected net gain of deferring the decision to migrate is the net gain of migrating when the destination's labor market is favorable, weighted by the probability that market conditions will be favorable. This gain can be higher than what would be enjoyed if migration took place immediately; hence postponement of the migration decision to the second period can make this person even better off. This is what Burda (1993, 1995) called the *option value of waiting* to migrate and it represents gains to procrastination.[25] An important caveat to Burda's analysis, however, is that there are also costs of waiting. For example, superior job opportunities in the destination are options with expiration dates as well. Waiting can sometimes result in the better job offers disappearing, and this risk must also be accounted for in the waiting decision.

[24] For example, Burda (1995) found that following a large spike immediately after reunification of West and East Germany, migration from the East to the West was surprisingly sluggish despite very large real wage differences. This type of pattern, where migration is sluggish, despite significant real income differentials between countries, has been found for other prominent cases of migration as well.

[25] Burda developed a formal theoretical model, deriving an expression for the option value of waiting, which is the excess of the value of a waiting strategy over the value of a strategy in which migration is undertaken immediately. He demonstrated that the value of the migration option is inversely related to the current wage gap, positively related to migration costs, has an ambiguous relationship with the discount rate, is inversely related to the wage gap when destination conditions are unfavorable, is positively related to the probability of unfavorable market conditions, and independent of the wage gap when conditions are favorable.

2.3.5 What happens when the family is the decision-making unit?

The earliest models of migration do not distinguish between personal and family decisions. Sjaastad's (1962) focus is on the individual and there is no analysis of how his migration may affect other persons close to him. The implicit assumption in early research on the migration decision is that if the migrant is part of a family, then the welfare of the rest of the family is unaffected by that person's decision to relocate. For a large proportion of internal and international moves, however, migration is indeed a family decision and everyone in the family is affected by it. Extensions of the Sjaastad (1962) model to include family ties date back to the mid-1970s. These extensions were designed to address two questions: (1) When the whole family migrates, how is the decision to migrate made when family members have conflicting interests? (2) Why would part of the family migrate while the other part stays at home? We discuss each of these questions below.

Work modeling family migration decisions began with Sandell (1977), Polachek and Horvath (1977), and Mincer (1978). These models assume that family members might have conflicting interests. If the entire family migrates, then relocation may enhance the well-being of some family members, but may reduce it for others. Even though the household head's income and job satisfaction may improve with relocation, other family members may suffer income losses, as well as psychological costs that result from leaving family and friends behind, adjusting to a new language and culture, etc. All other things being equal, family ties have a greater tendency to discourage migration than to encourage it.

Mincer's (1978) model of the family migration decision can be easily described. Suppose for simplicity that the household includes just two persons, a husband and a wife. Allow ΔI_H to be the change in the present value of the husband's income stream were he to relocate to another region or country, and let ΔI_W be the change in the present value of the wife's income stream were she to move with him. Note that ΔI_H is equivalent to the husband's "private" gains to migration, the gains he would enjoy if he were single and were deciding on his own to migrate. Similarly, ΔI_W measures the wife's private gains to migration; if she were not married, she would move if the private gains were positive.

Let us assume that this two-person family has two alternatives to choose from: (i) both migrate together; or (ii) both stay at the origin.[26] How then does the family decide whether or not to migrate? Mincer argues that the requirement for migration to take place is not that both persons have positive gains to migration (both ΔI_H and $\Delta I_W > 0$), but rather that the net sum of the family's gains is positive:

$$\Delta I_W + \Delta I_H > 0 \tag{1.4}$$

[26] We rule out the possibility that one person migrates, while the other stays behind, as is the case with many "commuting couples," i.e. couples where each spouse has a separate home connected with his/her workplace, but the two spend time together at one or both of those homes periodically.

The insight here is that the family may migrate even if the two persons have conflicting interests. What matters is not what migration does to each person's private gains, but to the joint gains of this multi-person decision-making unit. If the private gains to migration for each person are positively correlated, then each family member agrees whether or not to migrate and the decision is individually efficient. If the private gains to migration for each person are negatively correlated, however, then it will be privately inefficient for one of the family members to migrate, although it may be efficient for the family unit to do so. This may arise, for example, if the husband is in a different occupation than his wife and their labor markets are very different in both the origin and destination. In the event that the husband experiences a gain from migration, the wife a loss, but the joint gains are still positive, then the wife is a *tied mover*—she follows her husband even though her employment outlook is better at home. Suppose, in contrast, that the wife's loss from migration dominates the husband's gain. Then, he is a *tied stayer*, which means that he sacrifices superior employment opportunities available elsewhere because his wife is much better off in their current place of residence.

The Mincer model can quite easily be expanded to analyze other economic phenomena. For example, it has interesting and important implications for such factors as marital formation, marital stability, and the labor supply decision of the "tied" party. Of particular note for immigration economists, it can be used to describe the emergence of commuting couples as an alternative to the migration decision—that is, couples in which one partner chooses to commute great lengths instead of moving the entire family to a new destination. Moreover, expansions of the Mincer model that endogenize the tied partner's labor supply decision will deliver a richer set of migration and labor supply implications that depend upon both partners' opportunities. That is, such models can better describe whether one or both partners move, commute, leave the labor market, settle for an inferior job, find improved labor market opportunities, or remain at home in their current occupation of employment.

Another strand of literature emphasizing the role of the family in the migration decision emerged in the 1980s. This literature, beginning with Stark and Levhari (1982), Stark (1984, 1991), and Katz and Stark (1986), focuses mostly on explaining migration from developing to developed countries and emphasizes the role of migrant remittances. The premise of this literature is that the household chooses whether or not to send a family member away to work. The decision to send a family member overseas, for example, is essentially a "family portfolio diversification decision"—a decision to try to hedge against risk and to ease liquidity constraints. These researchers have suggested that migration: (i) is a response to various market failures experienced by families and households in developing countries; and (ii) provides an alternative source of capital for families and ensures stability in consumption.

The core feature of this collective decision-making model is that the family or household, unlike the individual, can control risk through diversification of household resources. Some members of the family, for example, can be assigned to work in the local economy, while others may be sent to work in foreign labor markets where conditions

are negatively correlated with local labor markets. If there is a downturn in the local labor market and the household faces a liquidity constraint, then having a family member working overseas in a strong labor market eases the constraint because the migrant can remit his income home. According to this literature, the decision to have family members migrate is a response to a lack of risk-hedging mechanisms such as crop insurance markets, futures markets, unemployment insurance, and capital markets. In fact, modern models of immigration exploit the insurance role of immigration as a household-level decision (as discussed in Section 3.2.2).

2.3.6 Migration as a response to relative deprivation

Stark (1984, 1991), Katz and Stark (1986), Stark and Taylor (1989, 1991), and Stark and Yitzhaki (1988) suggest that migrants are motivated to relocate by a desire to improve their income relative to members of their reference group. "Relative deprivation" models[27] argue that utility is influenced from knowing how one's income ranks relative to his peers.[28] In the context of migration, the reference group is assumed to be other income-earning people at the origin. A person's feelings of relative deprivation will be greater, all other things being equal, the lower his income ranks in the distribution. Furthermore, his utility will rise if his ranking in the income distribution rises, even if his absolute income stays the same.[29] If migration leads to higher absolute income elsewhere (adjusted for cost of living differences), then a person's utility rises because (i) consumption opportunities have improved and (ii) feelings of relative deprivation have been assuaged.

To illustrate these concepts more precisely, define $F(Y)$ as the cumulative income distribution in a person's reference group and $b[1-F(Y)]$ as the disutility felt from not having an income that is higher than Y^*. The relative deprivation $RD(Y^*)$ experienced with absolute income Y^* is

$$RD(Y^*) = \int_{Y^*}^{Y'} b[1-F(z)]\mathrm{d}z \tag{1.5}$$

[27] The concept of relative deprivation is due originally to the social psychologist W.C. Runciman (1996).

[28] Stark's (1984) example was that, "In a well-defined reference group—an army regiment or a university faculty—we are promoted from time to time on par with our peers. We are always aware of opportunities elsewhere, but decide not to pursue them. Assume that one fine day, arriving in a cheerful mood at the officers' or faculty club, we find out that everyone else has been promoted or tenured, but that we were not. In the army one would be likely to put in for an immediate transfer; in academic life—actively seek to offer elsewhere; one becomes *relatively deprived* (Stark's italics), resents it, and decides to 'migrate'" (pp. 210–211).

[29] It follows logically that feelings of relative deprivation will fall if the incomes of one's peers fall, but one's own income stays the same.

where Y' is the highest income earned by someone in the reference group. The individual's feelings of relative deprivation will fall (rise) if income rises above (falls below) Y^*. If migration results in a higher income of Y', then relative deprivation falls by an amount equal to the following:

$$RD(Y^*) - RD(Y') = \int_{Y^*}^{Y'} b[1 - F(z)]\mathrm{d}z - \int_{Y''}^{Y'} b[1 - F(z)]\mathrm{d}z \tag{1.6}$$

While the notion that relative deprivation motivates migration is very easy to understand and may be appealing, it is easily subject to criticism. First, the behavior predicted by relative deprivation models is often not consistent with what we tend to observe in migration behavior. According to relative deprivation models, the benefit of migration depends upon a person's position in the income distribution. This means that, for example, a person's relative deprivation improves if he/she moves from being in the 50th percentile of the income distribution in a high-income country to being in the 95th percentile of the income distribution in a low-income country. The model implies that if the gain in utility from the reduction in relative deprivation is high enough, the person would move from a high-income country to a low-income country. However, this is rarely observed. Indeed, most domestic and international migrants receive higher absolute real incomes in the destination, but on lower rungs of the income distribution compared to the origin. Second, relative deprivation is an explanation rooted in the migrant's tastes, which are impossible to measure. Third, many would argue that relative deprivation is a concept that cannot be extended to cases other than rural-to-urban migration in relatively poor countries.

Nonetheless, there are some potentially important testable implications of the relative deprivation hypothesis. First, it implies that characteristics of the income distribution of the migrant's source country will influence the decision to migrate. For example, if absolute income stays the same, but the variance of the distribution or the degree of positive skewness rise, this will alter utility and raise a person's incentive to migrate. Second, the relative deprivation hypothesis is capable of contradicting the traditional hypothesis that expected income differentials across two regions or countries must be positive in order to induce migration. Altogether, however, though feelings of relative deprivation are indeed a motive very different from the others discussed, they do not necessarily replace or contradict the human capital investment motive for migration. Concerns about relative deprivation may be best viewed as an additional motive for migration within a larger human capital investment model.

2.3.7 The influence of age on the migration decision

It is well known that international migrants tend to be young. The theoretical literature on the relationship between age and migration decisions is surprisingly thin, though some early literature provides insights. Becker (1964) argued that the propensity to migrate will tend to decrease with age because the expected net present value of benefits from

relocation will, due to greater duration of stay in the destination, be higher for younger persons. This implies that migration rates for persons from the lower (higher) end of the source region's age distribution will be higher (lower).[30]

Gallaway (1969), in contrast, suggested an ambiguous relationship between age and the probability of migration. On the one hand, older workers face higher costs of relocation because (i) they face a greater expected cost of not being able to fully transfer pension rights accumulated in retirement programs; (ii) the expected costs of liquidating physical investments in the origin are often higher; and (iii) they face greater psychic costs from uprooting themselves from long-held jobs and dwellings. On the other hand, migration will tend to be more affordable for older workers because they often earn more money and have more assets than young workers.

David (1974) suggested that seniority rights (which provide protection from the risk of layoffs) will be lost following a move. Schwartz (1976) emphasized the importance of psychic costs, arguing that as persons get older, they will invest more in relationships with family members and friends and the emotional costs of severing those relationships will be higher. Building upon Schwartz's work, Lundborg (1991) suggested that the demand for return visits will depend on the length of time spent at the destination, age at the time of migration, and the stock of prior migrants from the origin residing in the destination. An increasingly important form of internal and international migration is undertaken by the aged. This has important implications for the destination labor market—such as greater demand by the aged for caregiving services—as well as direct implications for the demographic distribution of the population.[31]

3. RECENT THEORETICAL ANALYSES OF WHY PEOPLE MIGRATE

The previous section provided a narrative survey and assessment of the theoretical literature on the determinants of migration through roughly 2000. In this section, we survey mostly post-2000 literature, but with a greater focus on the mathematical mechanics behind the models. Common between both sections, however, is recognition that human capital investment forms the center of migration models. The most recent models of the determinants of migration are characterized by their emphasis on the endogeneity

[30] Schlottmann and Herzog (1984) described this as "age selectivity of migration." They focused on how career and geographic mobility interact in influencing this age selectivity. Using data on interstate migration for 1965–70, they concluded that failure to account for this interaction will tend to overstate the negative influence of age on the probability of migration. It should be noted that Schlottmann and Herzog focused on this particular empirical issue and did not provide a theoretical model with any novel implications regarding the relationship between age and proclivity to migrate.

[31] Bodvarsson and Hou (2010) re-examined these theoretical issues using a model of the migration decision that captures all the above effects of age on the migration decision, in addition to other influences not discussed in the literature. They demonstrated that age will have an ambiguous effect on the proclivity to migrate; in some cases, the returns to migration can rise with age.

of the migration decision and of wages. Simple models treat migration as a static decision determined by exogenous wages that vary across different levels of human capital. More rigorous models treat migration as a dynamic decision with endogenous wage determination and human capital accumulation.

This section is divided into three subsections. In the first, we discuss somewhat less complex static models with endogenous migration, whereas the second discusses more complex models that add the feature of endogenous wages. The third subsection then explores dynamic models of migration.

3.1 A static human capital model with endogenous migration

As a framework for understanding recent theoretical developments on the determinants of international migration, we begin by describing a simple model where the migration decision is endogenous. The basic two-country framework will deliver an equation that pinpoints the migration rate between the two countries. We then add various components to the model, in line with recent work, to isolate the importance of various factors in affecting migration rates. For simplicity, we abstract from return and circular migration such that the model will produce migration rates from one region to another.[32]

Let us begin by assuming that the world economy consists of two regions (or countries), A and B. There is a continuum of agents in each region. Agents are heterogeneous with respect to skill level: they are either unskilled (u) or skilled (s). The precise definition of this dichotomy is not crucial for understanding the fundamental nature of theoretical models, but most researchers define skilled workers as those with at least some college education. The measure of agents with skill i is defined as Ω_i for $i=\{u,s\}$.

Agents maximize their utility over consumption and leisure. Each is endowed with one unit of time, which can be allocated to leisure activities or supplied to the firm. Agents choose a region in which to reside by comparing the expected value of living in each. For simplicity, assume that the expected value of living in region j depends on the after-tax wage rates, w_j^i, faced by an agent with skill level i. Thus, the level of human capital of an agent determines his/her wage rate at home and abroad. An agent of skill i solves the following maximization problem:

$$\max\left\{\lambda^i E\left[u\left(c_A^i, l_A^i\right)\right] + \left(1-\lambda^i\right)E\left[u\left(c_B^i, l_B^i\right)\right]\right\} \tag{1.7}$$

subject to the budget constraints:

$$c_A^i = w_A^i\left(1-l_A^i\right) \text{ and } c_B^i = w_B^i\left(1-l_B^i\right)$$

The agent chooses how much to consume, c_A^i, the amount of time to devote to leisure activities, l_A^i, and the probability of remaining in region A, λ^i for $\lambda^i \in [0,1]$. The after-tax wage rate, w_j^i, is taken as given.

[32] Later in the chapter, we will discuss how these important aspects of migration have been incorporated in models.

Expected utility depends on the migration decision, and hence the expected value of living in each region. Following Rogerson (1988), agents randomize over migration. The allocation of consumption and leisure will depend upon their region of residence. λ^i is defined as the probability that the allocation $\{c_A^i, l_A^i\}$ is realized (i.e., that the agent lives in region A), whereas $1-\lambda^i$ is the probability that the allocation $\{c_B^i, l_B^i\}$ is realized (that the agent lives in region B). Since there is a continuum of agents in each region, the representative agent of type i in each region determines the equilibrium fraction of type i agents living in region A. Similarly, the probability of living in region B for a specific type of agent is equivalent to the fraction of that type living in region B.

In equilibrium interior solutions (i.e., cases in which migration occurs at a rate less than 100%), an agent of skill i must be indifferent between living in region A and living in region B. Thus, it must be that $E[u(c_A^i, l_A^i)] = E[u(c_B^i, l_B^i)]$ for all i. In corner solutions, $\lambda^i = 0$ implies that the agent strictly prefers to live in region B; if $\lambda^i = 1$, the agent strictly prefers to live in region A.

3.1.1 Adding migration costs

The model represented in equation (1.7) abstracts from migration costs and policy, two factors that play important roles in affecting the flow of migrants between two regions. One puzzle, as discussed earlier, is that observed international migrant flows are small given the huge real income differentials across countries. This implies that significant migration costs must be present so that net benefits of migration are small once moving costs are taken into consideration. Following Chiswick (1999), migration costs may be explicit travel or admissions costs, or they may be implicit costs associated with time spent looking for a job, language and cultural assimilation, distance from family, a preference for home, or psychological costs (see Urrutia, 1998; Chiquiar and Hanson, 2005; Hunt, 2006; Clark et al., 2007; Hatton and Williamson, 2011; and Grogger and Hanson, 2011).[33] Migration costs may depend on the specific pair of origin and destination countries. For example, travel costs may depend on the distance between the two countries. Other types of migration costs, such as home preference, might only depend on the origin country. We will allow for the most general case in the model, which allows costs to depend on both the source and destination countries.

Costs can be explicitly modeled as an exogenous loss of utility associated with migration that depends on the region of origin and the destination.[34] If the costs are measured in utils, then the maximization problem for an agent of skill i migrating to region j becomes:

[33] Refer to Section 2.2 for a more thorough discussion of the various types of migration costs.

[34] Alternatively, costs can be measured in consumption goods, and hence put directly into the budget constraint, which becomes: $c_j^i = w_j^i(1-l_j^i) - \gamma_{A,B}^i$. Theoretically, it makes no difference.

$$\max\left\{\lambda^i E\left[u\left(c_A^i l_A^i\right)\right] + \left(1-\lambda^i\right)E\left[u\left(c_B^i l_B^i\right)\right] - \gamma_{A,B}^i\right\} \tag{1.8}$$

where $\gamma_{A,B}^i$ represents the costs of an agent of skill i from region A migrating to region B. Notice that migration costs depends on skill level, indicating that human capital is once again playing a role in the migration decision. Abstracting from endogenous labor supply for the moment by fixing $l_j^i = 0 \,\forall\, i,j$, the probability that individual i will migrate from region B to region A is:

$$\lambda^i = \text{Prob}\left[\left(E\left[u\left(c_A^i\right)\right] - E\left[u\left(c_B^i\right)\right] - \gamma_{A,B}^i\right) > 0\right] \tag{1.9}$$

Often, utility is assumed to be linear (Ortega and Peri, 2009; Beine et al., 2011; Grogger and Hanson, 2011; Simpson and Sparber, 2013), such that $\lambda^i = \text{Prob}[E[c_A^i - c_B^i - \gamma_{A,B}^i] > 0]$ or $\lambda^i = \text{Prob}[E[w_A^i - w_B^i - \gamma_{A,B}^i] > 0]$. Given distributional assumptions on w_j^i, the emigration rate can be estimated using equation (1.9). If wages are exogenous and observable, λ^i can be defined as the ratio of the number of immigrants from origin region B in destination A ($M_{A,B}$) over the native population of the source region B (M_B). That is, $\lambda^i = M_{A,B}/M_B$ is the migration rate from region B to A, and it depends explicitly on the after-tax wage (or income) differential between the source and destination region net of migration costs. Assuming wages are known for certainty (an assumption to be relaxed later), the baseline model of endogenous migration yields the following specification:

$$\text{Migration rate}_{A,B}^i = w_A^i - w_B^i - \gamma_{A,B}^i \tag{1.10}$$

Equation (1.10) reinforces the point that the migration decision depends critically on human capital via wages and migration costs. Also notice that equation (1.10) can be expressed in terms of migrant flows since the migration rate is the ratio of the flow of immigrants from region B to A over the size of the population of region B. Equation (1.10) can be rearranged so that migrant flows are on the left-hand side of the equation and population size is on the right-hand side. This is often the specification used in recent gravity models of immigration, including Karemera et al. (2000), Lewer and Van den Berg (2008), Ortega and Peri (2009), Mayda (2010), and Beine et al. (2011).[35]

3.1.2 Incorporating immigration policy

Laws usually restrict the flow of migrants across countries, and sometimes restrict movements across regions within a country as well. Immigration policy can be modeled directly in migration costs, whereby costs for certain types of immigrants (identified by skill level and/or region of origin) are higher for those who face tighter immigration restrictions. As immigration policy tightens for an individual from region i with skill j, γ_j^i in equation (1.10) would rise, implying that the migration rate would fall for that group of

[35] Depending on the distributional assumptions of wages, equation (1.10) may also be written in log form as in Borjas (1987) and Ortega and Peri (2009), for example.

individuals. This is in line with Clark et al. (2007), for example, which specifies immigration costs that vary with immigration policy and skill level. Various other mechanisms could be imbedded into the model to incorporate immigration policy, such as including a measure that represents the probability that an immigrant from region B will be allowed to stay in region A (as in Mayda, 2010). In this case, there is an immigration quota in the destination region, and once the quota is reached, new immigrants are not allowed to enter and those in the destination region may be sent home.

Migration costs are the most straightforward and common way of accounting for immigration policy in theoretical models. However, immigration policy can also be modeled in other ways. For example, suppose border enforcement restricts the movement of people across borders and adds to the costs of migration by increasing the time spent moving. The time spent migrating results in lost wages in the source and destination. Thus, border enforcement can be modeled as a tax on migrant wages (see Guzman et al., 2008). Alternatively, immigration policy can take the form of interior enforcement schemes such as employer sanctions (such as in Ethier, 1986), whereby employers in the destination face increased hiring costs, especially in the case of illegal immigrants. In this way, immigration policy is raising the input costs for firms that hire immigrants. If these costs are passed along to workers in the form of lower wages, they reduce the incentive to migrate. If instead employers absorb the costs, output prices could rise.

3.1.3 Incorporating self-selection

The simple model in equation (1.10) allows wage differentials and migration costs to vary across skill type, so immigration rates depend critically on the (exogenous) level of human capital. However, variation in migration rates can also arise due to the unobservable characteristics and skills that potential migrants possess. This is the key insight of Borjas's (1987) self-selection model. Unobservable characteristics imply that migration rates will vary not only by human capital, but will also depend upon the *distribution* of the returns to human capital (wages) in the origin and destination regions.[36]

In this framework, the migration decision is driven by both the average relative wage gain from immigrating and whether the immigrants' skills would be rewarded by moving abroad (that is, by where the potential immigrant stands in the wage distribution). Notice that the returns to immigration are not only captured by wages, but other factors such as wage inequality, the progressivity of income taxes, and redistribution via social insurance. Importantly, the variance of earnings and the transferability of skills across regions are important components in the immigration decision. Equation (1.10) can be rewritten as

[36] Also see Borjas and Bratsberg (1996), which examined selection consequences in a model that allows people of different exogenous skill levels to decide both whether to migrate and whether to return.

$$\text{Migration rate}^{i}_{A,B} = \frac{\ln\left(w^{i}_{A}\right) - \ln\left(w^{i}_{B}\right) - \gamma^{i}_{A,B}/w^{i}_{A}}{\sigma} \tag{1.11}$$

where σ represents the standard deviation of the error terms in destination and origin region wages, which depends on each region's earnings variance, the degree of skill transferability across regions, and the interaction of the origin and destination earnings variances.[37] Note that the migration rate depends on relative wage differences and migration costs ($\gamma^{i}_{A,B}$) relative to the destination wage rate (w^{i}_{A}).

The main finding of the Borjas model is that immigration occurs when the destination offers higher relative returns to the individual's skill set (i.e., human capital), assuming mean wages are higher. Consequently, differences in income inequality and the transferability of skills are important determinants of immigration. In Borjas (1991), earnings variance is driven in part by observable characteristics (such as education and experience) such that the migration decision varies by the mean education level in each region, for example. This extension allows the model to predict that migration rates rise (fall) with the mean (variance) education level of the origin region. This has important implications for what types of individuals—skilled or unskilled—have an incentive to migrate.

3.1.4 Blending self-selection and migration costs

Migrant selectivity can arise due to distributional assumptions on observed or unobserved wage components, as discussed above. However, selectivity can also arise under alternative specifications for migration costs. For example, different from Borjas (1987), Chiswick (1999) assumes that skills are observable and that wages in the origin and destination do not depend on labor market experience. Instead, migration costs that do not depend on the wage or skill level play an important role in determining which types of migrants have an incentive to migrate. The baseline model in this case is similar to equation (1.10) but with more elaborate explicit and implicit migration costs (that do not depend on skill level i).[38] Specifically, higher explicit migration costs yield a positive selection bias for those who earn the highest wages in the destination since their gains are large enough to offset the high costs.

Alternatively, Chiquiar and Hanson (2005) extend the Borjas model to incorporate migration costs that decrease in education with self-selection in observable levels of education. In this environment, high-skill migrants face relatively low migration costs. This, combined with the selectivity of skills, can explain why Mexican immigrants in the US are more educated than their counterparts in Mexico, but less educated than the average US citizen.

[37] We have simplified notation for ease of exposition. A complete description of the Borjas (1987) model can be found in Bodvarsson and Van den Berg (2009).

[38] Once again, we refer the interested reader to Bodvarsson and Van den Berg (2009), who offer a comprehensive description of the Chiswick (1999) model.

In recent work, Grogger and Hanson (2011) developed a model where absolute differences in earnings and fixed migration costs are the primary determinants of migration, rather than relative earnings and costs (as in Borjas, 1987, 1991). In fact, our simple model outlined in equation (1.10) is very close to the specification of Grogger and Hanson (2011), but they include an unobserved idiosyncratic component in utility. They also decompose the migration costs into fixed monetary costs (specific to the origin–destination pair) and a component that varies by skill. The model of Grogger and Hanson (2011) suggests that increases in the absolute differences in earnings between high- and low-skilled workers in destination regions lead to more migration, and the mix of migrants is more skilled.

3.1.5 Accounting for income inequality

In equation (1.10), after-tax wages are the relevant measure of income since the model represents an individual's migration decision. This is consistent with microeconomic models of migration that use wages as a measure of income (such as Borjas, 1987, 1991; Chiquiar and Hanson, 2005; Orrenius and Zavodny, 2005; and Hunt, 2006). However, the model can be generalized to the macro-level by incorporating aggregate measures of income (such as GDP). In fact, some measure of income in the origin and/or destination country is included in almost every model explaining international migration. Recently, Clark et al. (2007), Lewer and Van den Berg (2008), Lewer et al. (2009), Ortega and Peri (2009), and Mayda (2010) all incorporate per-capita GDP (in the origin and/or destination country) as a key determinant of cross-country immigrant flows.

In addition to average income,the individual's relative income position within a society forms part of the decision to migrate, as discussed in Section 2.3.6 above. This provides a theoretical foundation for the empirical observation that migration rates in the poorest regions are not necessarily the highest, and that migration rates increase with income inequality. Borjas (1987) shows that, conditional on mean wages, high-skill immigrants from low inequality locations prefer to move to relatively high inequality locations, while the low skilled will prefer relatively low inequality locations (although Chiswick (1999) points out that this finding emerges due to the lack of fixed migration costs in the model). Stark (1991) also discusses income inequality as a determinant of immigration. According to Rotte and Vogler (1998, p. 5), "There is a higher incentive to migrate if one is poor among rich than if one is poor among poor." Rotte and Vogler (2000), Brücker and Defoort (2006), Brücker and Schröder (2006), Clark et al. (2007), Ortega and Peri (2009), and Mayda (2010) are among others to present theoretical models incorporating income inequality as a determinant of bilateral migration flows.

3.1.6 Introducing credit and poverty constraints

An important literature has emerged on the role of credit constraints in migration. There are often significant costs to migration, but if perfect credit markets exist, migrants could

borrow to finance these costs. However, credit markets are imperfect, and more so in developing countries. In addition, future income streams are uncertain with migration, making financial contracts difficult to impose. Such constraints provide an explanation for a phenomenon regularly found in the empirical literature and discussed earlier—that pull effects are stronger determinants of migration than push effects (see Hunt, 2006; Pedersen et al., 2008; Warin and Svaton, 2008; Zaiceva and Zimmermann, 2008; and Mayda, 2010). That is, poverty constraints and imperfect capital markets might prevent source region income from affecting migration decisions since worsening conditions simultaneously increase the incentive to leave while decreasing the ability to do so.

Orrenius and Zavodny (2005) built on the Borjas (1987) selection model to incorporate features involving access to credit markets. In their model, migrants must save to cover migration costs. Since access to formal and informal credit markets varies with income, Orrenius and Zavodny (2005) assume that savings and hence the migration decision depends on the level of human capital. By imposing a cash-in-advance constraint, their model argues that tight credit constraints and insufficient savings to cover up-front migration costs have worked to limit unskilled Mexican migration to the US below levels that would otherwise occur.

Belot and Hatton (2008) and Hatton and Williamson (2011) also include poverty constraints as a determinant of emigration. The premise is that potential migrants from developing countries who live near the subsistence level will not be able to provide collateral for future earnings since the earnings will be acquired abroad, making it difficult for the lender to recoup loan payments. However, migrant networks may mitigate the poverty constraint via remittances (which will be discussed more below) and financial support upon arrival. For individuals with high migration costs and low levels of skills, the poverty constraint will more likely be binding. For individuals with access to larger migrant networks, the poverty constraint will less likely be binding.[39]

3.1.7 Accounting for unemployment

In the simple model above, expected wages depend upon both wages of employees and the probability of being employed in each period. Thus, unemployment could be incorporated directly into the model via expected wages and can affect migrant selectivity (Karemera et al., 2000; Pedersen et al., 2008). Migrants are predicted to move to high wage and low unemployment regions (Hunt, 2006). In addition, unemployment rates are higher among illegal migrants (than legal ones) since it is more difficult to obtain employment and employers often face sanctions for hiring undocumented migrants. Thus, the flow of illegal immigrants is likely to be more responsive to unemployment rates than would be the flow of legal immigrants. In addition, attitudes toward

[39] Shen et al. (2010) also included credit constraints to analyze the effect of migration and remittances on inequality.

immigration tend to move with business cycles. As the macroeconomy worsens, so does the pressure to hire native workers, making it more difficult for legal and illegal immigrants to find work.

3.1.8 Incorporating taxes and social insurance

The model in equation (1.10) consists of wages net of taxes and government transfers. Both components can be treated as independent factors that affect the migration decision, such that equation (1.10) can be expressed explicitly in terms of taxes and transfers, as in

$$\text{Migration rate}^{i}_{A,B} = \left[w^{i}_{A}\left(1-\tau^{i}_{A}\right) + T^{i}_{A}\right] - \left[w^{i}_{B}\left(1-\tau^{i}_{B}\right) + T^{i}_{B}\right] - \gamma^{i}_{A,B} \qquad (1.12)$$

where τ^{i}_{j}, T^{i}_{j} are the tax rate and government transfers respectively in region j for a worker of skill i. Once again, the migration rate depends critically on human capital via taxes and government transfers, in addition to wages and migration costs. Notice that the migration rate falls with the tax rate and rises with government transfers at the destination region B, and vice versa with respect to the origin region A, *ceteris paribus*.

Government transfers in the destination (as a form of social insurance) may act as a magnet for immigrants. Welfare payments can be viewed as a substitute for earnings during the time spent searching for a job, mitigating unemployment risk. The theoretical model of Borjas (1999), for example, suggests that given high migration costs, immigrants will geographically sort themselves into US states that offer generous welfare benefits, and more so than natives. Borjas and Trejo (1993) found that origin country characteristics explain a significant share of welfare participation rates in the United States. However, several studies have found no such effect, including Zavodny (1997), Urrutia (1998), and Pedersen et al.(2008).

Retirement or pension plans that expropriate wage income early in life and provide subsidies later in life have the potential to affect the migration decision at various stages of the life-cycle. For example, the degree of portability of pension plans may influence the migration decision. In addition, the generosity of public pension plans may act as an additional magnet for immigrants, or the high taxes needed to finance these generous systems may deter immigration. While there is literature that examines the role that immigration may play in financing pension plans (Storesletten, 2000; Krieger, 2005), little work exists that considers how pension plans (private or public) influence the migration decision.

Notice that in equation (1.12), if the government budget is assumed to balance in each region, the transfers T^{i}_{j} would drop out of the migration rate, such that only the tax rate matters in the migration decision. High tax rates among destination regions may detract immigrants by reducing the relative return to labor, but tax rates are typically highly (positively) correlated with public social expenditure, which may attract immigrants (according to the welfare magnet theory). Thus, it is not clear in equation (1.12) how taxation in both the origin and destination regions affects migration. In practical terms, tax rates are relevant only for relatively rich origin and destination countries in which income taxes

are collected, and for high-skilled individuals who face non-trivial income tax rates. However, little work has been done that isolates the effect of taxation on international migration. One exception is Andersen (2005), who analyzes the role that taxes play in reducing the incentive for skilled workers to emigrate.

3.1.9 Accounting for political institutions

Recent work has explored the role of political institutions in influencing immigrant flows. Following Mayda (2010), one way to embed political institutions into a theoretical migration model is to assume that pairs of destination and origin countries with similar institutions should have lower migration costs. Empirical work often proxies this by identifying countries with common colonial ties (Mayda, 2010) or measures of political freedom (Rotte and Vogler, 2000). Alternatively, Hatton and Williamson (2011) consider the role of political events in the origin country on immigration, including civil wars, upheavals, and abuse of human and civil rights. In their model, political rights are captured by potential immigrants having non-economic preferences for their origin country. Equation (1.10) can be rewritten as:

$$\text{Migration rate}^{i}_{A,B} = w^{i}_{A} - w^{i}_{B} - \gamma^{i}_{A,B} - z^{i}_{A,B} \tag{1.13}$$

where $z^{i}_{A,B}$ represents the compensating differential of an individual's non-economic preference to immigrate from region A to region B. For example, if the political situation in the origin is bad relative to the situation in the destination, $z^{i}_{A,B}$ could be negative, increasing the benefits of migration and hence the migration rate.

In addition, political rights, individual freedom, and political instability may affect the migration decision, as empirically documented in Karemera et al. (2000). These ideas stem from Borjas (1989), who discusses how political conditions may affect the non-random sorting of immigrants, especially for refugees. Migration might also affect corruption in the destination country. For example, Mariani (2007) developed a model of migration in which high-skilled workers moving abroad choose between rent-seeking behavior (i.e., corruptive activities) and productive activities. Mariani (2007) found that with endogenous migration, "There is even less room for a positive role of skilled labor mobility in the perspective of reducing rent-seeking (in the destination country)" (p. 627).

3.2 A static human capital model with endogenous migration and endogenous wages

Wages are exogenous in the models presented in the previous subsection as there is no general equilibrium adjustment when labor moves across countries (or regions). We now relax this assumption and allow the stock of unskilled labor in each region to depend upon three things: (i) the measure of unskilled agents in the world (Ω_u); (ii) the fraction of unskilled agents who migrate to that region (λ^u to region A and $1-\lambda^u$ to region B); and (iii) the amount of labor supplied by unskilled agents, $(1-l^u)$. Note that labor supply

is now elastic (above it was assumed to be inelastic). The quantities of unskilled labor in regions A and B are $U_A = \Omega_u \lambda^u (1 - l^u)$ and $U_B = \Omega_u (1 - \lambda^u)(1 - l^u)$. Similarly, the amount of skilled labor in each region depends on the measure of skilled agents (Ω_s), the fraction of skilled agents who migrate to that country (λ^s in country A and $1 - \lambda^s$ in country B), and the amount of labor supplied by skilled agents ($1 - l^u$) so that $S_A = \Omega_s \lambda^s (1 - l^s)$ and $S_B = \Omega_s (1 - \lambda^s)(1 - l^s)$.

In this framework, profit maximization of the representative firm delivers first-order conditions that determine wages. Depending on the specification of the production function $F_j(Z_j, U_j, S_j)$ in each region j where Z_j represents total factor productivity (TFP), unskilled and skilled labor could be complements or substitutes in production. Either way, diminishing marginal returns imply that an increased number of unskilled (skilled) workers in a region will reduce the unskilled (skilled) wage rate. Hence, unskilled wages in region j are now: $w_j^u = \partial F_j(Z_j, U_j, S_j)/\partial U_j$ and skilled wages are $w_j^s = \partial F_j(Z_j, U_j, S_j)/\partial S_j$. Migration will change the relative returns to skills that each region offers. For example, a large inflow of unskilled immigrants into a region will lower the unskilled wage rate, reducing the incentive for more immigration. In addition, TFP in each region will have important effects on the relative wage in each region. In equilibrium, individuals must still be indifferent to living in the origin and destination region, once migration costs are considered.

With endogenous wages, the model yields the following specification:

$$\text{Migration rate}^i_{A,B} = M\left(Z_A, U_A, S_A, Z_B, U_B, S_B, \gamma^i_{A,B}\right) \tag{1.14}$$

where $M(\,)$ is a function that depends on the production function specification $F(\,)$. Different from equation (1.10), the migration rate depends on the stock of unskilled and skilled labor in each region (which includes both domestic and foreign workers) and TFP differentials, but it does not depend on exogenous wage differentials.

This specification is consistent with the literature that considers schooling levels, demographics, and income levels as important determinants of immigration. For example, Hatton and Williamson (2011) focused on the supply side of immigration by considering how origin country demographics and education affect emigration from developing countries. In their specification, income depends on human capital levels in each country and an idiosyncratic component that may vary across countries. In addition, migration rates depend positively on the proportion of young adults in the population and on network effects (as discussed in Section 2.3.7). For example, Hunt (2006) claimed that younger people are more likely to migrate since they have (i) a longer time horizon to recoup migration costs; (ii) more time to benefit from good economic conditions abroad; (iii) less firm- and location-specific human capital than older people, lowering their costs of migration; and (iv) less on-the-job training and hence a lower opportunity cost of time, thus lowering their cost of human capital investment.

3.2.1 Accounting for network effects

Notice that with endogenous wages, the migration model suggests that the stock of workers (of each skill type) affects the return to migration. The stock of workers in the destination region is comprised of natives and previous migrants, and given the assumptions about the complementarities in production between workers of various skill levels and regions of origin, it is plausible that new immigrants benefit both directly and indirectly from having a network of migrants in the destination (as discussed in Section 2.3.2). This could be modeled explicitly, as a stock of workers from the same origin region, or via migration costs—migrants with a larger network may experience lower migration costs, increasing the net gain of migration (Carrington et al., 1996; Munshi, 2003; Pedersen et al., 2008; McKenzie and Rapoport, 2010). In fact, the pattern of migrant selectivity can heavily rely on the existence and use of networks. As a result, much of the immigration literature incorporates the stock of previous migrants from the origin country residing in the destination country as an important determinant in international migration, including work by Bartel (1989), Bauer and Zimmermann (1995), Zavodny (1997), Orrenius and Zavodny (2005), Clark et al. (2007), Mayda (2010), Hatton and Williamson (2011), Grogger and Hanson (2011), and Simpson and Sparber (2013).

3.2.2 Distinguishing between individual and household migration decisions

As noted in Section 2.3.5, most migration models focus on the individual's decision even though migration is often a family decision. If the family unit as a whole decides whether it will migrate or if it sends a member of the household abroad, then the model must incorporate the consumption, income, and migration costs of the entire family. Immigration is a form of insurance against household and macro-level shocks in the origin country. Remittances (which are discussed in more detail below) are the primary insurance mechanism for migrants and allow family members who migrate abroad the opportunity to send a portion of their income back home. In addition, families often help to relax the tight borrowing constraints that potential migrants face, especially to help finance the move. Recent work on migration as a family decision (Anam et al., 2008; Shen et al., 2010) indicates that real option theory can pinpoint the optimal migration time and length when households desire to diversify the location of family members in order to reduce income risk, for example.

3.2.3 The relationship between trade and migration

Iranzo and Peri (2009) provided a theoretical connection between trade and migration by expanding the Heckscher–Ohlin two-country trade model. Individuals are differentiated by skill, and countries possess different levels of technology. Firms in each country are monopolistically competitive, and gains from trade arise from increased product variety.

Technology complements skills, thereby implying that skilled agents from low-technology countries have the greatest incentive to migrate.

In the absence of trade and with low migration costs, Iranzo and Peri (2009) found that workers at the tails of the skill distribution have the incentive to migrate, consistent with the self-selection model. The presence of trade has small effects on the migration decision. Though this implies that the model offers predictions similar to more typical models of the determinants of migration, the implication sharply contrasts with the Heckscher–Ohlin model since differences in technology allow for factor price differences to persist, providing workers with a continued incentive to migrate across borders.

3.3 Dynamic models with endogenous migration and physical capital accumulation

In most of the discussion so far, migration has been treated as an event that occurs once, such that the benefits and costs of migration are experienced instantaneously (or at least are modeled as such). Developments in dynamic stochastic general equilibrium modeling have allowed researchers to model the migration decision within a more realistic setting. In this section, we highlight how the theory on international migration has evolved in recent years to incorporate the dynamics of the migration decision and show, once again, that human capital is a central component to this decision.

A few studies have endogenized migration in a dynamic general equilibrium framework. Examples include Galor (1986), Djajic (1989), Glomm (1992), Wong (1997), Urrutia (1998), Klein and Ventura (2009), and Mandelman and Zlate (2012). The standard dynamic model of endogenous migration extends the baseline model in equation (1.10) to incorporate physical capital accumulation. If physical capital is assumed to be mobile, then a fixed factor of production (such as land) with diminishing returns must be introduced to guarantee equilibrium in which labor and capital are moving across the two regions without moving costs (Wildasin, 1994; Simpson, 2001; Klein and Ventura, 2009). A life-cycle model of migration assumes that agents live for T periods, accumulate assets that are used as income streams in future periods, and make a migration decision at some point during their lifetimes that may be reversed in future periods to allow for return migration.

The utility an agent receives would then depend on her capital accumulation and migration decisions. For now, we assume that there are no borrowing/lending restrictions in the model and that there is no return migration. Also, human capital is exogenous (which we will relax below), such that agents are either unskilled or skilled. Agents compare the present value of lifetime income among four alternatives: (1) remain unskilled and stay in the home region; (2) remain unskilled and migrate; (3) invest in education and stay in the source region; or (4) invest in education and migrate. The value function for an agent represents the best choice of these four options. In this setting, agents face migration costs (which are once again specific to skill level within the region of origin).

With perfect capital mobility, the return to physical capital is equal across regions. In interior solution equilibria, an agent from region j must be indifferent to the four alternatives. Hence, the migration decision depends on the relative returns to skills in each region (net of taxes and government transfers), migration costs, the stock of unskilled and skilled workers in each region, the stock of physical capital in each region (K_A, K_B), the stock of land in each region (L_A, L_B), and the relationship between the production inputs (assuming endogenous factor prices). That is, the migration rate for individuals of skill i and age a would depend on:

$$\text{Migration rate}^{i,a}_{A,B} = M\left(Z_A, U_A, S_A, Z_B, U_B, S_B, \gamma^i_{A,B}, \tau^i_A, \tau^i_B, K_A, K_B, L_A, L_B\right) \quad (1.15)$$

where all variables are defined as before, i denotes skill level, A is the origin region and B is the destination, and assuming a government budget constraint holds in each region (the time subscripts are dropped for ease of exposition). If factor prices are exogenous, equation (1.15) would not depend on the inputs to production (Z_j, U_j, S_j, K_j, L_j) in region j but instead depends on exogenous factor prices (w_j, R, q_j), where R is the world interest rate (net of depreciation) and q_j is the return on land in each region. An important innovation in the dynamic setting from Klein and Ventura (2009) is that migration rates are heterogeneous with respect to age and hence can vary over the life-cycle. Note, however, that given the interrelatedness of the endogenous and exogenous variables over time, the researcher cannot obtain closed-form solutions for equation (1.15) under reasonable assumptions for utility, production, and human capital accumulation functions. Most often, the researcher must resort to computational methods to obtain solutions, following Urrutia (1998), Klein and Ventura (2009), and Mandelman and Zlate (2012).

Urrutia (1998) uses a dynastic overlapping generations model to analyze the effect of migration costs on the self-selection of immigrants, building on the work of Borjas (1987) and Chiswick (1999) in which human capital is exogenous. Migration costs, which include fixed costs and loss of ability (as in Chiswick, 1999), dictate the selectivity of migrants in a dynamic model of endogenous migration. When the fixed cost is relatively low, immigrants are selected from the bottom of the ability distribution. The opposite occurs if the fixed migration cost is relatively high.

The dynamic model of Klein and Ventura (2009) considered the role of TFP differentials across countries when barriers to entry exist. Barriers to entry include skill loss associated with migration and paying a fixed resource cost. In their framework, they find that modest productivity differentials lead to large increases in the migration rate. As a result, output differentials increase as capital chases labor into the country with the productivity advantage. They find that capital accumulation and mobility can accentuate the welfare losses of barriers to entry, consistent with the findings of Michael (2003). In fact, in their analysis of the transitional dynamics, Klein and Ventura (2009) found that the removal of barriers to entry has large and immediate effects on migration rates, but the effects dampen over time. The welfare gains associated with removing barriers to

entry are significant, and are much larger than the losses experienced by natives of the destination country.

3.3.1 Immigration policy in a dynamic framework

In a dynamic framework, the effects of immigration policy on migration decisions are more complicated. Guzman et al. (2008) developed an overlapping generations model of migration in which technological progress in the smuggling industry has important effects on migration and savings decisions. For example, when smugglers are more efficient, migrants spend less time evading border patrol and more time working and saving in the destination country. Higher savings results in higher wages abroad but lower returns to savings. The net effect on migration depends on which of these two effects is larger. Similarly, Guzman et al. (2008) showed how tighter immigration policy via more border enforcement (which is financed by higher lump sum taxes) has ambiguous effects on migration rates because it lowers the capital–labor ratio in the destination country. In their framework, the key mechanism is how smuggling and border enforcement affects savings, and how the effects on savings translate into important effects for the migration decision. Importantly, the model assumes that migrants can contribute to the destination country's capital stock, following Benhabib (1996).

3.3.2 Accounting for remittance behavior

In equation (1.15), the migration rate depends on the level of capital stock in each region. Migrants save throughout their lifetimes, and those savings can be used to finance consumption during retirement. However, migrants' savings are being increasingly used to finance contemporaneous consumption of the individual's family left back home (in the origin country). Worldwide, remittances represent a significant component of GDP for countries that are large suppliers of immigrants. In the model, remittances can be thought of as part of the savings decision, with that portion of the capital stock going back to the origin country. That is, the location of the capital stock matters, and must be accounted for in the model. The migration decision will depend on the expected remittance behavior of a potential migrant, as often dictated by other family members (requiring the model to be representative of a household rather than an individual). Or one can assume that remittances alleviate credit constraints of the migrant, lowering the costs of migration (as discussed below).

A recent paper by Shen et al. (2010) developed a dynamic model of the household's migration decision where households face liquidity constraints, bequests are used to smooth intergenerational income, and remittances equalize income within a household (of both migrants and non-migrants). The focus of the paper is to analyze how migration and remittance behavior affects income inequality in the origin country, but the framework allows for endogenous migration and remittance decisions.[40]

[40] A related paper by Rapoport and Docquier (2006) provided a review of the determinants of remittances.

Taking a different approach, Mandelman and Zlate (2012) quantified the extent to which immigration and remittances respond to business cycles in the origin and destination countries. Using a calibrated quantitative macroeconomic model, they found that immigration and remittance flows are procyclical such that in good times immigrant flows increase to the destination but in bad times they retreat. When the sunk costs of migration are lowered (which represent border enforcement), migrant flows are more responsive to business cycles. Important innovations in their paper consist of having both endogenous migration and remittance decisions, and assuming complex relationships between the inputs to production. For example, domestic and foreign labor are substitutes in production, whereas the capital of skilled and unskilled labor is imperfectly substitutable.

3.3.3 Dynamic models with human capital accumulation

Human capital is exogenous in the dynamic migration model described in equation (1.15). However, the model can be adjusted so that one may analyze the interaction between endogenous human capital accumulation and migration. In this environment, agents are endowed with an initial level of human capital, h_0^i (i.e. inherent ability), and benefit from the existing stock of human capital, H_t, in region j. Human capital evolves depending on the individual time investment (s_t^i, if labor supply is endogenous), explicit costs of financing education such as tuition fees (e_j), the migration decision (λ_t^i), and human capital depreciation via skill loss as a result of imperfect transferability of skills (δ_j^i). A human capital accumulation function could be specified as:

$$h_{t+1}^i = H\left(H_t^j, h_0^i, s_t^i, \lambda_t^i, e_j, \delta_j^i\right) \tag{1.16}$$

where H is a function that dictates the relationship between the various inputs of human capital over time. The standard assumptions regarding H are that human capital exhibits diminishing returns and that the inputs to production are complementary in the production of new human capital (Dustmann and Glitz, 2011). In this case, the migration rate would depend on education costs in each region (e_j) and migration costs, in addition to all of the other factors of production. The timing of the human capital decision is also important, since human capital accumulation could occur either simultaneously with or preceding the migration decision.

Chiswick and Miller (2011) considered the role of human capital investment for migration between two countries with similar income levels and a high degree of skill transferability. In the typical setting with imperfect skill transferability across countries, immigrants have the incentive to invest in their human capital following migration as a way to increase their earnings abroad and to catch up to natives. However, when skills are perfectly transferable across countries (i.e., two rich countries with similar culture and language), there is no incentive for human capital accumulation following migration. In this setting, migration is a "two-way street." The earnings of immigrants, which are

initially higher than natives, decline towards those of natives, allowing for the possibility of negative assimilation.

Dustmann and Glitz (2011) developed a comprehensive dynamic model with endogenous migration and human capital accumulation (but with exogenous wages and no savings decision). Their model allows for different types of migration patterns: the individual can migrate temporarily or permanently, and human capital can be acquired both at home and abroad. Hence, they incorporate migration decisions regarding the initial migration and the length of the migration. In this framework, migration decisions depend on the relative returns to skill in both the origin and destination, in addition to the length of migration (which, among other factors, affects the degree of skill transferability across countries). Similar to our dynamic framework (in equation (1.15)), the migration decision depends on the initial level of ability and the age of the migrant. Those with high ability and who migrate earlier in life experience steeper wage profiles and higher wage growth.

The Dustmann and Glitz (2011) model also can be used to analyze the flow of international students. For example, migrants who obtain skills abroad may be induced to return home to take advantage of the higher returns to skill. Earlier work by Rosenzweig (2006) developed a theoretical model of international student flows, where the transferability of skills across countries varies for individuals. For individuals who obtain schooling abroad, their skills are more easily transferable, leading to a higher return for skills and a higher likelihood of obtaining employment in the destination country.

3.3.4 Accounting for temporary, return, and circular migration

Most theories of migration treat it as an event that occurs only once. That is, individuals (or households) make the migration decision, and if they migrate, they remain in the destination country.[41] The dynamic model of equation (1.15) similarly ignores the possibility of multiple migration decisions. However, dynamic theoretical models of migration exist in which the initial migration is exogenous but the migrant living abroad can choose to move back home at some point in time. Galor and Stark (1990) provided a canonical two-period model, and found that the level of savings explicitly depends on the likelihood of returning home. Both Hill (1987) and Djajic and Milbourne (1988) developed life-cycle models of migration in which agents have a preference for location. Agents can choose total time allocated to working in the home and foreign country, in addition to the number of trips. Djajic and Milbourne (1988) focused on guest-worker migration and considered the interaction between savings and the length of stay in the foreign country. Recently, Dustmann and Mestres (2011) investigated the interaction between savings (physical capital accumulation) and return migration. The amount saved may affect when the agent finds it optimal to return home, and vice versa.

[41] Borjas and Bratsberg (1996) and Polachek and Horvath (1977) are early exceptions.

Several papers by Dustmann and co-authors have recently incorporated decisions regarding migration duration that aid in analyzing temporary migration (see Dustmann, 1997, 2003; Dustmann and Kirchkamp, 2002; Dustmann et al., 2011; Dustmann and Glitz, 2011; and Dustmann and Mestres, 2011). For example, Dustmann (2003) developed a dynamic model of return migration to determine how wage differentials between the origin and destination country affect the optimal migration duration. The stock of labor in each country depends not only on the migration decision and the net flow of migrants, but also on the duration of migration. Migration decisions are endogenous. As the stock of labor in each country adjusts, so do wages, which changes the migration decision. Durations decrease in response to increasing wage differentials.

Dustmann et al. (2011) allowed for the possibility that individuals obtain their human capital abroad and return home to capitalize on the relative high domestic returns to skill. That is, the location of human capital accumulation depends on the relative costs of skill acquisition and the rate at which skills are augmented in each country. They employ a dynamic Roy model where skills are two-dimensional such that skills evolve as migrants learn abroad (i.e. a "self-productive" strategy). As such, the selectivity of migrants depends on the returns to skill in each country and the composition of those who initially emigrate. The approach in Dustmann et al. (2011) is novel in that they obtained tractable solutions in a dynamic migration model and discussed how differences in skill transferability across countries can generate different migration outcomes, both in terms of emigration and return migration rates. This is quite promising for future work.

3.3.5 Dynamic models with human capital accumulation, circular migration, and brain drain

Dynamic models of migration and human capital formation are intimately tied to an important consequence of immigration feared by many policymakers—the potential exodus of high-skilled labor or what is often called "brain drain." In the dynamic migration model of Beine et al. (2008), individuals invest in education during their youth and choose to migrate in adulthood. Similar to the model above, human capital investment is costly and depends on ability and personal investment. In addition, individuals may face important credit constraints in acquiring education. To simplify the model, the authors assume that only skilled workers can migrate. Interestingly, the model predicts that losses from brain drain will be small as long as the skilled emigration rate does not get too high. Skilled migration raises the return on human capital in the source country, incentivizing non-migrants to invest in human capital. This "brain gain" helps offset the brain drain usually associated with skilled migration.[42]

[42] See Mountford (1997), Vidal (1998), and Beine et al. (2001, 2003) for further work indicating that the outflow of skills associated with brain drain may entice those at home to accumulate more skills. Fan and Yakita (2011) developed a two-period model of human capital to isolate the conditions when brain gain and brain drain occur.

Beine et al. (2008) are not alone in finding that brain drain fears may be overstated. Mayr and Peri (2009) combined the insights of Borjas and Bratsberg's (1996) model of return-migrant selection with those from the dynamic human capital acquisition literature in order to analyze joint education, migration, and return migration decisions under different policy environments. They found that both return migration and the education incentives of the sort envisioned by Beine et al. (2008) help to increase human capital levels in the source region. Furthermore, they demonstrated that the selection of return migrants depends upon both the initial selection of emigrants and whether wage premiums for having lived abroad vary across education level. Altogether, the evidence suggests that a richly specified model incorporating endogenous educational attainment and circular migration is necessary for a full understanding not only of the consequences of migration flows, but of their causes as well.

4. CONCLUSIONS AND IMPLICATIONS FOR FURTHER RESEARCH

In this chapter, we have surveyed theories of the economic determinants of migration. The chapter began by outlining the primary forces driving migration. Though some theories might stress the role of consumption, family decision-making, or various non-economic factors, the fundamental premise of most migration theories is that migration is driven by spatial differences in the net returns to factor supply and is a response to labor market disequilibrium. A big step in the development of migration theory was taken by Sjaastad (1962), who articulated a theory of migration as a form of human capital investment—people relocate because it adds to lifetime earnings.

In general, migration models are adept at describing both internal and international migration. That is, the human capital model is suitable for explaining both forms of labor flows. Nonetheless, authors have added specific features to the human capital model to focus on one form of migration over another. For example, Borjas (1987) applied Sjaastad's (1962) model to develop a theory of international migration, arguing that migration is not only influenced by net earnings differences between countries, but also by factors such as international differences in income inequality and the degree of skills transferability. Recently, Clark et al. (2007) extended this model to include the effects of immigration policy. Other authors have suggested that people move to assuage feelings of relative deprivation, as a solution to a household portfolio diversification problem, or to exploit migrant network effects flowing from the destination country.

The large literature examining the impact of international migration on destination and/or origin regions typically assumes that migration is exogenous (i.e., the flow of migrants across regions is fixed). However, many recent models have taken an important step in incorporating endogenous migration into theoretical models in order to fully account for both the determinants and consequences of migration. Such models might incorporate decisions regarding whether or not to migrate, the duration of residence

away from home, and whether to migrate again. Factors influencing these decisions include the role of migration costs, immigration policy, income and income inequality, networks, family ties, political institutions, and trade. This is certainly not a comprehensive list, but it instead represents a starting point for researchers interested in considering other possible factors.

In an open-economy model (with two or more regions), the migration decision can be paired with other decisions, such as savings and schooling decisions, to allow for more complex and realistic depictions of the various pressures associated with migration. As we have documented, advances in recent years have allowed for these possibilities by using dynamic models of endogenous migration. The migration decision has long-lasting effects not only for migrants and their families, but also for residents of the origin and recipient regions. It is important to sort out how these effects feed into the individual migration decision and the consequent impact on aggregate migrant flows. For example, the general equilibrium effects that occur as large quantities of skilled and unskilled workers move across countries can be significant, and can change depending on how the capital stock evolves.

Despite the robust literature on the theoretical determinants of migration, we believe that plenty of work remains to be done. For example, a model that incorporates relocation and income remittance as joint decisions has yet to be fully explored. If remitting is the reason behind migration, then the migration decision has to be made with that in mind. Factors that affect remittances, such as the macroeconomic conditions in both the origin and destination countries, can be isolated so that policymakers can better understand how migrant flows respond to them. In fact, the extent to which business cycles in both the origin and destination affect the magnitude and composition of migrant flows is still not fully understood. Mandelman and Zlate (2012) represent an excellent starting point for this work. Also, work by Ben-Gad (2008) suggests that the degree of capital–skill complementarity in the production technology has important welfare implications for natives due to immigration. These production complementarities may be important to the immigration decision, as they change how skill-specific wages respond to the flow of unskilled and skilled labor and physical capital across countries. Determining the extent to which assumptions regarding the production technology change the size and composition of migrant flows is an avenue that is worth pursuing in the future.

Certainly, immigration policy is dynamic in that it responds to political and economic pressures over time. Currently, there is no dynamic political economy model that considers how immigration policy evolves over time and how it affects the composition and magnitude of migrant flows. Immigrant flows clearly respond to changes in policy, so understanding these linkages should be crucial in developing immigration policy.

We conclude by submitting that international migration flows are probably best explained by a unified theory that combines the current model of international migration

as human capital investment with roles for household portfolio diversification, consumption interests, and migrant network effects. Theoretical issues that remain unaddressed or unresolved include:

1. The effects of migrant age on the propensity to migrate. Young migrants sacrifice less home-specific human capital and face a longer stream of potential earnings in destination regions. Nonetheless, older migrants might have greater propensities to migrate if, for example, they have larger assets to finance the migration decision or have larger general human capital endowments that can command high returns in the destination. More generally, there has been no thorough theoretical examination into how the source region's age distribution influences emigration rates.
2. The effects of exchange rates. Since many international migrants remit a portion of destination country earnings back to their home countries, exchange rates will influence the returns to migration. Exchange rates have not, however, appeared in theoretical models of migration.
3. The joint migration and remittance decision. The decision to migrate may occur at the same time as the remittance choice. How these joint decisions are made and their implications are not well understood.
4. Distinct determinants of illegal immigration. Very little theory has attempted to address whether the decision to be an illegal immigrant differs structurally from the decision to migrate legally.

The above list is not intended to be exhaustive, but includes topics that may be particularly fruitful at this time. While the lack of a unified theory might seem daunting, it is also very exciting, for it means that much innovative research lies ahead.

REFERENCES

Anam, M., Chiang, S., Hua, L., 2008. Uncertainty and international migration: An option cum portfolio model. J. Labor Res. 29 (3), 236–250.
Andersen, T.M., 2005. Migration, taxation and educational incentives. Econ. Lett. 87 (3), 399–405.
Anderson, J., 1979. A theoretical foundation for the gravity equation. Am. Econ. Rev. 69, 106–116.
Bartel, A.P., 1989. Where do the new United States immigrants live? J. Labor Econ. 7 (4), 371–391.
Bauer, T., Zimmermann, K.F., 1995. Integrating the East: The Labour Market Effects of Immigration, CEPR (Centre for Economic Policy Research) Discussion Paper No. 1235.
Beals, R., Levy, M., Moses, L., 1967. Rationality and migration in Ghana. Rev. Econ. Stat. 49, 480–486.
Becker, G., 1964. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. National Bureau of Economic Research, New York.
Becker, G., 1965. A theory of the allocation of time. Econ. J. 75, 493–517.
Beine, M., Docquier, F., Rapoport, H., 2001. Brain drain and economic growth: Theory and evidence. J. Dev. Econ. 64, 275–289, Elsevier.
Beine, M., Docquier, F., Rapoport, H., 2003. Brain Drain and LDCs' Growth: Winners and Losers, IZA (Institute for the Study of Labor) Discussion Paper No. 819.
Beine, M., Docquier, F., Rapoport, H., 2008. Brain drain and human capital formation in developing countries: Winners and losers. Econ. J. 118 (528), 631–652, Royal Economic Society.

Beine, M., Docquier, F., Oden-Defoort, C., 2011. A panel data analysis of the brain gain. World Dev. 39 (4), 523–532.
Belot, M.V.K., Hatton, T.J., 2008. Immigrant Selection in the OECD. CEPR (Centre for Economic Policy Research) Discussion Paper No. 571.
Ben-Gad, M., 2008. Capital-skill complementarities and the immigration surplus. Rev. Econ. Dynam. 11, 335–365.
Benhabib, J., 1996. On the political economy of immigration. Eur. Econ. Rev. 40 (9), 1737–1743.
Bhagwati, J., Srinivasan, T., 1974. On reanalyzing the Harris-Todaro model: Policy rankings in the case of sector-specific sticky wages. Am. Econ. Rev. 64, 1063–1093.
Bodvarsson, Ö., Hou, J.W., 2010. The Effects of Aging on Migration in a Transition Economy: The Case of China. IZA Discussion Paper 5070, http://www.iza.org.
Bodvarsson, Ö., Van den Berg, H.F., 2009. The Economics of Immigration: Theory and Policy. Springer, Heidelberg, Germany.
Borjas, G.J., 1987. Self-selection and the earnings of immigrants. Am. Econ. Rev. 77 (4), 531–553.
Borjas, G.J., 1989. Economic theory and international migration. Intern. Mig. Rev 23 (3), 457–485.
Borjas, G.J., 1991. Immigration and self-selection. In: Abowd, J.M., Freeman, R.B. (Eds.), Immigration, Trade and the Labor Market. University of Chicago Press, pp. 29–76.
Borjas, G.J., 1999. Immigration and welfare markets. J. Lab. Econ. 17 (4), 607–637.
Borjas, G.J., Bratsberg, B., 1996. Who leaves? The outmigration of the foreign-born. Rev. Econ. Stat. 78 (1), 165–176.
Borjas, G., Trejo, S.J., 1993. National origin and immigrant welfare recipiency. J. Publ. Econ. 50 (3), 325–344, Elsevier.
Brücker, H., Defoort, D., 2006. Inequality and the (self-)selection of international migrants: theory and novel evidence, IZA Discussion Paper 2052.
Brücker, H., Schröder, P.J.H., 2006. International Migration with Heterogeneous Agents: Theory and Evidence, IZA (Institute for the Study of Labor) Discussion Paper No. 2049.
Burda, M., 1993. The determinants of East-West German migration. Eur. Econ. Rev. 37, 452–461.
Burda, M., 1995. Migration and the option value of waiting. Econ. Soc. Rev. 27, 1–19.
Calvo, G., 1978. Urban unemployment and wage determination in LDCs: Trade unions in the Harris-Todaro model. Int. Econ. Rev. 19, 65–81.
Carrington, W.J., Detragiache, E., Vishwanath, T., 1996. Migration with endogenous moving costs. Am. Econ. Rev. 86 (4), 909–930.
Chiquiar, D., Hanson, G.H., 2005. International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States. J. Polit. Econ. 113 (2), 239–281.
Chiswick, B., 1999. Are immigrants favorably self-selected? Am. Econ. Rev. 89 (2), 181–185.
Chiswick, B.R., Miller, P.W., 2011. The "negative" assimilation of immigrants: A special case. Ind. Labor Relat. Rev. 64 (3), 502–525.
Clark, X., Hatton, T.J., Williamson, J.G., 2007. Explaining U.S. immigration, 1971–1998. Rev. Econ. Stat. 89 (2), 359–373.
Corden, W., Findlay, R., 1975. Urban unemployment, intersectoral capital mobility and development policy. Economica 42, 59–78.
David, P., 1974. Fortune, risk, and the micro-economics of migration. In: David, P., Reder, M. (Eds.), Nations and Households in Economic Growth. Academic Press, New York.
Deardorff, A., 1998. Determinants of bilateral trade: Does gravity work in a classical World? In: Frankel, J. (Ed.), Regionalization of the World Economy. University of Chicago Press, Chicago.
Djajic, S., 1989. Skills and the pattern of migration: The role of qualitative and quantitative restrictions on international labour mobility. Int. Econ. Rev. 30, 795–809.
Djajic, S., Milbourne, R., 1988. A general equilibrium model of guest-worker migration: A source-country perspective. J. Int. Econ. 25, 335–351.
Douglas, P., 1919. Is the new immigration more unskilled than the old? Reprinted in: Zimmermann, K., Bauer, T., 2002. The Economics of Migration. vol. 3. Elgar, Northampton, MA.
Dustmann, C., 1997. Return migration, uncertainty and precautionary savings. J. Dev. Econ. 52, 295–316.

Dustmann, C., 2003. Return migration, wage differentials, and the optimal migration duration. Eur. Econ. Rev. 47 (2), 353–369.
Dustmann, C., Glitz, A., 2011. Migration and education. In: Hanushek, E., Machin, S., Woessman, L. (Eds.), Handbook for the Economics of Education. vol. 4. North Holland, Amsterdam, The Netherlands, pp. 327–439.
Dustmann, C., Kirchkamp, O., 2002. The optimal migration duration and activity choice after re-migration. J. Dev. Econ. 67, 351–372.
Dustmann, C., Mestres, J., 2011. Savings, Asset Holdings, and Temporary Migration. IZA (Institute for the Study of Labor) Discussion Paper No. 5498.
Dustmann, C., Fadlon, I., Weiss, Y., 2011. Return migration, human capital accumulation and the brain drain. J. Dev. Econ. 95 (1), 58–67.
Ethier, W.J., 1986. Illegal immigration. Am. Econ. Rev. 76 (2), 258–262.
Fan, X., Yakita, A., 2011. Brain drain and technological relationship between skilled and unskilled labor: Brain gain or brain loss? J. Popul. Econ. 24 (4), 1359–1368, Springer.
Fields, G., 1979. Place-to-place migration: Some new evidence. Rev. Econ. Stat. 56, 21–32.
Gallaway, L., 1969. Age and labor mobility patterns. South. Econ. J. 36, 171–180.
Galor, O., 1986. Time preference and international labor migration. J. Econ. Theor. 38, 1–20.
Galor, O., Stark, O., 1990. Migrants' savings, the probability of return migration and migrants' performance. Int. Econ. Rev. 31 (2), 463–467.
Ghatak, S., Levine, P., Price, S., 1996. Migration theories and evidence: An assessment. J. Econ. Lit. 10, 159–198.
Glaeser, E., Shapiro, J., 2003. Urban growth in the 1990s: Is city living back? J. Reg. Sci. 43, 139–165.
Glomm, G., 1992. A model of growth and migration. Canadian Journal of Economics 24 (4), 901–922.
Goodman, J., 1981. Information, uncertainty, and the microeconomic model of migration decision making. In: De Jong, G., Gardner, R. (Eds.), Migration Decision Making: Multidisciplinary Approaches to Micro-Level Studies in Developed and Developing Countries. Pergamon Press, New York.
Gorter, C., Nijkamp, P., Poot, J., 1998. Regional and urban perspectives on international migration: An overview. In: Gorter, C., Nijkamp, P., Poot, J. (Eds.), Crossing Borders: Regional and Urban Perspectives on International Migration. Brookfield Ashgate, UK.
Graves, P., 1979. A life-cycle empirical analysis of migration and climate, by race. J. Urban Econ. 6, 135–147.
Graves, P., 1983. Migration with a composite amenity: The role of rents. J. Reg. Sci. 23, 541–546.
Green, G., Deller, S., Marcouiller, D., 2006. Amenities and Rural Development Theory, Methods, and Public Policy. Elgar, Northampton, MA.
Greenwood, M., 1975. Research on internal migration in the United States: A survey. J. Econ. Lit. 13, 397–433.
Greenwood, M., 1985. Human migration: theory, models and empirical studies. J. Reg. Sci. 25, 521–544.
Greenwood, M., 1997. Internal migration in developed countries. In: Rosenzweig, M., Stark, O. (Eds.), Handbook of Population and Family Economics. Elsevier, Amsterdam.
Grogger, J., Hanson, G.H., 2011. Income maximization and the selection and sorting of international migrants. J. Dev. Econ. 95 (1), 42–57, Elsevier.
Guzman, M.G., Haslag, J.H., Orrenius, P.M., 2008. On the determinants of optimal border enforcement. Econ. Theor. 34 (2), 261–296.
Hansen, M., 1940. The Mingling of the Canadian and American Peoples. Russel & Russel, New York.
Harris, J., Todaro, M., 1970. Migration, unemployment and development: A two-sector analysis. Am. Econ. Rev. 60, 126–142.
Hatton, T.J., Williamson, J.G., 2011. Are Third World emigration forces abating? World Dev. 39 (1), 20–32.
Hicks, J., 1932. The Theory of Wages. Macmillan, London.
Hill, J.K., 1987. Immigrant decisions concerning duration of stay and migratory frequency. J. Dev. Econ. 25 (1), 221–234.
Hugo, G., 1981. Village-community ties, village norms, and ethnic and social networks: A review of evidence from the third World. In: Dejong, J., Gardner, R. (Eds.), Migration Decision Making: Multidisciplinary Approaches to Micro-Level Studies in Developed and Developing Countries. Pergamon Press, New York.

Hunt, J., 2006. Staunching emigration from East Germany: Age and the determinants of migration. J. Eur. Econ. Assoc. 4 (5), 1014–1037.
Iranzo, S., Peri, G., 2009. Migration and trade: Theory with an application to the Eastern-Western European integration. J. Int. Econ. 79 (1), 1–19.
Jerome, H., 1926. Migration and Business Cycles. National Bureau of Economic Research, New York.
Karemera, D., Oguledo, V.I., Davis, B., 2000. A gravity model analysis of international migration to North America. Appl. Econ. 32 (13), 1745–1755.
Katz, E., Stark, O., 1986. Labor migration and risk aversion in less developed countries. J. Labor Econ. 4 (1), 134–149.
Klein, P., Ventura, G., 2009. Productivity differences and the dynamic effects of labor movements. J. Monetary Econ. 56, 1059–1073.
Krieger, T., 2005. Public Pensions and Immigration: A Public Choice Approach. Elgar, Northampton, MA.
Kuznets, S., Rubin, R., 1954. National Bureau of Economic Research Occasional Paper 46. Immigration and the Foreign Born. National Bureau of Economic Research, New York.
Lall, S., Selod, H., Shalizi, Z., 2006. Rural–Urban Migration in Developing Countries: A Survey of Theoretical Predictions and Empirical Findings, World Bank Policy Research Working Paper 3915.
Lancaster, K., 1966. A new approach to consumer theory. J. Polit. Econ. 74, 132–157.
Lewer, J.J., Van den Berg, H., 2008. A gravity model of immigration. Econ. Lett. 99 (1), 164–167.
Lewer, J.J., Pacheco, G., Rossouw, S., 2009. Do Non-Economic Quality of Life Factors Drive Immigration. IZA Discussion Paper No. 4385.
Linneman, H., 1966. An Econometric Study of International Trade Flows. North-Holland, Amsterdam.
Lucas, R., 1997. Internal migration in developing countries. In: Rosenzweig, M., Stark, O. (Eds.), Handbook of Population and Family Economics. Elsevier, Amsterdam.
Lundborg, P., 1991. An interpretation of the effects of age on migration: Nordic migrants' choice of settlement in Sweden. South. Econ. J. 58, 392–405.
Mandelman, F., Zlate, A., 2012. Immigration, remittances and business cycles. J. Monetary Econ. 59 (2), 196–213, Elsevier.
Mariani, F., 2007. Migration as an antidote to rent-seeking. J. Dev. Econ. 84 (2), 609–630.
Massey, D., Garcia Espana, F., 1987. The social process of international migration. Science 237, 733–738.
Massey, D., Arango, J., Graeme, H., Kouaouchi, A., Pellegrino, A., Taylor, J., 1993. Theories of international migration: A review and appraisal. Popul. Dev. Rev. 19, 431–466.
Mayda, A.M., 2010. International migration: A panel data analysis of the determinants of bilateral flows. J. Popul. Econ. 23 (4), 1249–1274.
Mayr, K., Peri, G., 2009. Brain drain and brain return: Theory and application to Eastern–Western Europe. B. E. J. Econ. Anal. Pol. 9 (1), Article 49.
McCall, B., McCall, J., 1987. A sequential study of migration and job search. J. Labor Econ. 5, 452–476.
McKenzie, D., Rapoport, H., 2010. Self-selection patterns in Mexico-U.S. migration: The role of migration networks. Rev. Econ. Stat. 92 (4), 811–821.
Michael, M.S., 2003. International migration, income taxes and transfers: A welfare analysis. J. Dev. Econ. 72, 401–411, Elsevier.
Mincer, J., 1978. Family migration decisions. J. Polit. Econ. 86 (5), 749–773.
Molho, I., 1986. Theories of migration: A review. Scot. J. Polit. Econ. 33, 396–401.
Mountford, A., 1997. Can a brain drain be good for growth in the source economy. J. Dev. Econ. 53 (2), 287–303.
Munshi, K., 2003. Networks in the modern economy: Mexican migrants in the U.S. labor market. Q. J. Econ. 118 (2), 549–599.
Orrenius, P.M., Zavodny, M., 2005. Self-selection among undocumented immigrants from Mexico. J. Dev. Econ. 78, 215–224, Elsevier.
Ortega, F., Peri, G., 2009. The Causes and Effects of International Migration: Evidence from OECD Countries 1980–2005. NBER. Working Paper No. 14833.
Pedersen, P.J., Pytlikova, M., Smith, N., 2008. Selection and network effects: Migration flows into OECD countries 1990–2000. Eur. Econ. Rev. 52, 1160–1186.

Pickles, A., Rogerson, P., 1984. Wage distribution and spatial preferences in competitive job search and migration. Reg. Sci. 18, 131–142.
Polachek, S., Horvath, F., 1977. A life cycle approach to migration: Analysis of the perspicacious peregrinator. In: Ehrenberg, R. (Ed.), Research in Labor Economics. JAI Press, Greenwich, CT.
Rapoport, H., Docquier, F., 2006. The economics of migrants' remittances. Handbook on the Economics of Giving, Reciprocity and Altruism. Elsevier.
Ravenstein, E., 1889. The laws of migration. J. Roy. Stat. Soc. 52, 241–305.
Rogerson, R., 1988. Indivisible labor, lotteries and equilibrium. Journal of Monetary Theory and Policy 21 (1), 3–16.
Rosen, S., 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. J. Polit. Econ. 82, 34–55.
Rosenzweig, M., 2006. Global wage differences and international student flows. In: Collins, S., Graham, C. (Eds.), Brookings Trade Forum 2006: Global Labor Markets? Brookings Institution Press, Washington, DC.
Rotte, R., Vogler, M., 1998. Determinants of International Migration: Empirical Evidence for Migration from Developing Countries to Germany. IZA (Institute for the Study of Labor) Discussion Paper No. 12.
Rotte, R., Vogler, M., 2000. The effects of development on migration: Theoretical issues and new empirical evidence. J. Popul. Econ. 13 (3), 485–508, Springer.
Runciman, W., 1996. Relative Deprivation and Social Justice. Routledge and Kegan Paul, London.
Sandell, S., 1977. Women and the economics of family migration. Rev. Econ. Stat. 59, 406–414.
Schlottmann, A., Herzog, H., 1984. Career and geographic mobility interactions: Implications for the age selectivity of migration. J. Hum. Resour. 19, 72–86.
Schultz, T., 1961. Investment in human capital. Am. Econ. Rev. 51, 1–17.
Schwartz, A., 1973. Interpreting the effects of distance on migration. J. Polit. Econ. 81, 1153–1169.
Schwartz, A., 1976. Migration, age, and education. J. Polit. Econ. 84, 701–719.
Shen, I., Docquier, F., Rapoport, H., 2010. Remittances and inequality: A dynamic migration model. Journal of Economic Inequality 8 (2), 197–220, Springer.
Shields, G., Shields, M., 1989. The emergence of migration theory and a suggested new direction. J. Econ. Surv. 3, 277–304.
Simpson, N., 2001. Redistributive Policies and Migration. Ph.D. Dissertation, University of Iowa.
Simpson, N., Sparber, C., 2013. The short- and long-run determinants of unskilled immigration into U.S. states. South. Econ. J. 80 (2), 414–438.
Sjaastad, L., 1962. The costs and returns of human migration. J. Polit. Econ. 70, 80–93.
Smith, A., 1776. An Inquiry into the Nature and Causes of the Wealth of Nations, 1937 edition. Modern Library, New York.
Stark, O., 1984. Rural-to-urban migration in LDCs: A relative deprivation approach. Econ. Dev. Cult. Change 32, 475–486.
Stark, O., 1991. The Migration of Labor. Basil Blackwell, Oxford.
Stark, O., Levhari, D., 1982. On migration and risk in LDCs. Economic Development and Culture Change 31 (1), 191–196.
Stark, O., Taylor, J., 1989. Relative deprivation and international migration. Demography 26, 1–14.
Stark, O., Taylor, J., 1991. Migration incentives, migration types: The role of relative deprivation. Econ. J. 101, 1163–1178.
Stark, O., Yitzhaki, S., 1988. Labor migration as a response to relative deprivation. J. Popul. Econ. 1, 57–70.
Storesletten, K., 2000. Sustaining fiscal policy through immigration. J. Polit. Econ. 108 (2), 300–324.
Taylor, J., 1986. Differential migration, networks, information and risk. In: Stark, O. (Ed.), Research in Human Capital and Development. In: vol. 4. JAI Press, Greenwich, CT.
Tiebout, C., 1956. A pure theory of local expenditures. J. Polit. Econ. 64, 416–425.
Tinbergen, J., 1962. Shaping the World Economy. Twentieth Century Fund, New York.
Todaro, M., 1969. A model of labour migration and urban unemployment in less developed countries. Am. Econ. Rev. 59, 138–148.
Todaro, M., 1976. Internal Migration in Developing Countries. International Labour Organization, Geneva.
Urrutia, C., 1998. On the Self-Selection of Immigrants. Universiad Carlos III de Madrid, Mimeo.

Vidal, J.P., 1998. The effect of emigration on human capital formation. J. Popul. Econ. 11 (4), 589–600.
Warin, T., Svaton, P., 2008. European migration: Welfare migration or economic migration? Global Econ. Q. 8 (3), 1–30.
Wildasin, E.D., 1994. Income redistribution and migration. Canadian Journal of Economics 27, 637–656.
Willis, R., 1973. A new approach to the economic theory of fertility behavior. J. Polit. Econ. 81 (supplement), 514–565.
Wong, K., 1997. Endogenous growth and international labor migration: the case of a small open economy. In: Jensen, B., Wong, K. (Eds.), Dynamics, Economic Growth, and International Trade. University of Michigan Press, Ann Arbor, pp. 289–336.
Yap, L., 1977. The attraction of cities: A review of the migration literature. J. Dev. Econ. 4, 239–264.
Zaiceva, A., Zimmermann, K.F., 2008. Scale, diversity, and determinants of labour migration in Europe. Oxf. Rev. Econ. Pol. 24 (3), 427–451.
Zavodny, M., 1997. Welfare and the locational choices of new immigrants. Economic Review – Federal Reserve Bank of Dallas 1997, 2–10, Second Quarter 1997.
Zipf, G., 1946. The [P(1)P(2)/D] hypothesis; On the intercity movement of persons. Am. Sociol. Rev. 11, 677–686.