DISCUSSION PAPER SERIES

IZA

IZA DP No. 5070

The Effects of Aging on Migration in a Transition Economy: The Case of China

Örn B. Bodvarsson
Jack W. Hou

July 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Effects of Aging on Migration in a Transition Economy: The Case of China

Örn B. Bodvarsson
St. Cloud State University
and IZA

Jack W. Hou
California State University, Long Beach
and Nankai University

Discussion Paper No. 5070
July 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5070
July 2010

ABSTRACT

The Effects of Aging on Migration in a Transition Economy: The Case of China*

China has been experiencing two major demographic sea changes since the late 1970s: (i) Internal migration, primarily rural-to-urban, on a scale that dwarfs all other countries at any time in history; and (ii) a shift in its age distribution. The basic question posed in this paper is: How are aging and migration related in post-reform China? We argue that there is probably two-way causality: Shifts in the origin region's age distribution induce changes in the scale and structure of migration, but out- (in-) migration shifts the origin's (destination's) age distribution. We examine theoretically and empirically the relationship between origin age distribution and interprovincial migration in China using province-level census data for 1985-2005. The goal of the paper is two-fold: (i) To develop a more refined theoretical model that explains how a migrant's age affects his/her likelihood of migration; and (ii) to obtain unbiased estimates of the effect of age on the interprovincial migration rate. Our theory section is motivated by the observation that, while most researchers recognize the importance of including age in theoretical and empirical models of migration, the exact reasons for why age affects migration have not been analyzed very thoroughly. We model the migration decision and demonstrate that there is an ambiguous relationship between age and the likelihood of migration. Implications of the theory are tested with an extended modified gravity model using OLS and 2SLS.

JEL Classification: J61, J11

Keywords: internal migration, age distribution, reforms

Corresponding author:

Örn B. Bodvarsson
Department of Economics
Department of Management
St. Cloud State University
St. Cloud, MN 56301-4498
USA
E-mail: obbodvarsson@stcloudstate.edu

* We thank Shuming Bao for invaluable assistance with data collection, as well as Ergan Xu, Robert Girtz, Asm Ulah, Li Lu, and Matthew Jensen for assistance with preparation of the data for analysis. We bear sole responsibility for any errors and request that we not be quoted without permission.

I. INTRODUCTION

In the history of human migration, there are only a handful of episodes frequently characterized as famous or extraordinary. Probably the most famous Western episode is the "Great Atlantic Migration," the movement of about 55 million people from Europe to the Americas and Australia between 1850 and 1914. The most famous episode of migration within the USA is the "Great Black Migration" of approximately 6.5 descendants of African slaves who left the American South and headed to cities in the North.[1] Receiving much less attention in the West, though, is an episode that began in the 1980s and which dwarfs all other episodes: the "Great Chinese Internal Migration." According to data from the China Census, approximately 33 million Chinese moved within or across provinces, with roughly one-third comprising flows of people across provincial lines. During 1995-2000 over 121 million persons moved, with approximately three-fourths of these moving within the same province. During 2000-05 nearly 195 million persons moved, approximately one-third between provinces. These surges primarily include persons moving without permission (the "floating population") from rural-to-urban areas and they have been focused on Eastern coastal cities.

Thanks largely to the 1990, 2000, and 2005 Chinese censuses and the intensification of Western style market reforms, researchers can now study the Great Chinese Internal Migration using Western models. A small, mostly empirical, literature on the determinants of internal migration in China has emerged. Its focus has been to examine the extent to which migration flows are driven by regional differences in labor markets and has identified two broad factors[2]:

---

[1] More recent famous episodes include: (i) the migration of Eastern Europeans to Western Europe following the expansion of the EU; (ii) the reversal of Ireland from being a net sender of immigrants for many years to, very recently, a large net receiver; (iii) the migration of several million Iraqi refugees to Syria and other Middle Eastern countries; (iv) the huge movements of persons born and raised in the former East Germany to the Western part of what is now united Germany; and (v) the estimated 10-11 million undocumented immigrants in the U.S., many from Mexico and Central America.

[2] The literature can be conveniently divided into studies utilizing micro-data obtained from special household surveys (see, for example, Liang (2001), Liang and White (1996,1997), Zhao (1997,1999a, 1999b, 2002, 2003) and a few studies utilizing province-level (see, for example, Fan (2005), Lin, Wang and Zhao (2004), Poncet (2006), Bao, Hou and Shi (2006)), and Bao, Bodvarsson, Hou and Shi (2008a, 2008b,

(1) Growth in regional income inequality due to the comprehensive economic reforms, a boom in China's export markets, and a surge in foreign and domestic investments; and (2) A dramatic decline in migration costs due to substantial improvements in the country's transportation infrastructure, deregulation of migration, and rapid growth in migrant communities. However, there is another major demographic change whose relationship to migration has not been fully recognized or properly analyzed – changes in the country's age distribution. Post-reform China is aging because of global prosperity, post-reform structural changes to China's economy, improved health care, the one child policy, and other factors. The question we ask: What is the relationship between changes in China's age distribution and post-reform migration surge; How have changes in the age distribution influenced internal migration patterns and, in turn, is the age distribution endogenous to migration?

To motivate our inquiry, consider Tables 1 and 2, which show snapshots since 1982 of two distinct measures of each province's age distribution – the share of population aged 15-29 and the age dependency ratio (ADR).[3] The share of population aged 15-29 is particularly relevant to the analysis of migration because that is the group considered most likely to relocate. Table 1 reveals that this particular group's share of the population rose very slightly from 1982 to 1990, but has declined since. In 1990, the youth share of the population was just over 30 percent, whereas it now ahs fallen to just over 21 percent. The likely primary cause of this post-1990 decline is the

---

forthcoming). We should also point out that in 2002, an entire issue of the journal *Urban Studies* was devoted to empirical papers on China's growing migration and urbanization. We particularly wish to highlight the studies of Chen and Coulson (2002) on the determinants of urban migration, Liang, Chen and Gu (2002) on the effects of rural industrialization on internal migration, Li and Zahniser (2002) on the determinants of temporary rural-to-urban immigration, Goodkind and West's (2002) study on the floating population.

[3] The ADR measures the relative size of the population that is not working, hence dependent upon the workforce for financial or in-kind support. The ratio is computed using the following formula:

$$ADR = \frac{\text{number of persons aged 0 to14} + \text{number of persons aged 65 or higher}}{\text{number of persons aged 15 - 64}}$$

one-child policy. However, accelerating prosperity in China has increased the opportunity costs of having children, which in the absence of the one child policy would have likely provided considerable disincentive for couples to have large families. An important question is: All other things equal, has the decline in the share of the population most likely to migrate resulted in lower migration rates?

Table 2 illustrates that, contrary to what many observers, particularly in the West, might expect, the age dependency ratio in China has fallen. In 1982, for every 100 persons of working age (15-64), there were on average of approximately 62 persons too young or too old to participate in the workforce. By 1990, the latter number had fallen to just under 50, by 2000 it was approximately 42 and it is now under 40. It is interesting that even though China's population has gotten older, on average the fraction of population not of working age has fallen. The decline in ADR is likely due to the long term effects of the one-child policy, which has reduced the fraction of the population consisting of children, teenagers, and very young adults. While improved health care and long term prosperity has increased the fraction of the population that is elderly, the rate of decline in the share of the young has in absolute value been greater than has the rate of increase in the old. Over the same period, the share of the working age population has risen.[4] Another important question then is: How has the long term decline in ADR influenced internal migration patterns? For example, has a decline in the fraction of the dependent population contributed *ceteris paribus* to an increase in the scale of migration?

How can previous literature linking age and migration, both in the West and China, help in answering the questions above? We consider three categories of previous work:

(a) *Western economic theory*. The theoretical relationship between age and migration has received very little attention in Western migration literature. A few scholars have made some

[4] Note that the average share of provincial population aged 15-64 was 66.79% in 1990, 70.39% in 2000 and 71.91% in 2005.

valuable points, though. Becker (1964) argued that the propensity to migrate will tend to decrease with age. The reason is that the expected net present value of the 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 home region's age distribution will be higher (lower). This has been the prevailing view and in most empirical studies where age is included as an explanatory variable, the strategy has been to test for a negative relationship between the rate of migration and a migrant's age.

Other explanations have been given for the effects of age, though. David (1974) suggested that seniority rights (which provide protection from the risk of layoffs) and nontransferable pension benefits will be lost following a move. Older workers will have a larger proportion of their wealth tied up in these specific assets and will have relatively more to lose following a move, hence migration rates for older persons will be lower, all other things equal. Schwartz (1976) makes a strong argument for the importance of psychic costs of migration. He characterizes these costs as the "agony of severing [] relations" with family members and friends. Schwartz argues that as persons get older, they will invest more in relationships with family members and friends and were they to relocate, the emotional costs of severing those relationships will be higher. The "agony" Schwartz describes can be assuaged by return visits to the origin and older persons will have a greater demand for return visits.

Lundborg (1991) developed Schwartz's point further by suggesting that the demand for return visits will depend on 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. On the one hand, older migrants value return visits more, but on the other, as time passes after the move the migrant invests in new social relations and the demand for return visits will fall. Furthermore, the larger is the migrant network the less homesick the migrant will feel and his/her demand for return visits will be lower. Schwartz provides two additional testable implications: (1) The deterrent effect of distance on migration is higher for older than for younger persons; and (2) There will be a U-shaped relationship between the sensitivity of migration to the size of the migrant community and migrant age;

(b) *Empirical Western literature on age and migration*. There have been numerous tests of the effect of age on migration, mostly of the Becker (1964) hypothesis, done in the West since the 1970s. In the interest of space, we do not provide an exhaustive survey here. Generally, the results from these studies have been quite mixed.[5] In some studies, it is found that younger migrants have a higher propensity to migrate, whereas other studies find the opposite. Some studies find no statistically significant relationship, while quite a few other studies simply omit age as a regressor. It is difficult to tie together the diverse empirical findings with respect to age across these studies because: (i) there is considerable diversity in empirical specifications and the types of data sets used; and (ii) te lack a unifying theory that is capable of accounting for the diversity of results. Our assessment is that Western empirical literature on age and migration would benefit from a unified theory and a meta-analytic study;

(c) *Empirical work on age and migration in China*. A majority of researchers doing empirical work on the determinants of migration in China have included age as an explanatory variable. The results from these studies have also been mixed.[6] An important result that emerges from a

---

[5] Bowles (1970) studied out-migration rates of black and white workers from the American south and found that for both groups, migration propensities were higher for younger workers, especially black workers. Using 1960 U.S. Census data, Schwarz (1976) found that migration rates were higher for younger persons and were the highest for well-educated young workers. Navratil and Doyle (1977) used 1970 U.S. Census data to estimate in-migration rates by race-sex cohorts (an aggregate flow model), as well as the likelihood of an individual migrating (a logit model using microdata), during 1965-70. For the aggregate flow model, they and found a positive relationship between age and migration rates for white males, black males and black females, but no relationship for white females; For the logit model the likelihood of migrating was higher for persons in all sex-race cohorts. Schlottmann and Herzog (1981) found strong evidence of an inverse relationship between the likelihood of migration and age using U.S. Census data for 1965-70, a result echoed by Goss and Paul (1986), who used Panel Study of Income Dynamics (PSID) micro-data for 1974-75. Lundborg (1991) found no evidence of a relationship between migration rates and age for migration between Scandinavian countries. Finally, Clark, Hatton and Williamson (2007) found no relationship between U.S. immigration rates during 1971-98 and the share of source country population aged 15-29.

[6] In a series of studies utilizing small, household surveys in specific areas, Zhao found a negative relationship between age and the propensity to migrate (Zhao (1999a, 1999b, 2003), as well as a positive relationship (Zhao (1997)). A negative effect of age on the propensity to migrate was also found by Zhu

majority of these studies is an inverse U-shaped relationship between age and the propensity to migrate.[7] Specifically, the propensity to migrate rises up to approximately the 25-30 age range, then falls thereafter. This type of result has generally not been found in the Western literature. Zhu (2002) has suggested that one reason for observing a negative age coefficient (beginning around the late twenties) in the Chinese case is that unskilled manual workers and older workers are disadvantaged with respect to their physical strength.

Other researchers have adopted the views of Lundborg (1991) and Schwartz (1976) that older migrants face higher psychic costs of migration. Zhao (1997) has suggested that the positive relationship between age and migration propensity that occurs through the mid-to-late twenties could be due to Hukou restrictions on migration, which may be especially constraining for very young persons living in rural areas. She points out that these persons may remain in rural areas for a while (or first enlist in the military) to gain favor from local officials in order to be considered for relocation when urban recruitment opportunities arise. Therefore, the apparent quadratic relationship between age and migration suggests that the Chinese case is more complex than the Western case due to institutional factors.[8]

In this paper, we seek to make a number of important contributions to understanding the complex relationship between aging and migration in China. First, we present a theoretical model of the migration decision, applicable to both the Western and Chinese cases, that is capable of

---

(2002), Shi et al (2007) and Wu (2008). We should also mention that Zhao (2002) found a higher tendency for older migrants to return home.

[7] See Shi and Bao (2006), Liang, Chen and Gu (2002), Zhao (1999a), Liang and White (1996, 1997), Li and Zahniser (2002), Ma and Liaw (1997) and Hare (1999).

[8] We can suggest a number of more casual explanations, applicable to both the Western and Chinese cases, for why migration propensities fall with age. One is that younger persons are on average healthier. Since the act of migration, particularly from a rural to an urban area, involves a relatively substantial investment of resources and there is a greater financial risk from getting sick and requiring hospitalization in the destination, particularly if one is part of the floating population (unauthorized migrants do not have access to free local medical care), older persons may find it much more costly to migrate. Another explanation is that younger persons are less risk averse, more adventuresome, and more entrepreneurial. All these explanations are consistent with and complement the general hypothesis from the basic human capital model that migration rates and age will be inversely related. We have not seen these explanations incorporated into formal theoretical or empirical models of migration, though.

generating the diversity of predictions compatible with what has been found in the empirical literature. Our model brings together the Becker (1964) explanation of the effects of aging with other explanations such as psychic costs and loss of firm-specific human capital. We then test numerous implications of the model on a panel data set encompassing three periods of interprovincial migration in China – 1985-90, 1995-2000, and 2000-05. Ours is the first study to examine the relationship between age and migration in China using a panel approach and aggregate data. Utilizing two measures of age distribution in the origin province – the share of persons aged 15-29 and the age dependency ratio – we find that age distribution is an important determinant of the scale of migration and can affect migration often in conflicting ways.

II. AGE AND THE DECISION TO MIGRATE: THEORY

The theoretical model below incorporates elements from the models of Schwartz (1976), Naskoteen and Zimmer (1990) and Lundborg (1991). While the model describes the behavior of an individual prospective migrant, it has immediate implications for the study of aggregate migration flows. For simplicity, we assume just one potential destination. The decision to migrate is influenced by three broad factors: (1) Differences in age-earnings profiles the migrant faces in the origin and destination; (2) The costs of maintaining investments in social relations at the origin; and (3) Direct migration costs, including the costs of obtaining Hukou in the destination.

Suppose a risk-neutral worker of age a, who plans to retire at time T, is contemplating a move from province i to province j. The decision to relocate is facilitated by a calculation of the expected net present value (NPV) of the benefits of relocation:

(1) $$E[NPV(t)] = \int_a^T [E(Y^j) - E(Y^i) - C^{ij}(t)]e^{-rt}dt \text{ ,}$$

where: Y^i = earnings per period available in the origin province;
Y^j = earnings per period available in the destination province;
C^{ij} = costs of migrating from provinces i to j;
r = discount rate.

We contend that it is not just spatial differences in expected levels of pay that matter to the prospective migrant, but expected spatial differences in age-earnings profiles. A more precise characterization of the migration problem is to say that the destination offers the migrant an age-earnings profile different from the profile available in the origin because: (a) earnings available with no labor market experience are different from what would be earned in the origin; (b) the returns to general human capital are different; (c) the returns to specific human capital are different; and (d) specific assets are lost following a move. If economic conditions in the destination are stronger, then the vertical intercept of the prospective migrant's age-earnings profile in the destination is likely to be higher. In other words, even if the migrant has no labor market experience, then at a given education level he/she can expect higher compensation in the destination. For the same reason, returns to general and specific human capital are likely to be higher, meaning that the age-earnings profile will be steeper than in the origin.

The loss of specific assets could be a significant reason for spatial differences in age-earnings profiles. A migrant may switch occupations following the move. For example, consider the case of a rural dweller in China whose human capital investments are primarily in agriculture. This person contemplates a move to a large city to work in manufacturing or construction. While he/she may be able to transfer general skills, specific human capital investments will be lost.

Given the above considerations, the expected stream of earnings in each location during a period is assumed to depend upon five components: (i) the wage that would be paid if the worker had zero labor market experience (we call this the "baseline wage," W_B); (ii) the amounts of general and specific human capital accumulated from prior periods; (iii) the price received for supplying a unit of general human capital (x); (iv) the price received for supplying a unit of specific human capital (s); and (v) the probability of securing employment ($1-\pi$, where π is the risk of unemployment).

For simplicity, we will assume that at any working age the worker acquires one unit of general human capital and one unit of specific human capital during each period in either location.[9] General human capital acquired in one location is perfectly transferrable to another location, while specific human capital cannot be transferred. That prompts a question: What exactly is specific human capital in this context – is it *firm*-specific (acquired while working for a particular employer) or *job/occupation* – specific (acquired while employed in a particular occupation or job assignment)? By definition, migration means a switch in employers, hence the abandonment of firm-specific human capital in one location and the initiation of investment in firm-specific human capital in another location. The migrant might perform the same job assignment or be in the same occupation in the destination, or might switch job assignments or occupations. While there may be some value to making a qualitative distinction between the skill set acquired within a firm versus the skill set acquired in an occupation or job assignment, we will not make a distinction between the two in this model. Instead, we will view specific human capital as firm-specific.

The prospective migrant is assumed to have graduated from school at age a_g and up to this point has accumulated $(a – a_g)$ units of general human capital. Furthermore, he/she is assumed to have taken the current job (in the origin) at age a_k $(a_g < a_k < a)$ and to have accumulated $(a – a_k)$ units of specific human capital. While the baseline wage, the reward for supplying general human capital, and the reward for specific human capital are likely to differ within and across locations, we will assume for simplicity that they are constant across the worker's lifetime. It follows then that at age a, the expected earnings in each location are the following:

[9] This is a departure from the traditional view of on-the-job training, which is that human capital investments taper off with age. As long as the rate of human capital accumulation is the same in either location, the predictions of the model do not depend upon the rate being constant or non-linear. It is possible that human capital could be acquired at a faster rate in one of the locations, e.g. at a faster rate in the destination if technology there is more advanced. We will not pursue that idea for present purposes, although that would be a worthwhile extension of the model.

$$E(Y^i) = \int_a^T [(1-\pi^i)(W_B^i + x^i + s^i)e^{-rt}dt] + x^i(a-a_g) + s^i(a-a_k)$$

(2)

$$E(Y^j) = \int_a^T [(1-\pi^j)(W_B^j + x^j + s^j)e^{-rt}dt] + x^j(a-a_g)$$

The costs of migration are assumed to depend upon the frequency of needed travel to the origin, the costs of transportation (which depend directly on distance) and the costs of securing Hukou in the destination. Following Lundborg (1991, pp. 395-6), we assume that the frequency of needed travel to the origin (TRIPS) depends upon the amount of time spent at the destination, the migrant's age at migration, and the relative size of the migrant community in the destination. Following Lundborg (1991), the function measuring desired return trips may be written as:

(3) TRIPS = g(t-a, a, MS_{ij}),

where: t-a = number of years at the destination
MS_{ij} = the stock of prior migrants from i residing in j.

With subscripts as derivatives, it is assumed that $g_{t-a} < 0$ because, as suggested by Schwartz (1976) and Lundborg (1991), as the migrant becomes more familiar and comfortable with the destination, investments in new social relationships occur at the destination and return trips to the origin have increasingly less value to him. In contrast, $g_a > 0$ because the older one is at the time of migration the more that will have been invested in social relationships at the origin, hence the greater will be the desire to make return visits.[10] We also adopt Lundborg's assumption that $g_{ms} < 0$ because a larger community of countrymen at the place of destination will lower psychic costs and the frequency of needed return visits.[11] The costs of return trips to the origin over the

[10] Compared to those who migrate when they are relatively young, older migrants are more likely to have developed longer term and deeper friendships at the origin, cultivated relationships with extended family members more extensively, and to leave children behind at the origin. To use Schwartz's (1976) terminology, the "agony" of severing those sorts of ties will be more intense for older migrants and they will feel a greater yearning to return home more frequently.

[11] Lundborg also points out that for migrants, particularly younger ones who are active in the market for marriage partners, a larger migrant network in the destination will make it more likely that a marriage partner with ties to the origin can be found at the destination. This will tend to reduce the number of trips that need to be made to the origin.

migrant's lifetime in the destination will equal the number of trips made times the direct cost of a return trip to the origin, which discounted to the present are

$$(4)\ \int_a^T [g(t-a, a, MS_{ij})C_{ij}(D_{ij})]e^{-rt}dt,$$

where C_{ij} = the round-trip direct cost of travelling between origin and destination;
D_{ij} = distance between origin and destination

It is assumed that $\frac{\partial C}{\partial D} > 0.$

Equation (4) quantifies the psychic costs of migration. There will also be the costs of the initial move (moving oneself and one's possessions, switching dwellings, etc.) and the costs of securing local registration in the destination. The costs of the initial move will depend upon geographic distance. It is assumed that older migrants will generally face higher costs of securing Hukou because they will have a lower likelihood of being included in those groups that are considered strong candidates for local Hukou, i.e. students admitted to university, military conscripts, and marriage partners. These factors are captured by the function below, measuring the direct costs of the initial move M:

(5) $M = m(D_{ij}) + H_{ij}(a)$,

where: H_{ij} = the costs of a resident of province i switching local registration to province j.

It is assumed that $\frac{\partial H}{\partial a} > 0.$

Combining equations (2) through (5), the expected net present value of migration is

$$(6)\ E[NPV(t)] = \int_a^T [(1-\pi^j)(W_B^j + x^j + s^j) - (1-\pi^i)(W_B^i + x^i + s^i)]e^{-rt}dt - \int_a^T [g(t-a, a, MS_{ij})C_{ij}(D_{ij})]e^{-rt} - m(\mathrm{D_{ij}}) - H_{ij}(a)) + (1-\pi^j)(x^j - x^i)(a-a_g) - (1-\pi^i)(s^i)(a-a_k).$$

Equation (6) illustrates that the net benefits to migration are influenced by five factors: (i) destination/source differences in baseline wages, returns to general human capital, returns to specific human capital, and unemployment risk; (ii) the expected psychic costs of migration, which are in turn are influenced by the needed frequency of return trips to the origin; (iii) the direct costs of moving and of securing local registration; (iv) the expected extra value the migrant can extract from transferring his/her accumulated general human capital to the destination; and (v) the expected loss in specific human capital acquired in the origin that is suffered as a result of relocation.

An additional cost facing the migrant is the expense of obtaining Hukou, which we assume varies inversely with age. Persons admitted to universities in another province are granted local registration, as are military conscripts. Another way of obtaining Hukou is to marry someone registered in the destination province. University students, military conscripts, and candidates for marriage are usually younger than the mean. In addition, because registration is a requirement for having access to social medical insurance in the destination, the costs of not having Hukou will be higher for older persons because they are at greater risk of needing medical care.

Since there is just one destination, the migration decision is a yes/no decision; The worker will migrate if equation (6) is positive. Our interest is in how age influences the migration decision. As a first step, we integrate equation (6),

$$(7)\ E[NPV(t)] = [(1-\pi^j)(W_B^j + x^j + s^j) - (1-\pi^i)(W_B^i + x^i + s^i)][(\frac{1}{r})(\frac{1}{a}e^{-ra} - \frac{1}{T}e^{-rt})] - \\ [g(t-a, a, MS_{ij})C_{ij}(D_{ij})][(\frac{1}{r})(\frac{1}{a}e^{-ra} - \frac{1}{T}e^{-rt})] + (1-\pi^j)(x^j - x^i)(a-a_g) - (1-\pi^i)(s^i)(a-a_k) - m(D_{ij}) - H_{ij}(a),$$

and then differentiate equation (7) with respect to age:

$$(8)\ \frac{\partial E[NPV(t)]}{\partial a} = -[\frac{(1-\pi^j)(W_B^j + x^j + s^j) - (1-\pi^i)(W_B^i + x^i + s^i)}{r}][e^{-ra}(\frac{1}{a^2} + \frac{r}{a})] + [\frac{g(t-a, a, MS_{ij})(C_{ij}(D_{ij}))}{r}][(\frac{1}{a^2} + \frac{r}{a})e^{-ra}] \\ -[\frac{\partial g}{\partial a}C_{ij}(D_{ij})](\frac{1}{a}e^{-ra} - \frac{1}{T}e^{-rt}) + (1-\pi^j)(x^j - x^i) - (1-\pi^i)(s^i) - \frac{\partial H_{ij}}{\partial a}.$$

Expression (8) has a number of important implications:

A. **The marginal effect of age on the net benefits of migration depends upon six factors, which are reflected in the six terms on the RHS of expression (8):**

1. The first term (negative) embodies Becker's explanation that older (younger) migrants will have a lower (higher) incentive to migrate because the stream of potential earnings gains in the destination is shorter (longer). Overlooked in the literature, though, is that the size of this effect depends upon spatial differences in baseline wages, returns to each type of human capital, and unemployment risks. Furthermore, the derivative of this term with respect to age is positive, indicating that the negative effect of a shorter term horizon in the destination on earnings increases at a decreasing rate with age;

2. The second term (positive) captures the idea that because the time horizon in the destination is shorter, an older migrant will make fewer trips to the origin and return migration costs will thus be smaller. Note that the savings in return migration costs will be larger, the greater is distance. The derivative of this term with respect to age is negative, indicating that savings in migration costs rise with age, but only at a diminishing rate;

3. The third term (negative) measures the increase in return migration costs facing older migrants because they will have stronger ties to family and friends in the origin. This effect is more negative as distance gets larger;

4. The fourth term (positive) has been completely overlooked in the literature. Older migrants will bring more general human capital to the destination. If the returns to general human capital are greater in the destination, then older migrants will, all other things equal, benefit more from transferring this human capital to the destination than will younger migrants. The extra benefit of transfer will be larger the larger is the premium to general human capital in the destination and the lower is unemployment risk;

5. The fifth term (negative) represents the loss to the migrant from abandoning investments in specific human capital made in the origin. The loss will be greater the lower is unemployment risk in the destination and the higher is the return to specific human capital in the origin;

6. The sixth term (negative) measures the extra cost to an older migrant of securing local registration in the destination;

B. The marginal effect of age on the incentive to migrate can be positive, negative, or zero, depending upon the sizes of the six effects discussed above. Most of the literature discussing age and migration presumes that older migrants always have a lower incentive to migrate. Our analysis above suggests that they could actually have a stronger incentive to relocate. Specifically, the sum of the second and fourth terms on the RHS of equation (8) could dominate the absolute value of the sum of the other four (negative) effects. This could occur if the frequency of return trips needed is relatively high (thus the cost savings from fewer return trips for older migrants is relatively high) and/or the premium paid in the destination for general human capital is relatively high. Therefore, the sign on the marginal effect of age is ultimately an empirical issue and the reason for the mixed results on the age coefficient in empirical studies may be due to the ambiguous theoretical relationship between age and the expected net returns to migration;

3. There will be an optimal age of migration. Our analysis implies that there will be some age that balances the expected marginal gains from waiting a year to migrate with the expected losses. Equation (8) provides the specific conditions for determining the optimal age of migration. The optimal age of migration is where the expected marginal benefit of being an older migrant (the sum of the second and fourth terms on the RHS of equation (8)) equals the expected marginal loss (the absolute value of the sum of the first, third, fifth and sixth terms):

$$(9)\ [\frac{(1-\pi^j)(W_B^j+x^j+s^j)-(1-\pi^i)(W_B^i+x^i+s^i)}{r}][e^{-ra}(\frac{1}{a^2}+\frac{r}{a})]+[\frac{\partial \mathrm{g}}{\partial \mathrm{a}}\mathrm{C}_{\mathrm{ij}}(\mathrm{D}_{\mathrm{ij}})](\frac{1}{\mathrm{a}}\mathrm{e}^{\text{-ra}} - \frac{1}{\mathrm{T}}\mathrm{e}^{\text{-rt}})+(1-\pi^i)(s^i)+\frac{\partial H_{ij}}{\partial a}$$

$$=[\frac{g(t-a,a,MS_{ij})(C_{ij}(D_{ij}))}{r}][(\frac{1}{a^2}+\frac{r}{a})e^{-ra}]+(1-\pi^j)(x^j-x^i).$$

The left-side term in equation (9) is the expected marginal gain from being an older migrant, whereas the right-side term is the expected marginal loss. While functional forms for g( ), C( ), and H( ) are required for an optimal numerical value for age, some general implications for optimal age of migration can be inferred from equation (9):[12]

(i) *The optimal age of migration will tend to be higher the greater is the premium paid in the destination for general human capital (the higher is x^j) and the greater is the frequency of needed return trips to the origin (the higher is the value of g for a given value of age), all other things equal*. If employers in the destination province provide a much higher premium for general human capital, but only a modest premium for specific human capital (and only a modestly higher baseline wage), older migrants will benefit substantially from transferring their general human capital to the destination. If, at the same, time there is a strong need for return trips to the origin, then the savings in return migration costs from being an older migrant will be relatively large. Under these conditions, people in the origin province may have a greater tendency to put off migration till an older age;

(ii) There will be a greater tendency for migrants to be younger when there are relatively large spatial differences in baseline wages and returns to specific human capital, the desire to make

---

[12] These implications require that $\frac{\partial^2 E(NPV(t))}{\partial a^2} < 0$, which will be satisfied if the derivative of the LHS of equation (9) with respect to age is smaller in absolute value than the derivative of the RHS of the equation. That condition can easily be satisfied with parameter restrictions.

return trips home is sensitive to age ($\partial g/\partial a$ is relatively large), the marginal cost of securing Hukou in the destination rises rapidly with age ($\partial H/\partial a$ is relatively large), and distance is large. Our model provides a theoretical explanation for why migrants moving long distances from rural to urban provinces and who switch occupations following a move tend to be younger. The long distance implies high costs of return migration, hence only those persons with relatively small investments in ties to family and friends back home will move. It is generally more costly for rural migrants to obtain urban Hukou and those costs are likely to rise rapidly with age. These conditions will tend to discourage older persons from moving.

There is another implication of our model that relates to optimal age of migration. Because optimal age is influenced by the same factors that influence the decision to migrate, the migration decision and age at migration are jointly determined. The traditional view of age in the context of migration is that age influences migration. Our model suggests two-way causality between age at migration and migration rates. For example, suppose that returns to specific human capital at the origin rise, resulting in both less out-migration and the average age of migrants to fall. Consequently, the age distribution shifts in the origin, specifically mean age of residents rises and there is likely to be a change in the variance of age. Therefore, out-migration rates and age distribution in the origin are both endogenous.

While it is beyond the scope of the theoretical model to pursue a simultaneous equations model from which closed form solutions for optimal age and the likelihood of migration are derived, this insight from the model has an immediate and important empirical implication: In the estimation of an equation for the migration rate, it is important to account for the effects of age on migration, what could be called a “feedback effect.” Accounting for this feedback effect is especially important when working with aggregate data, for there is the strong possibility that out-migration from a province will alter its age distribution.

D. Age will interact with other determinants of migration. The traditional determinants of migration rates are distance, migrant networks, and spatial differences in earnings and unemployment risk. Our model implies age interactions with these and other determinants:

Distance. *The marginal effect of distance on the expected net benefits of migration is negative, but can be less or more negative for older migrants*:

$$(10)\ \frac{\partial E[NPV(t)]}{\partial D_{ij}} = -g(t-a,a,MS_{ij})\frac{\partial C_{ij}}{\partial D_{ij}}(\frac{1}{r})(\frac{1}{a}e^{-ra}-\frac{1}{T}e^{-rt})-\frac{\partial M}{\partial D_{ij}}<0$$

$$\frac{\partial^2 E(NPV(t))}{\partial D_{ij}\partial a} = -[\frac{\partial g}{\partial a}\frac{\partial C_{ij}}{\partial D_{ij}}(\frac{1}{r})(\frac{1}{a}e^{-ra}-\frac{1}{T}e^{-rt})-(\frac{1}{a^2}e^{-ra}+\frac{r}{a}e^{-ra})g(t-a,a,MS_{ij})\frac{\partial C_{ij}}{\partial D_{ij}}(\frac{1}{r})] \geq or < 0.$$

Greater distance taxes the migrant for two reasons. First, greater distance means higher out-of-pocket costs of transportation, switching residences, and all the other direct costs of the initial relocation (the second term on the RHS of equation (10)). Second, greater distance increases the costs of return migration for a given number of return trips (the first term on the RHS of equation (10)). Age doesn't impact out-of-pocket costs created by distance, but it does impact the effects of distance on the costs of return migration. Because older migrants place a higher priority on return visits ($\partial g/\partial a > 0$), a given increase in distance will impose higher costs on an older migrant than a younger migrant. This "distance stings more" effect on an older migrant is reflected in the first term in brackets in the cross-partial derivative in (10) above. On the other hand, an older migrant has a shorter time horizon in the destination, so will be making fewer return trips than a younger migrant. This effect is reflected in the second term in brackets in the cross-partial derivative. The cross-partial derivative will net be negative if, for example, the frequency of return trips is very sensitive to age (∂g/∂a is very high). The ambiguity of the sign of the cross-partial derivative in (10) stems from the fact that the expression in brackets includes the difference between two positive terms and it is unclear which expression is larger. This result conflicts with the findings of Schwartz (1976) and Lundborg (1991), who show that the negative marginal effect of distance unambiguously strengthens with age;

Migrant stock. *The positive marginal benefits of a larger migrant stock in the destination diminish with migrant age:*

$$(11)\ \frac{\partial E[NPV(t)]}{\partial MS_{ij}} = -(\frac{\partial g}{\partial MS_{ij}})(C_{ij} D_{ij})(\frac{1}{ra} e^{-ra}) > 0$$

$$\frac{\partial^2 E[NPV(t)]}{\partial MS_{ij} \partial a} = (\frac{\partial g}{\partial MS_{ij}})(C_{ij} D_{ij})(\frac{1}{a^2} + \frac{r}{a}) e^{-ra} < 0$$

Migrants benefit from a larger migrant network in the destination because a larger network assuages the “agony” of being separated from family and friends back in the origin. The migrant’s need for return trips to the origin is lowered and he/she saves on return migration costs. Older migrants will save less, though, because their time horizon in the destination is shorter. Therefore, we predict a negative interaction between migrant stock and age;

Other interaction effects include:

Destination unemployment risk. *The risk of higher unemployment in the destination deters migration, but the deterrent effect can be either weaker or stronger for older migrants:*

$$(12)\ \frac{\partial E[NPV(t)]}{\pi^j} = [-(W_B^j + x^j + s^j)][(\frac{1}{r})(\frac{1}{a} e^{-ra} - \frac{1}{T} e^{-rt})] - \pi^j (x^j - x^i)(a - a_g) < 0$$

$$\frac{\partial^2 E[NPV(t)]}{\partial \pi^j \partial a} = [(W_B^j + x^j + s^j)][(\frac{1}{a^2} e^{-ra} + \frac{r}{a} e^{-ra})] - \pi^j (x^j - x^i) \geq or < 0.$$

Origin unemployment risk. *Higher risk of unemployment in the origin encourages migration, but the encouragement factor can be stronger or weaker for older migrants:*

$$(13)\ \frac{\partial E[NPV(t)]}{\pi^i} = [(W_B^i + x^i + s^i)][(\frac{1}{r})(\frac{1}{a}e^{-ra} - \frac{1}{T}e^{-rt})] + s^i(a - a_k) > 0$$

$$\frac{\partial^2 E[NPV(t)]}{\partial \pi^i \partial a} = -[(W_B^i + x^i + s^i)][(\frac{1}{a^2}e^{-ra} + \frac{r}{a}e^{-ra})] + s^i \geq or < 0.$$

Returns to general human capital. *Higher returns to general human capital in the destination (origin) encourages (discourages) migration, but the strength of encouragement (discouragement) can be strengthened or weakened for older migrants:*

$$(14)\ \frac{\partial E[NPV(t)]}{\partial x^j} = [(1-\pi^j)][(\frac{1}{r})(\frac{1}{a}e^{-ra} - \frac{1}{T}e^{-rt})] + (1-\pi^j)(a - a_g) > 0$$

$$\frac{\partial^2 E[NPV(t)]}{\partial x^j \partial a} = -[(1-\pi^j)][(\frac{1}{r})(\frac{1}{a^2}e^{-ra} + \frac{r}{a}e^{-ra})] + (1-\pi^j) \geq or < 0.$$

$$(15)\ \frac{\partial E[NPV(t)]}{\partial x^i} = -[(1-\pi^i)][(\frac{1}{r})(\frac{1}{a}e^{-ra} - \frac{1}{T}e^{-rt})] - (1-\pi^j)(a - a_g) < 0$$

$$\frac{\partial^2 E[NPV(t)]}{\partial x^j \partial a} = [(1-\pi^j)][(\frac{1}{r})(\frac{1}{a^2}e^{-ra} + \frac{r}{a}e^{-ra})] - (1-\pi^j) \geq or < 0.$$

Returns to specific human capital in the origin. *An increase in the return to specific human capital in the origin will reduce the incentive to emigrate, but the strength of this reduction can rise or fall with age:*

$$(16)\ \frac{\partial E[NPV(t)]}{\partial s^i} = -(1-\pi^i)[(\frac{1}{r})(\frac{1}{a}e^{-ra} - \frac{1}{T}e^{-rt}) + (a - a_k)] < 0$$

$$\frac{\partial^2 E[NPV(t)]}{\partial s^i \partial a} = (1-\pi^i)[(\frac{1}{r})(\frac{1}{a^2}e^{-ra} + \frac{r}{a}e^{-ra}) - 1] \geq or < 0$$

To summarize, our theoretical model provides some important guidelines for our empirical specifications. First, a complete empirical specification should include empirical counterparts for the six components of earnings included in equation (7), age at migration, migrant stock, distance, general human capital in the origin and destination, specific human capital in the origin, barriers to securing local registration in the destination, unemployment risk in both locations, and time spent in the destination.[13] Second, there need to be interactions between the various determinants of migration and age. Third, because the decision to migrate and age at migration are jointly

[13] One should also include some measure of the discount rate, but most studies usually do not include any such measure in empirical specifications.

determined, the empirical strategy should take account of the possible "feedback effect" of migration rates on age distribution in the origin. This could be accomplished by a simultaneous equations estimation procedure. Fourth, because of the non-linear functional form of equation (7), a double-log equation for the migration rate would be most appropriate.

III. EMPIRICAL MODEL

The theoretical model above implies a double-log empirical specification where the dependent variable is the log of the migration rate ($\ln(M_{ij})$), defined as the number of persons moving from province i to province j as a percentage of all persons moving out of province i.[14] In using this specification for the case of China, we include explanatory variables from an assortment of studies, including Lin, Wang and Zhao (2004), Bao, Hou and Shi (2006), Poncet (2006), and Bao, Bodvarsson, Hou, and Zhao (forthcoming). Our empirical model extends previous research in several important ways. First, we use two alternative measures of the origin province's age distribution as an explanatory variable. Second, we include interactions between age distribution and some key explanatory variables. Third, ours is the first study we know of that accounts for the possibility that the origin province's end-of-period age distribution and out-migration are simultaneously determined. Finally, in contrast to most previous studies of China, ours is a panel study spanning three important periods of migration.[15]

Our empirical specification, whose structure is implied by the model in the previous section, is an extension of a panel regression equation due to Bao, Bodvarsson, Hou and Zhao (forthcoming). These authors in turn base their specification on earlier work by Greenwood (1969), Lin, Wang and Zhao (2004) and Poncet (2006). We take Bao, Bodvarsson, Hou and Zhao's equation and add measures of the origin province's age distribution, interactions between age distribution and other explanatory variables implied by the theory. We also add a control for

[14] This specification, widely used in the literature, is due originally to Greenwood (1969).
[15] The only other study to do this for China is Bao, Bodvarsson, Hou and Zhao (forthcoming).

migration flows between adjacent provinces, which has also been used by Poncet (2006). Our specification is described by this double-log equation for the interprovincial migration rate:

$$(17)\ \mathrm{Ln}(\mathrm{M}_{\mathrm{ijt}}) = \alpha_0 + \alpha_1 Ln(Y_{jit}) + \alpha_2 D_{ij} + \alpha_3 Ln(1-u_{it}) + \alpha_4 Ln(1-u_{jt}) + \alpha_5 Ln(\mathrm{Hukpercent}_{it}) + \alpha_6 Ln(\mathrm{Age}_{\mathrm{jt}}) + \sum_{x=1}^{q}\beta_x Ln(\mathrm{Z}_{xt}) + \sum_{t=1}^{T-1}[(\lambda_t \mathrm{Period}_t) + \sum_{w=1}^{s}\psi_w[Ln(\mathrm{Age}_{\mathrm{jt}})Ln(\mathrm{V}_{wt}))] + \sum_{p=1}^{Z-1}\theta_p \mathrm{Province}_p + \varepsilon_{ijt},$$

where:

$\mathrm{Y_{jit}}$ = the ratio of destination province to origin province income in period t;

$\mathrm{D_{ij}}$ = Geographic distance between provinces;

$\mathrm{u_{jt}}$ ($\mathrm{u_{it}}$) = the probability of being unemployed in the destination (origin) province in period t;

$\mathrm{Hukpercent_{it}}$ = the probability of securing local Hukou in the destination

$\mathrm{Age_{it}}$ = a measure of the origin province’s age distribution;

$\mathrm{Z_{xt}}$ = all other controls for perceived quality of provincial life, in period t;

$\mathrm{V_{wt}}$ = a subset of controls in Z

$\mathrm{Period_t}$ = time period during which migration occurred, where there are T periods;

$\mathrm{Province_p}$ = origin province fixed effect, where there are Z provinces;

$\varepsilon_{\mathrm{ijt}}$ = random error term;

and the α, β, λ, and θ parameters are coefficients to be estimated. We hypothesize that $\alpha_1 > 0$, $\alpha_2 < 0$, $\alpha_3 > 0$, $\alpha_4 < 0$, $\alpha_5 > 0$, and both α_6 and ψ_w can be > 0 or < 0.

Because of the potential for two-way casuality between migration and age distribution, we take two approaches to estimating equation (17). Our first approach is to estimate an OLS version where the age distribution variable is measured at the beginning of the period. Our reasoning is that migration patterns during the period will be influenced by predetermined characteristics of the age distribution. Thus, OLS is fine because there is no potential for two-way causality between age distribution and migration. However, the really interesting question is how migration and age distribution influence one another during the period? This is a question about two-way causality and it can only be answered effectively with simultaneous equation estimation.

Consequently, our second approach is use two-stage least squares (2SLS), where: (a) age distribution (measured after the beginning of the period of migration) is regressed on two instruments, as well as a number of the controls that are part of equation (17); and (b) the migration equation (less one control) is estimated with predicted values for age distribution obtained in (a) being substituted for actual values.

The two measures of age distribution used in this study are: (i) the share of the origin province's population aged 15-29[16]; and (ii) the origin province's age dependency ratio (ADR). The former has been used in a number of important migration studies, e.g. Clark Hatton and Williamson (2007), as an indicator of the sending country's age distribution. Recall that our theoretical model implies that migration rates will be influenced by interaction effects between age distribution and several key determinants of migration – distance, the relative size of the migrant network in the destination, the likelihoods of employment in each province, and relative income in the destination. Accordingly, for each panel and estimation strategy we will estimate equation (17) first without these five interaction terms and then we'll estimate the equation with the interactions included.

Following Bao, Bodvarsson, Hou and Zhao (forthcoming), we measure the probability of securing Hukou in the destination as the lagged relative frequency of registered households. It is assumed that prospective migrants know the historical relative frequencies of registered households in destination provinces and have adaptive expectations about barriers to entry. When the likelihood of securing Hukou rises, perceived benefits to migration will rise and that will lead to a higher migration rate.

The controls comprising the vector Z (the x's) included for each period are the following (hypothesized signs in parentheses):

(i) Log size of the migrant community residing in the destination province that previously migrated from the origin province, as a percent of the destination's

[16] Note that this measure has been used in a number of prominent migration studies, e.g. Clark, Hatton and Williamson's (2007) study of U.S. immigration.

population (> 0);

(ii) Log ratio of real FDI per capita in the destination province to real FDI per capita in the origin province (>0);

(iii) Log ratio of real domestic fixed asset investment (FAI) per capita in the destination province to real domestic FAI in the origin province (>0);

(iv) Log percentages of population enrolled in the origin province's universities (< 0) and the destination province's universities (> 0)

(v) Log ratio of the share of manufacturing employment in the destination to the share of manufacturing employment in the origin (> 0)

(vi) Log ratio of the urban share of the destination province's population to the urban share of the origin province's population (> 0)

(vii) Log ratio of the destination province's minority population share to the origin province's minority population share (> 0 or < 0);

(viii) Log ratio of mean yearly temperature in the capital city of the destination province to mean yearly temperature in the capital city of the origin province (> 0);

(ix) Dummy equaling one if the migration flow is between adjacent provinces (>0).

The instruments used in the age distribution equation are: (1) the number of doctors per 10,000 persons in the origin province ("doctor density"); and (2) the share of non-Han population in the origin province.[17] We conjecture that doctor density is an indicator of the level of public health spending in the province. When doctor density is higher, residents will ultimately be healthier, infant mortality will be lower, and death rates, particularly among the elderly, will be lower. This will ultimately result in a shift in the age distribution. The share of non-Han population is used as an instrument to account for the possibility that the age distribution may be influenced by ethnic characteristics of the origin population.

Other explanatory variables included in the age distribution equation include the urban population share in the origin province, temperature in the origin province, income in the origin province, and educational attainment in the origin province. Populations in more urban provinces could be healthier than those in rural provinces, health and longevity may depend upon climate,

[17] Note that the non-Han population share is the omitted control in the 2SLS version of equation (17).

and higher income, as well as more-educated persons, may have greater longevity. All these factors may, in different ways, influence the origin province's age distribution.

IV. DISCUSSION OF DATA

Our data are drawn from two major sources. For the 1985-90 and 1995-2000 periods, we expand the data set used by Lin, Wang and Zhao (2004) in their study of interprovincial migration.[18] Data for 2000-05 are taken from University of Michigan's *China Data Online* website (http://www.chinadataonline.org/). For all regressions we omitted observations for which the migration rate was zero. Like Lin, Wang and Zhao (2004) and other researchers, we exclude Tibet from our sample because of data gaps and we treat Chongqing as part of Sichuan. Our sample thus includes 29 provinces, each a prospective destination and point of origin.

A major drawback of the 1985-90 period is that information about the size of the community of migrants from the origin who reside in the destination is not available. The reason is that the 1990 semi-decennial census was the first to include questions about change in residence and measures of migrant stock during 1990 would require information about migration flows prior to that year. However, information about migrant networks is available when estimating migration rates for 1990 and beyond. Consequently, we produced two sets of estimates: (i) estimates for the full panel (three periods) with no control for past migration; and (ii) estimates for a smaller panel comprising the later two periods only (which does include a control for past migration). There are 2,385 usable observations in the full panel, of which 765 come from the first period and 790 from each of the later two periods. The smaller panel has 1,535 observations and includes 784 for 1995-2000 and 751 for 2000-05. The reason fewer observations for each period are used in the smaller panel is because observations for which past migration is zero were eliminated.

[18] Note that we replaced Lin, Wang and Zhao's (2004) calculations of the dependent variable with our own calculations. The reason is that there are some inaccuracies in the series used by Lin, Wang and Zhao, which they acknowledged in communications with us.

Tables 3 - 5 show summary statistics for all variables used in our regressions for each of the three migration periods. Starting from the top of each table, we describe the variable, the data source from which the variable is taken, and trends apparent in the data:

(i) Gross interprovincial migration rate. For the 1985-90, 1995-2000 and 2000-05 periods, respectively, migration rates are calculated from samples comprising 1% of the 1990 population census, 0.95% of the 2000 census,[19] and 1% of the 2005 census. In the 1990 (2000, 2005) census, respondents were asked to report on migration activities during 1985-90 (1995-2000, 2000-05, respectively). Consequently, migration rates during each decade were calculated for only the second half of each decade. The mean volume of emigration from a province surged from over 355,000 persons during 1985-90 to over 1,075,000 during 1995-2000 and over 2,200,000 during 2000-05.[20] Note that mean provincial population rose by 9.44% between 1990 and 2000 and by 5.86% between 2000 and 2005;

(ii) *The percentage of population aged 15-29 in the origin province*. These data were calculated from Census information on provincial age distributions. In our OLS estimation of equation (18), we used 1982 (1995, 2000) data on the age 15-29 population share for the 1985-90 (1995-2000, 2000-05) periods of migration; For the 2SLS estimation we used 1990 (2000, 2005) data on the age 15-29 population share for the 1985-90 (1995-2000, 2000-05) periods of migration. Note that

[19] As pointed out by Lin, Wang and Zhao (2004), there is a small difference between the 1990 and 2000 censuses with respect to how migration is defined. If a person is observed to change residence *and* to change their household registration (a situation officially called "Hukou migration"), then this movement is officially classified as "migration" in both censuses. If, however, the person is observed to change residence without changing registration ("non-Hukou migration"), then the movement is classified as "migration" only if the migrant has been away from the place of registration for a minimum period of time. In the 2000 census, this period is 6 months, but in the 1990 census it is one year. To account for this change in classification between the two periods, the migration numbers in both periods were standardized by discounting the 2000 numbers by a small amount, approximately 5%. For further details, see Lin, Wang and Zhao (2004, page 593).

[20] There are likely to be discrepancies in the calculations of these numbers between decades, for the reasons discussed in the preceding footnote.

Tables 3-5 show summary statistics for the mean values of 1982 and 1990 (1995 and 2000, 2000 and 2005). The tables illustrate that the national age distribution has shifted over time towards older age groups, as the number of persons aged 15-29 has fallen from approximately an average of 30% of provincial population during 1985-90 to below 24% during 2000-05;

(ii) *Age dependency ratio (ADR)*. In our OLS estimation of equation (17), we used 1982 (1995, 2000) data on ADR for the 1985-90 (1995-2000, 2000-05) periods of migration; For the 2SLS estimation we used 1990 (2000, 2005) data on ADR for the 1985-90 (1995-2000, 2000-05) periods of migration. Tables 3-5 show summary statistics for the mean values of 1982 and 1990 (1995 and 2000, 2000 and 2005).

(ii) *The historical relative frequency of persons with local Hukou*. This is the ratio of the registered population to total (registered + unregistered) population at year's end. For the 1985-90 (1995-2000, 2000-05) period, we use the mean annual proportion of persons with Hukou during 1980-84 (1990-94, 1995-99, respectively). We use the lagged proportion of persons with Hukou because there is very likely to be two-way causality between the migration rate and the contemporaneous proportion of registered persons in the destination. By using the lagged proportion of persons with Hukou we avoid potential problems with simultaneous equations bias;

(iii) Size of the community of migrants from the origin who reside in the destination. An ideal measure of the size of the destination's migrant community is the current stock of migrants from the origin as a percentage of current population in the destination. Unfortunately, unlike data sets in the USA and many European countries, this type of migrant stock measure is unavailable for China. Therefore, we used past relative flows. We measured the relative size of the destination's migrant network with relative migrant flows during the half-decade ending five years prior to the migration period. For the 1995-2000 (2000-05) period migrant network was calculated by the

ratio of total flows from origin to destination during 1985-90 (1990-95) to the destination's population in 2000 (2005). There are several reasons for this approach. First, it is presumed that the stock of previous migrants is proportional to the size of the previous flow of migrants. Second, by lagging past flows by 5 years, we hopefully reduce the risk of serial correlation. As Tables 4 and 5 show, the estimated provincial migrant stock averaged approximately 1.35 million persons for the 1995-2000 migration period and approximately 1.22 million persons for the 2000-05 period. The reduction in migrant stock could be due to return migration;

(iv) Real annual FDI and FAI per capita. For each period, we used mean annual real FDI (FAI) per capita during 1980-84 when regressing 1985-90 migration flows, 1990-94 mean annual real FDI (FAI) per capita when regressing 1995-2000 migration flows and 1995-99 mean real FDI (FAI) per capita when regressing 2000-05 migration flows. We lagged investment spending because it typically takes time for migration to respond to changes in spending on investment projects. Furthermore, since there is very likely to be two-way causality between investment and migration, by regressing migration rates on lagged investments we avoid potential problems with simultaneous equations bias. We adjusted the investment series for cost of living differences between the two decades, as well as across provinces within each decade, using national government measures of provincial CPI and calculating both series at 1985 price levels. For most of the provinces, FDI numbers were available for each year, but for some there were missing years. For several provinces, no investment data were available for 1980-84, so we used the earliest year available as a proxy for that period. Therefore, our coefficient estimates for the early period may be influenced by measurement error in some parts of the investment series. Note that the FDI series is in USA dollars, whereas the fixed asset investment series is in Chinese Yuan;

(v) The share of manufacturing employment. Manufacturing is classified in China as a "Secondary" industry in China and construction is one of its components. There is considerable variation across the country with respect to the dominance of manufacturing in the provincial labor market;

(vi) The share of the province's population that is non-Han. Because data on Han population shares for 1990 are not available, we used 2000 data to proxy minority those shares for the first two migration periods. For the most recent migration period, we used information on Han population shares from the 2005 census;

(vii) Mean real per capita income. Due to lack of available data for consecutive years during the 1980s and 1990s, income data only for 1989 (1999) were used to measure average annual income for the 1985-90 (1995-2000) periods. For 2000-05, though, we use annual mean incomes. All income data are adjusted for cost of living differences using provincial CPI measures;

(viii) Mean level of educational attainment. Educational attainment was measured as the percentage of the population aged 22-60 enrolled in universities in 1990 (for the 1985-90 period), in 2000 (for the 1995-2000 period), and in 2005 (for the latest period). For all three periods, a large majority of a typical province's adult population was not enrolled in universities, due to substantial barriers to to post-secondary education in China. However, as reforms deepened and barriers to access fell, the percentage of the population enrolled at universities rose at an increasing rate, from over 3% in 1990 to nearly 9% in 2005. Note also that the variance of enrollment rose at an increasing rate, a likely explanation for rising income inequality in China.

Data on the remaining variables are from Lin, Wang and Zhao (2004). Please refer to their paper for details on data sources and measurement of these variables.

V. COEFFICIENT ESTIMATES

Tables 6 through 9 provide OLS and 2SLS estimates of different versions of equation (17). Tables 6 (7) provide OLS (2SLS) estimates for the full panel, which exclude past migration as an explanatory variable. Because of the omission of this key variable, there is the distinct possibility of omitted variables bias. The advantage of the full panel, however, is that it allows us to get a better sense of whether or not there has been long term structural change in migration in post-reform China. Tables 8 (9) provide OLS (2SLS) estimates for the panel comprising the last two periods only, which includes migrant networks. The advantage of this smaller panel is the reduced likelihood of omitted variables bias due to the inclusion of migrant networks.

It is important to emphasize at the outset that our theory utilizes the chronological age of the individual, whereas our empirical work utilizes two distinct measures of a spatial unit's age distribution. It seems reasonable to assume that as average migrant age in a province rises, the relative size of the population aged 15-29 will fall and the ADR will rise. Therefore, our theory would imply that the signs of the marginal effects of the age 15-29 population share and ADR can be positive or negative. Therefore, in assessing our regression results, we are not predisposed to a particular sign on either of the age distribution measures.

We must also emphasize that interpretation of the numerical coefficients in Tables 6 and 9 requires some care due to the double-log functional form for the regression equations and because some of the independent variables are ratios. Note that each coefficient is an estimated migration *elasticity*, the percentage change in the relative flow of persons moving from province i to province j (out of all persons moving from i). Furthermore, some coefficients will be estimates of the percentage change in the migration rate when there is a one percentage change in a ratio. For example, the coefficient on the destination/origin income ratio measures the estimated percentage change in the migration rate when *relative* destination income changes by 1%. Note that all estimated equations in both tables include origin province fixed effects and time period controls, and are corrected for heteroskedasticity.

(A) *The effects of origin province age distribution on out-migration*

When one compares Tables 6 and 7, it is apparent that there is greater consistency in the results for age distribution when 2SLS estimation is performed (Table 7). Furthermore, for either estimation strategy many coefficient estimates vary dramatically with the inclusion of age interactions. According to Table 6, when interactions are excluded, there is no relationship between the 15-29 population share and out-migration. However, when four interactions are added,[21] the youth population share has a strong, positive effect on out-migration. Equation II in Table 6 predicts that a 1 percentage point increase in the youth population share raises the level of out-migration by nearly 174 percent. According to Table 7, when interaction terms are excluded, the elasticity of migration with respect to the youth population share is 8.73, meaning that a 1 percentage point increase in the share raises the level of out-migration by 8.73 percent. Note that when interaction terms are added in Table 7, the sign of the coefficient on the youth population share is positive but insignificant. In the main, the OLS results in Tables 6 and 7 indicate a positive relationship between the share of the origin's population aged 15-29 and the scale of out-migration, a result which generally supports the findings in Western literature.

The results in Tables 6 and 7 for the age dependency ratio (ADR) are stronger and generally suggest a negative relationship between the fraction of the non-working population in the origin province and the scale of out-migration. According to Table 6, when interaction terms are excluded, there is no apparent relationship between ADR and out-migration, but the relationship is negative and strong when interactions are added. Specifically, equation IV in Table 6 predicts that a one percentage point decrease in ADR will raise the out-migration rate by 113.93 percent. The results in Table 7 suggest a consistently negative relationship, with the strength of the relationship dramatically higher when interactions are added (equation IV).

[21] The previous section mentioned five interactions, the fifth being the interaction between age and past migration. For the full panel, that interaction was not included owing to migrant networks not being an explanatory variable.

The results for the estimated coefficients on the interaction terms are generally very supportive of the theoretical model. For example, equation II in Table 6 shows that the marginal effect of the youth population share on out-migration will be smaller the greater is distance and the marginal effect of the share will be smaller the higher are the odds of employment in the origin. As another example, equation IV in Table 7 indicates that the marginal effect of ADR on out-migration will be greater the lower is the risk of unemployment in the destination and the more attractive are income opportunities in the destination relative to the origin.

The results for the smaller panel (Tables 8 and 9) are more mixed. According to Table 8, with or without interactions there is no relationship between the youth population share and out-migration. According to Table 9, however, there is a relationship but its sign depends on whether or not interactions are included. Since an empirical specification with interaction terms is more consistent with the underlying theory, we view the results for equation II in both tables to be more relevant. However, those results indicate a negative relationship between the youth population share and out-migration, not consistent with what has been found in previous literature. The results for ADR in Tables 8 and 9 tend to point to a negative marginal effect. For example, equation IV in Table 8 predicts that a one percentage point decrease in ADR will raise the out-migration rate by 241.03 percent, whereas equation III in Table 9 shows a much milder elasticity. There is reasonably strong evidence in both tables of interaction effects between age distribution and migration. For example, both tables show a positive interaction between the relative size of the migrant network and ADR. This means that the sensitivity of the out-migration rate to ADR is larger the bigger is the size of the migrant network in the destination.

(B) *Other results*

Below we highlight a number of important patterns for our other results:

(i) The addition of interaction terms can change many coefficient estimates, sometimes dramatically. For example, note from Table 9 that the coefficient on distance goes from the hypothesized negative sign to a positive (and significant) sign when interactions are added to equation I. This happens frequently across Tables 6-9 and is likely due to multicollinearity between the age distribution measure and its interactions, as well as multicollinearity between the four (Tables 6 and 7) or five (Tables 8 and 9) explanatory variables that are interacted with age and those same explanatory variables;

(ii) Tables 8 and 9 illustrate the strong and robust effects of migrant networks on the scale of interprovincial migration. Across different specifications, the elasticity of migration with respect to past migration flows averages about 0.63, implying that a one-point increase in past migration flows (as a percentage of current population in the destination) is estimated to cause the migration rate to rise by approximately 0.63%. Of equal importance is that when past migration flows are omitted from the regressions (Tables 6 and 7), some coefficients change dramatically and the adjusted R-squared falls from approximately 75% to between 55% and 60%. These results illustrate that failure to control for migrant networks is likely to lead to omitted variables bias;

(iii) For the full panel, the odds of obtaining Hukou in the destination appear not to affect migration flows, but they do affect flows in the smaller panel. Referring to Table 9, equations II and IV predict that a one percentage point increase in the odds of Hukou will raise the destination's in-migration rate by between 2.29 percent and 5.13 percent. This supports previous findings of Bao, Bodvarsson, Hou and Zhao (forthcoming);

(iv) The "classic" determinants of migration – distance and spatial differences in income, unemployment risk, and climate – consistently affect the scale of migration flows in the hypothesized directions. For example, the coefficient estimates for spatial differences in provincial mean temperatures are consistently positive and robust across Tables 6 through 9;

(v) For the full panel only, higher FDI in the destination relative to the origin encourages more in-migration. We find that for both Tables 6 and 7, across all specifications the coefficient estimate on the FDI variable is positive and significant, with the elasticity of migration averaging approximately 5.5%. This adds to earlier findings by Bao, Bodvarsson, Hou and Zhao of the stimulating effects of FDI on in-migration;

(vi) Migration between adjacent provinces is consistently stronger than between more remote provinces. A very robust result across all the tables is that the dummy variable which controls for flows between provinces that share common border is positive and significant. This supports the findings of Poncet (2006). Note, though, that the migration elasticities are considerably lower for the small panel regressions;

(vii) The degree of urbanization and share of adult population enrolled in universities affect migration rates considerably more in the smaller panel regressions; Spatial differences in industry mix (measured by percentage of provincial employment in the manufacturing sector) and minority population shares do not appear to affect migration rates.

VI. CONCLUDING REMARKS

We view this study as making two important contributions to the literature on migration in China, as well as the literature on the relationship between migrant age and the scale of migration. First, we link together two major long term demographic shifts in China – the post-reform surge in internal migration and the shifting of China's age distribution that resulted from the one child policy and the massive structural changes in the economy. We find strong evidence that the Great Chinese Internal Migration episode has been influenced by shifts in the national age distribution. Recognizing the possibility that migration is endogenous to age distribution, we used both OLS and 2SLS to estimate the effects of age distribution on migration. We found that the OLS and 2SLS estimates differed significantly for some specifications, indicating the strong likelihood of a simultaneous relationship. We believe our results for the effects of age distribution are important because they clearly demonstrate that shifts in the age distribution in China are capable of generating often sizeable changes in migration rates. China's aging appears to have very important implications for labor mobility.

The second contribution of the paper, which is important for both the case of China and the Western migration literature at large, is that we analyzed in considerable detail the theoretical relationship between migrant age and the net benefits of migration. With the exception of Lundborg (1991), who did derive some testable implications for age in his migration model, we believe our study goes the farthest in examining how exactly age affects the expected benefits and costs of relocation. We contend that the traditional Becker (1964) view that younger migrants always have a higher likelihood of moving is just one of a set of important, conflicting explanations. We argue that it is important to also consider the effects of age on psychic costs, out-of-pocket costs, the loss of specific assets, and spatial differences in the returns to different types of human capital. Furthermore, for the China case one must also consider the effects of age on the costs of securing local registration. Our model demonstrated that when one blends all these explanations together, it is not generally true that older migrants have a lower propensity to migrate. This may explain the diversity of results for the age variable across both Western and Chinese studies.

We recommend that the next steps in this research are: (1) to apply the theoretical model to Western cases of internal migration; (2) test the model on micro data from carefully designed household surveys in different regions and in different policy environments; and (3) use the results obtained from estimating our migration equations to carefully construct a model and test of age distribution with the ultimate goal of testing a general equilibrium theory of migration and age distribution. Furthermore, as new waves of census data become available in China, studies of structural change in migration, particularly as it relates to aging, will become more feasible.

VII. REFERENCES

Bao, S., Bodvarsson, Ö., Hou, J., & Y. Zhao (2008). The Deregulation of People Flows in China: Did the Structure of Migration Change? *The Indian Journal of Labour Economics*, 51, 587-602.

Bao, S., Bodvarsson, Ö., Hou, J., & Y. Zhao (2008). Migration in China During 1985-2000: The Effects of Past Migration, Investments and Deregulation. *The Chinese Economy*, 42, 7-28.

Bao, S., Bodvarsson, Ö, Hou, J., & Y. Zhao (Forthcoming). The Regulation of Migration in a Transition Economy: China's Hukou System. Accepted in *Contemporary Economic Policy*.

Bao, S., Hou, J., & Shi, A. (2006). Migration and Regional Development in China. In S. Bao, S. Lin and C. Zhao (Eds.), *Chinese Economy after WTO Accession*. Aldershot, UK: Ashgate.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis.* New York: Columbia University Press.

Bowles, S. (1970). Migration as Investment: Empirical Tests of the Human Investment Approach to Geographical Mobility. *Review of Economics and Statistics*, 52, 356-62.

Chen, A., & Coulson, N.E. (2002). Determinants of Urban Migration: Evidence from Chinese Cities. *Urban Studies*, 39, 2189-97.

Clark, X., Hatton, T., & Williamson, J. (2007). Explaining U.S. Immigration, 1971-1998. *Review of Economics and Statistics*, 89, 359-73.

David, P.A. (1974). Fortune, Risk and the Micro-Economics of Migration. In P.A. David and M.W. Reder (Eds.), *Nations and Households in Economic Growth*. New York, NY: Academic Press.

Fan, C.C. (2005). Modeling Interprovincial Migration in China, 1985-2000. *Eurasian Geography and Economics* 46, 165-84.

Goodkind, D., & West, L.A. (2002). China's Floating Population: Definitions, Data and Recent Findings. *Urban Studies*, 39, 2237-50.

Goss, Ed., & C. Paul (1986). Age and Work Experience in the Decision to Migrate. *Journal of Human Resources*, 21, 397-405.

Greenwood, M.J. (1969). An Analysis of the Determinants of Geographic Mobility in the United States. *Review of Economics and Statistics,* 51,189-94.

Hare, D. (1999). 'Push' versus 'Pull' Factors in Migration Outflows and Returns: Determinants of Migration Status and Spell Duration Among China's Rural Population. *Journal of Development Studies*, 35, 45-72.

Li, H., & Zahniser, S. (2002). The Determinants of Temporary Rural-to-Urban Migration in China. *Urban Studies*, 39, 2219-35.

Liang, Z. (2001). The Age of Migration in China. *Population and Development Review*, 27, 499-524.

Liang, Z., Chen, Y.P., & Gu, Y. (2002). Rural Industrialisation and Internal Migration in China. *Urban Studies,* 39, 2175-87.

Liang, Z., & White, M.J. (1996). Internal Migration in China, 1950-88. *Demography*, 33, 375-84.

Liang, Z., & White, M.J. (1997). Market Transition, Government Policies, and Interprovincial Migration in China: 1983-1988. *Economic Development and Cultural Change,* 45, 321-39.

Lin, J., Wang, G., & Zhao, Y. (2004). Regional Inequality and Labor Transfers in China. *Economic Development and Cultural Change*, 52, 587-603.

Lundborg, P. (1991). An Interpretation of the Effects of Age on Migration: Nordic Migrants' Choice of Settlement in Sweden. *Southern Economic Journal*, 58, 392-405.

Ma, Z., & Liaw, K. (1997). Explaining Hierarchical and Interprovincial Migrations of Chinese Young Adults by Personal Factors and Place Attributes: A Nested Logit Analysis. *Mathematical Population Studies*, 6, 217-39.

Naskoteen, R.A., & Zimmer, M. (1980). Migration and Income: The Question of Self-Selection. *Southern Economic Journal*, 46, 840-51.

Navratil, F., & J. Doyle (1977). The Socioeconomic Determinants of Migration and the Level of Aggregation. *Southern Economic Journal*, 43, 1547-59.

Poncet, S. (2006). Provincial Migration Dynamics in China: Borders, Costs and Economic Motivations. *Regional Science and Urban Economics*, 36, 385-98.

Schwarz, A. (1976). Migration, Age, and Education. *Journal of Political Economy*, 84, 701-19.

Schlottmann, A., & H. Herzog (1981). Employment Status and the Decision to Migrate. *Review of Economics and Statistics*, 63, 590-98.

Shi, A., & Bao, S. (2007). Migration, Education and Rural Development: Evidence from China 2000 Population Census Data. *Journal of Chinese Economic and Business Studies*, 5, 163-77.

Shi, X., Heering, N., & Futian, Q. (2007). Choices Between Different Off-farm Employment Sub-categories: An Empirical Analysis for Jiangxi Province, China. *China Economic Review*, 18, 438-55.

Sjaastad, L. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70, 80-93.

Wu, Z. (2008). Self-selection and Earnings of Migrants: Evidence from Rural China. *Discussion Papers in Economics and Business from Osaka University,* No. 08-24.

Wu, Z. (2009). Rural Labor Migration in China: An Overview. *Osaka Economic Papers*, 59, 52-70.

Zhao, Y. (1997). Labor Migration and Returns to Rural Education in China. *American Journal of Agricultural Economics*, 79, 1278-87.

Zhao, Y. (1999a). Labor Migration and Earnings Differences: The Case of Rural China. *Economic Development and Cultural Change*, 47, 767-82.

Zhao, Y. (1999b). Leaving the Countryside: Rural-to-Urban Migration Decisions in China. *American Economic Review,* 89, 281-86.

Zhao, Y. (2002). Causes and Consequences of Return Migration: Recent Evidence from China. *Journal of Comparative Economics,* 30, 376-94.

Zhao, Y. (2003). The Role of Migrant Networks in Labor Migration: The Case of China. *Contemporary Economic Policy*, 21, 500-11.

Zhu (2002). The Impact of Income Gaps on Migration Decisions in China. *China Economic Review*, 13, 213-30.

TABLE 1
Share of population aged 15-29 by province and year

PROVINCE	1982	1990	1995	2000	2005
Beijing	33.9%	28.8%	23.0%	27.7%	25.8%
Tianjin	32.6	26.7	21.6	24.5	24.3
Hebei	30.4	27.8	23.9	24.6	24.1
Shanxi	29.7	30.7	25.7	24.7	21.2
Inner Mongolia	31.4	32.5	28.5	27.1	21.4
Liaoning	33.2	29.9	24.9	22.9	19.3
Jilin	32.8	32.3	27.0	25.2	21.7
Heilongjiang	31.9	33.2	28.8	26.1	20.7
Shanghai	33.5	22.4	18.4	24.4	25.6
Jiangsu	29.9	29.7	25.9	22.5	20.1
Zhejiang	30.0	30.2	24.0	24.9	20.8
Anhui	26.9	33.1	28.5	23.6	18.4
Fujian	29.3	30.7	27.2	29.3	23.8
Jiangxi	27.5	31.6	26.9	26.7	18.7
Shandong	29.8	29.7	24.9	23.5	20.7
Henan	27.9	31.0	27.1	24.0	21.9
Hubei	29.8	31.1	24.4	25.4	18.6
Hunan	28.4	31.6	25.0	24.6	20.2
Guangdong	29.4	30.2	24.8	32.4	28.5
Guangxi	27.5	29.5	26.7	26.4	21.1
Hainan	30.0[i]	30.0	26.8	27.6	24.9
Sichuan	26.4	33.7	27.1	23.0	14.9
Guizhou	25.2	31.6	29.4	25.1	18.8
Yunnan	26.9	32.0	30.6	28.4	23.0
Shaanxi	29.9	30.4	25.3	24.3	20.2
Gansu	28.6	34.5	29.0	24.5	20.4
Qinghai	27.3	35.2	33.5	28.6	23.6
Ningxia	28.2	32.6	30.0	28.3	23.5
Xinjiang	27.0	31.9	30.3	29.0	25.1
MEAN	**29.49%**	**30.84%**	**26.52%**	**25.84%**	**21.77%**

Source: China Census (from China DataOnline)

TABLE 2
Age dependency ratio by province and year

PROVINCE	**1982**	**1990**	**1995**	**2000**	**2005**
Beijing	38.3%	36.1%	37.8%	28.2%	26.7%
Tianjin	46.8	41.2	42.3	33.6	28.8
Hebei	56.6	53.5	53.7	42.5	34.9
Shanxi	62.2	50.5	52.3	47.2	40.9
Inner Mongolia	64.3	48.0	45.1	36.5	33.3
Liaoning	50.4	40.7	40.2	34.3	31.6
Jilin	56.2	44.3	40.0	33.3	28.5
Heilongjiang	62.2	43.6	39.0	32.4	28.9
Shanghai	34.4	38.1	40.0	31.1	26.5
Jiangsu	52.7	44.0	43.2	39.8	35.9
Zhejiang	55.1	43.1	43.7	36.9	36.0
Anhui	67.3	51.0	51.6	49.4	49.7
Fujian	69.1	57.6	57.5	42.2	37.4
Jiangxi	77.9	58.4	57.9	47.4	50.5
Shandong	57.3	48.8	47.0	40.8	34.9
Henan	67.0	54.1	54.6	49.3	41.6
Hubei	60.5	51.4	54.7	41.3	39.0
Hunan	63.7	50.5	53.4	42.0	40.5
Guangdong	64.3	55.9	61.4	43.4	40.4
Guangxi	75.1	63.4	61.3	50.4	49.9
Hainan	30.0[i]	30.0	61.6	42.6	46.3
Sichuan	65.2	40.6	45.1	42.9	47.5
Guizhou	83.6	59.5	56.4	56.6	57.6
Yunnan	77.4	57.6	51.9	47.2	46.3
Shaanxi	60.3	51.6	52.9	45.1	39.2
Gansu	66.2	47.1	48.8	47.4	44.2
Qinghai	76.2	51.1	47.0	45.8	43.1
Ningxia	80.0	59.4	54.7	48.9	46.9
Xinjiang	76.2	58.6	53.4	46.9	42.9
MEAN	**61.94%**	**49.3%**	**49.94%**	**42.66%**	**39.65%**

Source: China Census (from China DataOnline)

Table 3
Summary Statistics for 1985-90 period
765 observations

Variable	Mean	Standard Deviation	Maximum	Minimum
Migration rate x 100[i]	3.775%	6.982%	79.336%	0.018%
Percentage of population aged 15-29 x 100[ii]	30.19%	1.12%	32.55%	27.95%
Age dependency ratio[ii]	56.37%	8.96%	71.55%	36.25%
Mean annual percentage of households with *Hukou* status during 1980-84 x 100	98.40%	1.412%	99.73%	94.94%
Real Mean Annual FDI Per Capita during 1980-84[iii]	$US 1.544	$US 5.947	$US 31.75	$US 0.0038
Real Mean Annual Fixed Asset Investment (FAI) Per Capita during 1980-84[iii]	163.77 Yuan	132.84 Yuan	518.71 Yuan	40.888 Yuan
Railway distance between capital cities	1,630.76 Kilometers	1.87 Kilometers	6,313.21 Kilometers	137 Kilometers
Real annual per capita income[iv]	510.95 Yuan	183.11 Yuan	1084.5 Yuan	340.53 Yuan
Percentage of adult population enrolled in universities	3.05%	3.01%	1.02%	16.29%
Unemployment rate x 100	1.178%	0.705%	4.11%	0.28%
Manufacturing share of employment x 100	23.44%	12.10%	59.3%	9.47%
Urban share of population x 100	31.03%	16.17%	73.44%	14.87%
Mean yearly temperature	14.113 C	5.176 C	24.517 C	4.608 C
Minority population share x 100	12.28%	16.06%	59.43%	0.31%
Doctors per 10,000 persons	5.40	2.33	9.21	0.87

[i]In the 1990 Census, a migrant is defined as someone who has moved from other towns or townships and has lived in this place for more than one year but less than five years.
[ii]Calculated as the mean value for 1982 and 1990
[iii] Computed using average annual CPI for 1980-84
[iv] Computed using income and average annual CPI for 1989 only

TABLE 4
Summary Statistics for 1995-2000 period
790 observations

Variable	**Mean**	**Standard Deviation**	**Maximum**	**Minimum**
Migration rate x 100[i]	3.589%	7.230%	87.317%	0.014%
Mean annual percentage of households with *Hukou* status during 1990-94 x 100	90.38%	5.443%	96.01%	74.97%
Percentage of population aged 15-29 x 100[ii]	26.18%	2.15%	31.05%	21.40%
Age dependency ratio x 100[ii]	42.66%	6.65%	44.18%	33.00%
Past migration flows during 1985-90	1,351,400	3,439,300	44,320,000	10,000
Real Mean Annual FDI Per Capita during 1990-94[iii]	$US 16.14	$US 24.25	$US 92.73	$US 0.58
Real Mean Annual Fixed Asset Investment (FAI) Per Capita during 1990-94[iii]	871.66 Yuan	717.63 Yuan	3393.2 Yuan	229.7 Yuan
Real annual per capita income[iv]	1,069 Yuan	442.2 Yuan	2,451.5 Yuan	605.26 Yuan
Percentage of adult population enrolled in universities	5.92%	3.58%	3.13%	20.5%
Unemployment rate x 100	4.40%	2.41%	9.64%	1.36%
Manufacturing share of employment x 100	22.83%	9.82%	49.25%	9.17%
Mean yearly temperature	14.113C	5.176C	24.517C	4.608C
Urban share of population x 100	40.20%	18.56%	90.67%	18.63%
Minority population share x 100	12.28%	16.06%	59.43%	0.31%
Doctors per 10,000 persons	6.71	3.13	13.24	0.96

[i] In the 2000 Census, a migrant is defined as someone who has moved from other towns or townships and has lived in this place for more than one year but less than five years.
[ii] Calculated as the mean of 1995 and 2000
[iii] Computed using average annual CPI for 1990-94
[iv] Computed using income and average annual CPI for 1999 only

TABLE 5
Summary Statistics for 2000-05 period
790 observations

Variable	Mean	Standard Deviation	Maximum	Minimum
Migration rate x 100[i]	3.655%	7.387%	87.32%	0.01%
Migration flows during 1995-2000	1,218,100	3,531,800	45,360,000	10,000
Mean annual percentage of households with *Hukou* status during 1995-99 x 100	88.57%	6.726%	95.85%	67%
Percentage of population aged 15-29 x 100[ii]	23.79%	2.33%	30.45%	18.90%
Age dependency ratio x 100[ii]	44.10%	7.25%	57.10%	27.45%
Real Mean Annual FDI Per Capita during 1995-99[iii]	$US 44.64	$US 66.57	$US 253.05	$US 1.15
Real Mean Annual Fixed Asset Investment (FAI) Per Capita during 1995-99[iii]	2,452.8 Yuan	2,441.5 Yuan	12,705 Yuan	646.5 Yuan
Percentage of provincial FAI attributable to transportation infrastructure x 100	9.2%	3.06%	17.19%	3.98%
Real annual per capita income[iv]	5,122.3 Yuan	2,632.7 Yuan	13,484 Yuan	2,614.6 Yuan
Percentage of adult population enrolled in universities	8.69%	4.92%	28.05%	4.03%
Unemployment rate x 100	3.14%	1.50%	7.17%	1.21%
Manufacturing share of employment x 100	22.83%	9.82%	49.25%	9.17%
Mean yearly temperature	14.27C	5.24C	25.1C	4.70C
Urban share of population x 100	40.20%	18.56%	90.67%	18.63%
Minority population share x 100	12.83%	16.47%	60.13%	0.31%
Doctors per 10,000 persons	6.62	3.24	13.93	0.92

[i] In the 2005 Census, a migrant is defined as someone who has moved from other towns or townships and has lived in this place for more than one year but less than five years.
[ii] Calculated as mean of 2005 and 2000
[ii] Computed using average annual CPI for 1995-99
[iii] Computed using average annual income and average annual CPI for 2000-05

TABLE 6
OLS Results for Full Sample (1985-90, 1995-2000 and 2000-05 migration periods)
Dependent variable = log gross interprovincial migration rate
(Age distribution measures are for beginning of migration period)
(Standard Errors in Parentheses; ** denotes significant at 1%, * significant at 5%)

REGRESSOR	I	II	III	IV
Log share of population 15-29	-0.5366 (0.32)	173.97** (53.85)		
Log Age dependency ratio			0.2007 (0.33)	-113.93** (39.97)
Log odds of obtaining Hukou in destination	-0.3596 (1.28)	-0.3780 (1.234)	-0.3707 (1.23)	-0.8169 (1.27)
Log distance	-0.924** (0.06)	3.9423** (1.06)	-0.9225** (0.06)	1.1528* (0.58)
Log dest./origin income ratio	1.8006** (0.12)	0.1172 (1.37)	1.7936** (0.121)	-1.4702 (0.93)
Log percentage share of adults in origin enrolled in university	0.0054 (0.14)	0.0395 (0.142)	-0.0054 (0.14)	0.1896 (0.15)
Log percentage share of adults in destination enrolled in university	-0.1629 (0.09)	-0.1862** (0.088)	-0.1649 (0.09)	-0.1829* (0.09)
Log odds of employment in origin	-3.246 (2.60)	101.81** (34.03)	-5.3889** (2.29)	-53.327* (32.04)
Log odds of employment in destination	7.6217** (1.60)	17.716 (31.71)	7.6321** (1.60)	-40.786 (25.06)
Log ratio of dest./origin urban population shares	-0.0709 (0.12)	-0.1425 (0.13)	-0.0635 (0.124)	-0.1567 (0.13)
Log ratio of dest./origin mean temperatures	0.4259** (0.07)	0.3818** (0.069)	0.4351** (0.069)	0.3595** (0.07)
Log ratio of dest./origin manufacturing employment shares	0.0378 (0.91)	-0.0302 (0.091)	0.0453 (0.09)	-0.0549 (0.09)
Log ratio of dest./origin minority shares	-0.0056 (0.018)	-0.042 (0.02)	-0.0047 (0.018)	-0.0153 (0.018)
Log ratio of dest./origin per capita fixed asset investment	-0.079** (0.026)	-0.0792** (0.026)	-0.0763** (0.026)	-0.0793** (0.026)

Log ratio of dest./origin per capita FDI	0.0579** (0.017)	0.0601** (0.017)	0.0553** (0.017)	0.0514** (0.017)
Dummy = 1 if flow to adjacent province	0.8349** (0.091)	0.8328** (0.09)	0.8361** (0.091)	0.7897** (0.091)
Early period dummy	-0.1480 (0.23)	0.0129 (0.25)	-0.1971 (0.26)	-0.2759 (0.26)
Late period dummy	-0.1486 (0.095)	-0.1477 (0.095)	-0.0719 (0.10)	-0.1952 (0.12)
Log share of population 15-29 x Log distance		-1.4798** (0.32)		
Log Age dependency ratio x Log distance				-0.5449** (0.151)
Log share of population 15-29 x Log odds of employment in origin		-32.454** (10.59)		
Log age dependency ratio x Log odds of employment in origin				13.773 (8.67)
Log share of population 15-29 x Log odds of employment in destination		-3.3553 (9.67)		
Log age dependency ratio x Log odds of employment in destination				12.082 (6.36)
Log share of population 15-29 x Log dest./origin income ratio		0.5555 (0.43)		
Log age dependency ratio x Log dest./origin income ratio				0.8811** (0.242)
Constant	-8.844 (14.27)	-570.71** (173.4)	-1.6571 (14.27)	425.24** (148.0)
Adjusted R-squared	0.5700	0.576	0.5696	0.5756
SSE	2359	2322.2	2361.4	2324.5
Sample size	2,385	2,385	2,385	2,385

TABLE 7
2SLS Results for Full Sample (1985-90, 1995-2000 and 2000-05 migration periods)
Dependent variable = log gross interprovincial migration rate
(Age distribution measures are for end of migration period)
(Standard Errors in Parentheses; ** denotes significant at 1%, * significant at 5%)

REGRESSOR	I	II	III	IV
Log share of population 15-29	8.7382** (1.152)	26.285 (65.45)		
Log age dep. ratio			-3.0115** (0.849)	-96.244* (47.06)
Log odds of Hukou in destination	0.3498 (1.24)	0.3105 (1.38)	-0.5157 (1.24)	-1.0037 (1.24)
Log distance	-0.9414** (0.057)	1.2514* (0.72)	-0.9253** (0.058)	1.4404* (0.673)
Log dest./origin income ratio	1.9675** (0.12)	0.6626 (1.11)	1.7679** (0.12)	-3.2965** (1.08)
Log share of adults in origin enrolled in university	0.0342 (0.14)	0.0056 (0.14)	-0.0032 (0.14)	0.1603 (0.14)
Log share of adults in destination enrolled in university	-0.079 (0.09)	-0.087 (0.09)	-0.1303 (0.09)	-0.1753* (0.09)
Log odds of employment in origin	-2.7794 (2.13)	64.652 (41.03)	-5.359** (2.12)	-14.90 (34.88)
Log odds of employment in destination	6.4507** (1.585)	-52.987* (29.2)	7.0933** (1.593)	-60.012 (24.35)
Log ratio of dest./origin urban population shares	-0.2552* (0.125)	-0.3089** (0.127)	-0.1151 (0.124)	-0.1872 (0.13)
Log ratio of dest./origin mean temperatures	0.3996** (0.067)	0.3956** (0.0677)	0.4411** (0.067)	0.3711** (0.067)
Log ratio of dest./origin manufacturing employment shares	-0.0326 (0.08)	-0.0347 (0.08)	0.0551 (0.08)	-0.0019 (0.08)
Log ratio of dest./origin per capita fixed asset investment	-0.0825** (0.026)	-0.0793** (0.026)	-0.0857** (0.026)	-0.0887** (0.025)
Log ratio dest./origin per capita FDI	0.0556** (0.017)	0.0502** (0.017)	0.0584** (0.017)	0.0575** (0.017)
Dummy for flow to adjacent province	0.8076** (0.087)	0.8177** (0.089)	0.8281** (0.089)	0.7684** (0.09)

Early period dummy	-1.5574** (0.287)	-1.604** (0.294)	0.4435 (0.29)	0.3843 (0.30)
Late period dummy	1.4254** (0.22)	1.5272** (0.24)	-0.3305** (0.11)	-0.4068** (0.117)
Log share population 15-29 x Log distance		-0.6752** (0.221)		
Log Age dependency ratio x Log distance				-0.6389** (0.182)
Log share of population aged 15-29 x Log odds of employment in origin		-20.883 (12.75)		
Log age dependency ratio x Log odds of employment in origin				3.8977 (10.04)
Log share of population aged 15-29 x Log odds of employment in destination		18.226* (9.01)		
Log age dependency ratio x Log odds of employment in destination				17.548** (6.43)
Log share of population aged 15-29 x Log dest./origin income ratio		0.4212 (0.35)		
Log age dependency ratio x Log dest./origin income ratio				1.3865** (0.293)
Constant	-40.025** (14.17)	-93.876 (209)	13.717** (13.87)	348.76* (164.1)
Adjusted R-squared	0.5799	0.5819	0.5719	0.5785
SSE	2305.6	2290.7	2349.5	2309.2
Sample size	2,385	2,385	2,385	2,385

TABLE 8
OLS Results for later two periods
Dependent variable = log gross interprovincial migration rate
Age distribution measured at beginning of migration period
(Standard Errors in Parentheses; ** denotes significant at 1%, * significant at 5%)

REGRESSOR	I	II	III	IV
Log percentage of population 15-29	0.4956 (0.37)	-54.228 (62.96)		
Log ADR			-0.2655 (0.49)	-241.03** (75.01)
Log past migration	0.5382** (0.022)	0.7319 (0.62)	0.5377** (0.021)	-0.7704 (0.38)
Log odds of Hukou in destination	2.7622** (1.25)	2.9034** (1.231)	2.7102** (1.246)	2.9346** (1.251)
Log distance	-0.3341** (0.06)	3.9651** (1.42)	-0.3335** (0.06)	-2.2057** (0.94)
Log dest./origin income ratio	1.2424** (0.113)	5.3293** (1.87)	1.2307** (0.113)	3.6498** (1.19)
Log share of adults in origin enrolled in university	0.2126 (0.18)	0.1440 (0.174)	0.1979 (0.18)	0.1932 (0.18)
Log share of adults in destination enrolled in university	-1.032** (0.107)	-1.0122** (0.108)	-1.0305** (0.108)	-1.0218** (0.108)
Log odds of employment in origin	2.5126 (3.14)	-53.11 (36.12)	3.3882 (3.14)	-192.8** (57.83)
Log odds of employment in destination	7.90** (1.74)	18.215 (36.01)	7.9586** (1.738)	3.1001 (27.88)
Log ratio of dest./origin urban population shares	0.6252** (0.12)	0.5917** (0.117)	0.6290** (0.117)	0.5760** (0.117)
Log ratio of dest./origin mean temperatures	0.6644** (0.079)	0.6566** (0.079)	0.6619** (0.079)	0.6191** (0.081)
Log ratio of dest./origin manufacturing employment shares	0.0232 (0.09)	0.0288 (0.09)	0.0257 (0.08)	0.0191 (0.08)
Log ratio of dest./origin minority shares	-0.0308 (0.017)	-0.0384** (0.017)	-0.0309 (0.017)	-0.0354* (0.017)
Log ratio of dest./origin per capita fixed asset investment	-0.0436 (0.038)	-0.0409 (0.037)	-0.0715 (0.037)	-0.0422 (0.037)

Log ratio of dest./origin per capita FDI	0.0123 (0.031)	0.0110 (0.03)	0.0150 (0.031)	0.0134 (0.03)
Dummy for flow to adjacent province	0.2638** (0.074)	0.2727** (0.075)	0.2651** (0.074)	0.2742** (0.076)
Late period dummy	0.1593 (0.103)	0.1579 (0.103)	0.0952 (0.13)	0.1985 (0.14)
Log share of population aged 15-29 x Log distance		-1.3302** (0.44)		
Log ARD x Log distance				0.4965* (0.25)
Log share of population aged 15-29 x Log past migration		-0.0601 (0.19)		
Log ADR x Log past migration				0.3439** (0.10)
Log share of population aged 15-29 x Log odds of employment in origin		17.255 (11.19)		
Log ADR x Log odds of employment in origin				51.321** (15.21)
Log share of population aged 15-29 x Log odds of employment in destination		-3.2693 (11.01)		
Log ADR x Log odds of employment in destination				1.2078 (7.18)
Log share of population aged 15-29 x Log dest./origin income ratio		-1.2468* (0.578)		
Log ADR x Log dest./origin income ratio				-0.6243* (0.312)
Constant	-53.849 (17.33)	122.89 (204.7)	-55.071** (18.38)	865.13** (285.2)
Adjusted R-squared	0.753	0.7757	0.7527	0.7567
SSE	815.35	803.67	813.36	800.53
Sample size	1,535	1,535	1,535	1,535

TABLE 9
2SLS Results for later two periods
Dependent variable = log gross interprovincial migration rate
Age distribution measured at end of migration period
(Standard Errors in Parentheses; ** denotes significant at 1%, * significant at 5%)

REGRESSOR	I	II	III	IV
Log share population 15-29	4.1047** (0.862)	-3.3974** (1.11)		
Log ADR			-3.2446** (0.714)	4.5963** (1.246)
Log past migration	0.5425** (0.022)	1.5486** (0.349)	0.5434** (0.022)	-0.5104** (0.024)
Log odds of Hukou in destination	1.5624 (1.26)	5.1252** (1.266)	1.4280 (1.27)	2.2889* (1.24)
Log distance	-0.2993** (0.058)	1.7333** (1.05)	-0.2935** (0.058)	-1.9922** (0.847)
Log dest./origin income ratio	0.9824** (0.112)	5.3579** (1.305)	0.9651** (0.113)	2.3148* (1.229)
Log share of adults in origin at university	0.5183** (0.178)	0.2201 (0.174)	0.5712** (0.176)	0.5912** (0.176)
Log share of adults in destination at university	-1.0089** (0.11)	-0.9661** (0.108)	-0.9978** (0.108)	-0.9826** (0.106)
Log odds of employment in origin	0.7133 (2.99)	-12.554 (29.68)	1.1064 (3.01)	25.713 (23.75)
Log odds of employment in destination	8.0203** (1.73)	8.7741 (28.51)	7.9777** (1.728)	-18.408 (24.17)
Log ratio of dest./origin urban population shares	0.6400 (0.12)	0.5101** (0.116)	0.6315** (0.119)	0.5698** (0.117)
Log ratio of dest./origin mean temperatures	0.7463** (0.08)	0.7346** (0.076)	0.7495** (0.079)	0.7101** (0.08)
Log ratio of dest./origin manufacturing employment shares	0.2012** (0.076)	0.0555 (0.07)	0.2067** (0.076)	0.1518* (0.074)
Log ratio of dest./origin per capita fixed asset investment	-0.0851** (0.038)	-0.0727 (0.037)	-0.0918** (0.038)	-0.0801** (0.037)
Log ratio dest./origin per capita FDI	0.0354 (0.031)	0.0087 (0.03)	0.0393 (0.031)	0.0357 (0.03)

Dummy for flow to adjacent province	0.2859** (0.074)	0.2741** (0.073)	0.2893** (0.074)	0.2924** (0.075)
Later period dummy	0.7507** (0.197)	2.1482** (0.229)	-0.2539* (0.11)	-0.2005 (0.109)
Log share of population 15-29 x Log distance		-0.6552* (0.335)		
Log ADR x Log distance				0.4638* (0.233)
Log share of population aged 15-29 x Log past migration		-0.3236** (0.112)		
Log ADR x Log past migration				0.2841** (0.065)
Log share of population aged 15-29 x Log odds of employment in origin		4.4224 (8.917)		
Log ADR x Log odds employment in origin				-9.327 (6.522)
Log share of population aged 15-29 x Log odds of employment in destination		-0.7052 (8.87)		
Log ADR x Log odds employment in destination				7.0708 (6.47)
Log share of population aged 15-29 x Log dest./origin income ratio		-1.2665** (0.418)		
Log ADR x Log dest./origin income ratio				-0.3295 (0.33)
Constant	-53.942** (17.70)	-43.371 (17.96)	-29.012 (16.59)	-15.528 (16.96)
Adjusted R-squared	0.7439	0.7659	0.7542	0.7538
SSE	846.58	771.15	848.04	811.09
Sample size	1,535	1,535	1,535	1,535