Appendix Discussion and Additional Tables and Figures

for

Migrant Opportunity and the Educational Attainment of Youth in Rural China

Alan de Brauw
International Food Policy
Research Institute

John Giles
The World Bank
& IZA

The following pages include additional documentation on enrollment patterns and duration of schooling, a discussion of returns to education among migrants and results from estimating the impact of migration on the probability of high school enrollment using an alternative identification strategy. Also included is supplementary evidence in additional appendix figures and tables referenced in the body of the main paper.

A.1. Evidence on Age of School Enrollment, Duration in School and Incidence of Repeated Grades

China's Compulsory Education Law, passed in 1986, mandated that all children enroll in school at age six regardless of gender, but allows enrollment to be postponed to age seven in those localities (urban and rural) where enrollment at age six is not feasible.[1] In its aspiration for early start dates, the implementation of the Compulsory Education Law bears striking similarity to other legal and institutional reforms in China (ranging from governance over land, to employment protections and protections over intellectual property, and as exploited in this paper, the distribution of national identity cards). In education, as in all of these other areas, legislation is used to set out a future national goal, but provinces, counties and sometimes even local communities (villages and urban neighborhoods) are granted flexibility to implement these changes as local conditions evolve. The high incidence of an age seven start date for school enrollment is evident from both census data and other surveys conducted in the early 2000s. China's 1990 national census shows that only 17 and 16 percent of male and female rural children who turned 6 by August 1990, respectively, were enrolled in school (Appendix Figure A4). By contrast, the enrollment rates of seven year olds were 60 and 55 percent for boys and girls, respectively.

We also note that data from a six-province survey collected by the Chinese Center for Agricultural Policy (CCAP) in 2000 lends further support to our assumption that children start school at age seven. Geographically, the CCAP sample overlaps with our sample as it is drawn from the same region of the country.[2] The CCAP survey form asked respondent household members to report the age each child began primary school, and the modal year of respondents answering in 2000 was seven (Appendix Table A.3). CCAP repeated their survey (in the same villages, with the same form) in 2009, and the modal answer was again seven. Another well-known survey (The Gansu Study of Families and Children) focusing on education in a poorer province, Gansu, enumerated the first grade start date in rural households as well, and the overwhelming majority of students surveyed in 2000 and 2004 also reported starting elementary school at age seven.

---

[1] 《中华人民共和国义务教育法》(1986年)，第五条，凡年满六周岁的儿童，不分性别、民族、种族，应当入学接受规定年限的义务教育。条件不具备的地区，可以推迟到七周岁入学。(Compulsory Education Law of the People's Republic of China, 1986, Article 5. All children who have reached the age of six shall enroll in school and receive compulsory education for the prescribed number of years, regardless of sex, nationality or race. In localities where this is not possible, enrollment may be postponed to the age of seven.) The Articles of Compulsory Education Law could be found here (both the English version and the Chinese version): http://www.lawinfochina.com/display.aspx?lib=law&id=1166&CGid.

[2] See de Brauw et al (2002) for a description of the CCAP survey.

By the time of the 2000 population census, progress on enrolling children in elementary school was significant, with 72 and 71 percent of boys and girls who were then age six, respectively, enrolled in school. Thus, between 1990 and 2000 there was considerable expansion of rural schools and success in reducing the age at which children started school to six, but most of the children in the RCRE sample used in the analysis of this paper were born earlier. The youngest students in the RCRE sample were of age to finish middle school in 2002 and decide about whether to enter high school or not in 2003, and these students would have entered first grade by 1994. Only 18 percent of our sample entered elementary school between 1991 and 1994, while the other 82 percent started primary school in 1990 or before.

Evidence on the timing of efforts to meet goals for lower enrollment ages suggests that the change was not smooth between 1990 and 2000. Efforts to lower enrollment rates were pursued in urban areas first, and then much of the focus on rural school enrollment ages occurred after the mid-1990s with other efforts to expand services in rural and more remote areas. Given evidence from the CCAP surveys, the Gansu Study of Families and Children, and enrollment rates at age six and seven from the 1990 census, there appears to be ample evidence to support our assumption that for most children in our sample, the first year of primary school enrollment was at age seven.

There will be some variation around this starting age, and as suggested by one of our critical readers, it is important to think about both how it will affect our estimates and the robustness checks that are worth performing. For identification, it is most important that our instrument for migrant employment is uncorrelated with unobserved differences in ability. First, phase in of the compulsory education law suggests that much variation in starting age is at the community level, and this is evident in the data. This dimension of variation is unrelated to student ability. Second, students starting school at age six or age eight likely make decisions about entering high school one year before or one year after the date we assume in this paper. We know whether the decision to enroll in another year of schooling was made, and being off by a year will mean that the predicted increase in migrant share from the village will not be perfectly mapped into the year of the decision. The quartic in years-since-IDs is a smooth function and we do not see extremely sharp jumps in the predicted migrant share variable from one year to the next, so it is unlikely that the instrument is systematically related to within village variation in unobserved ability, or that it will introduce bias to our analyses.

A second assumption on the duration of primary school is important for assigning the year of potential high school enrollment. In some parts of rural China, primary school lasted for five years for much of the period, whereas other regions achieved the six-year mandate much

earlier. The RCRE supplemental survey did not directly ask whether villages have five or six year primary schools, but when we examine completed years of schooling at the village level, it is fairly straightforward to discern whether completed schooling patterns are consistent with five or six year primary schools. We found that in some villages most children completed either 6, 9 or 12 years of school; as middle and high school each last three years, and these patterns were consistent with six year primary schools. In other villages, most children completed 5, 8, or 11 years of schooling, consistent with five year primary schools.[3] Using this information, we coded all of the villages as five or six year primary school villages. To illustrate our assumption, we show average enrollment rates for each grade level in five and six year primary school villages conditional on completing the previous grade (Table A.4). Until the decision to enter grade 9 or grade 10, transition probabilities from one grade to the next are well over 90 percent. The first significant decision node is clearly either the decision to move from grade 8 to grade 9 (in five year villages) or from grade 9 to grade 10 (in six year villages). We measure the decision to enroll in high school with a variable that includes the decision to enter grade 10 conditional on completing grade 9 for six year primary school villages and the decision to enter grade 9 conditional on completing grade 8 for five year primary school villages.[4]

A final concern involves handling repeated years of schooling or skipped grades. Although the supplemental survey did not ask explicitly about repeats or skips, the protocol for the supplemental survey required respondents to report years of schooling completed and the common interpretation is to answer in terms of the level of schooling completed. Examination of the CCAP data, which asked explicitly about repeats, suggests their inclusion does not affect the general distribution of educational attainment. Therefore our findings should be robust to any errors in the measurement of years of schooling.

A.2. An Alternative Instrument for Migration: GDP Growth in Potential Migrant Destinations as a Proxy for Labor Demand

An alternative approach to identifying migration involves using labor demand, or shocks at migrant destinations.[5] Measures of labor demand may identify effects of migration in sending

[3]In the one village in which our method was indeterminate, we assumed that the village had a five year primary school. Our results are robust to recoding the village as one with a six year primary school.

[4]All of our estimation results are robust to analyzing the grade 10 enrollment decision conditional on grade 8 completion.

[5]Several researchers have recently used variants of this approach. In looking at the effects of parent migration on child time allocation in Mexico, Antman (2011) uses shocks to weighted city-level employment in sectors thought to employ migrants in destination cities. Yang and Martinez (2006) and Yang (2008) use exchange rate shocks to identify the effects of remittances to the Philippines on household investment, poverty headcounts, and agricultural production, respectively. In all three cases, the authors weight the destination unemployment rate or exchange rate by the fraction of migrants from the source going to specific destinations.

communities if they only affect other household level outcomes through their effect on migrant labor supply decisions. While the prospective migrant observes expected wages at a set of potential destinations, and then uses that information to choose where to migrate, the full set of wage offers and potential destinations faced by any individual migrant are unobserved by the analyst. Therefore, implementing this approach involves assumptions regarding potential destinations and use of a proxy for labor demand.

First, both the RCRE surveys and the 1995 and 2005 agricultural censuses suggest that nearly 70 percent of rural migrants find work within their home provinces, we matched villages to the nearest city outside their home counties. In using the nearest city, we pick up the nearest urban labor market signal for potential migrants, and implicitly assume that it would be the best potential proxy for news about labor demand for young, first-time migrants. We use measures of city level GDP from the National Bureau of Statistics (CNBS), and construct year-to-year city-level growth rates.[6]

We test several potential specifications of instruments based on city GDP growth rates. Signals of labor market conditions may be picked up over the year and could show up in either lagged or current growth in city GDP. Thus, it is unclear whether year on year growth measured in year t, or $t-1$, is exerts stronger influence on the migration decision, and so we tried both. The strength of the growth signal might also depend upon the distance from the village to the nearest city, but as it is not clear that the relationship between distance and signal strength would be linear, we tested four different relationships between GDP growth and distance. First, we used the GDP growth rate in the nearest city itself, without controlling for the distance to the nearest city. Second, we divide the GDP growth rate by the distance, and in the third specification, we divide by the distance squared.

None of the above distance weightings above account for the fact that the effects of signal may vary non-linearly with distance. Individuals who live in relatively close proximity to a growing urban area may be able to commute for work, as opposed to migrating, even if the city is located in another county. Alternatively, after some distance X from the village, the effect of signals related to economic growth may decline more rapidly with distance. We thus posit that city GDP growth might have its strongest pull on village residents at that distance X, and that

[6]While we would prefer to use information on employment by industry sector, these data series were incomplete for many cities and in many years, and further, only include formal sector employment, and migrants are frequently employed in the small scale informal sector. For these reasons, we make use of the more complete city-level GDP from the series, from which we construct year to year growth rates.

this effect will decline for values greater or less than X. We therefore also experiment with an instrument that is defined as the growth rate divided by the squared difference between the distance D from the village to the nearest city and X, or $\frac{g}{(D-X)^2}$.

To test the plausibility of the various instrument specifications, we regress the share of the migrants from the village on each instrument described above, along with a full set of village dummies, and a full set of province-year dummy variables (Appendix Table A.8, columns 1-3). When using the contemporaneous growth rate, we use data from 1991 to 2003, and when we use lagged growth, we use data from 1993 to 2003. In the table, each cell represents a separate regression, and we list the numerator in the row (either contemporaneous or lagged GDP growth) and the denominator in the column. We find a significant relationship between the migrant share and the lagged GDP measure, but only when we control for distance either linearly or quadratically in the denominator. However, the F-statistics suggest that these would be very weak instruments and unlikely to be appropriate for estimating the second stage.

We next explore whether there is an optimal distance X from the village in which the employment signal would be strongest, as discussed above. We re-specify the denominator as $(X-D)^2$, and we perform a grid search over possible values of D to find the strongest relationship between the share of migrants in the village workforce and the instrument. For both GDP growth and lagged GDP growth, we find a strongly significant relationship between the migrant share and the instrument for values of D centered around 100 km. As it is generally implausible that rural residents could commute this distance over the period under study, or even much more than 25 km, this finding makes intuitive sense. We thus use a value of D=100km in column 4 of Appendix Table A.8.

We find that the growth instrument has the strongest relationship with the share of migrants from the village when we use contemporaneous growth. The cluster robust F-statistic is 7.64, which is below 10, but stronger than in other specifications. While one might prefer this specification based on instrument strength, we had some expectation that migration should follow the previous year's growth.

Thus, we first report the second stage result with no additional covariates and using lagged city GDP growth (Appendix Table A.9). When no additional control variables are included, the estimated coefficient is negative and significant at the 1 percent level (column 1). Perhaps most importantly, the point estimate (-2.12) is quite similar to the point estimate reported in Table 10 with no covariates (-2.50), and is within the range found when also controlling for additional covariates. From the perspective of demonstrating the validity of the quartic in years since IDs

were issued as an instrument, this result offers some confirmation as the result is similar using a different instrument.

When we add covariates to the regression (columns 2-5), the estimated coefficient on migrant share remains negative; however, it is no longer significantly different from zero. Point estimates are slightly smaller in magnitude than the point estimates found using the quartic in years since IDs were issued, but they are consistent with estimates in Tables 6 and 7. Apart from levels of significance, the Anderson-Rubin test statistics indicate that we cannot reject the joint hypothesis that the instrument does not affect the migrant share and that the migrant share does not affect school enrollment. When we use the quartic in years-since-IDs were issued, this test is consistently rejected.

We can be somewhat more sanguine about results using contemporaneous city GDP growth in potential destinations (Appendix Table A.10). Estimated coefficients on the share of the village population working as migrants range from -2.16 to -2.28, and the Anderson-Rubin test statistics suggest that while weak, the instruments are sufficiently strong and yield results consistent with the years-since-IDs instruments. As these results provide no evidence against the years-since-IDs instrument, we view them as offering confirmation. Yet, as they identify effects of migration over short distances, we view this instrument as identifying the effects of relatively local migration. We prefer the years-since-IDs instrument used in the main text as it identifies the network outside the village without strong assumptions about the work locations of current and past migrant residents from the village.

A.3. Evidence on Returns to Education among Migrants from RCRE Villages

To understand how returns in the urban labor market might in influence decisions about enrollment in high school, we use a module from the 2003 round of the RCRE survey that was designed to study the returns to migrant employment. For individuals who had out-migrated from RCRE households in 2003, the RCRE survey collected information on earnings, the cost of migration and the number of days individuals worked as migrants. Using a sample of all adult children of the household head and spouse who were between 15 and 50 years old, we estimate the net returns to education for migrants using a Heckman selection model. Our objective is to assess whether the pattern of returns to education are consistent with the observed decline in high school enrollment in those villages where it became easier to migrate. For the selection equation, we use household land per capita and demographic characteristics (e.g. household size, number of laborers, number of elderly in the household, the household

dependency ratio, the male/female ratio and number of children under 5). On average, we find that an additional year of education has a return of 2.9 percent (Appendix Table A.11, model 1).

To separately estimate the returns to years of schooling for primary and middle school, high school, and post-high school education, we introduce a linear spline in model 2. We find a higher return to primary, middle school and post-secondary years of schooling than to high school. Specifically, we estimate an average return of 4.0 percent to primary and middle school, but only a statistically insignificant 0.3 percent return to a year of high school (Appendix Table A.11, model 2). The return to post-secondary education, consistent with findings Heckman and Li (2004), is higher at 4.7 percent, but also statistically insignificant.

These estimated returns to schooling are much below recent estimates of returns among urban workers, may be biased downward as family member estimates of daily wages of migrants are likely to be imprecisely estimated.[7] Moreover, the selection coefficient on years of high school education is negatively related to whether an individual is a migrant or not, implying that individuals who stay in the village are more likely to go to high school. Returns to high school for rural migrants in urban areas are low, while individuals with a high school education in rural areas are able to qualify for more lucrative positions in village or township government or as managers (or owners) of local enterprises.

References

Antman, Francisca. 2011. "The Intergenerational Effects of Paternal Migration on Schooling and Work: What Can We Learn from Children's Time Allocations?" *Journal of Development Economics*, 96(2): 200-208.

de Brauw, Alan, Jikun Huang, Scott Rozelle, Linxiu Zhang, and Yigang Zhang. 2002. "The Evolution of China's Rural Labor Markets during the Reforms," *Journal of Comparative Economics* 30(2): 329-353.

Heckman, James J. and Xuesong Li. 2004. "Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000," Pacific Economic Review 99(3): 155-71.

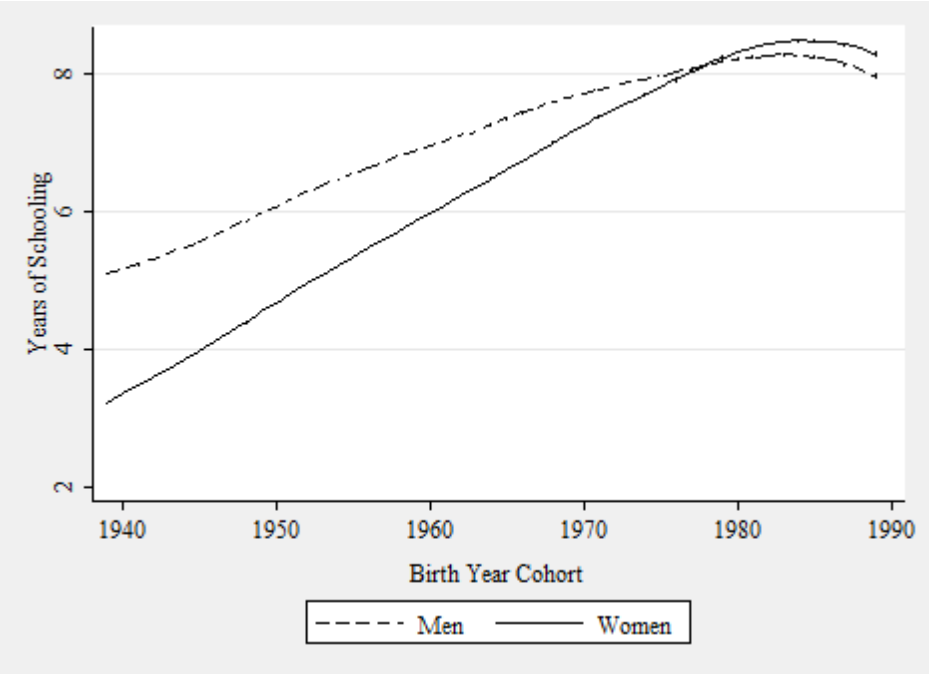
Yang, Dean and Claudia Martinez. 2006. "Remittances and Poverty in Migrants' Home Areas: Evidence from the Philippines," pgs 81-121 in *International Migration, Remittances, and the*

[7]Parents may not know how much of the time children living away from home were working, and may not have precise information on earnings.

Brain Drain, Maurice Schiff and Caglar Ozden (eds.) Trade and Development Series. Washington, D.C.: World Bank; Houndmills, U.K. and New York: Palgrave Macmillan.

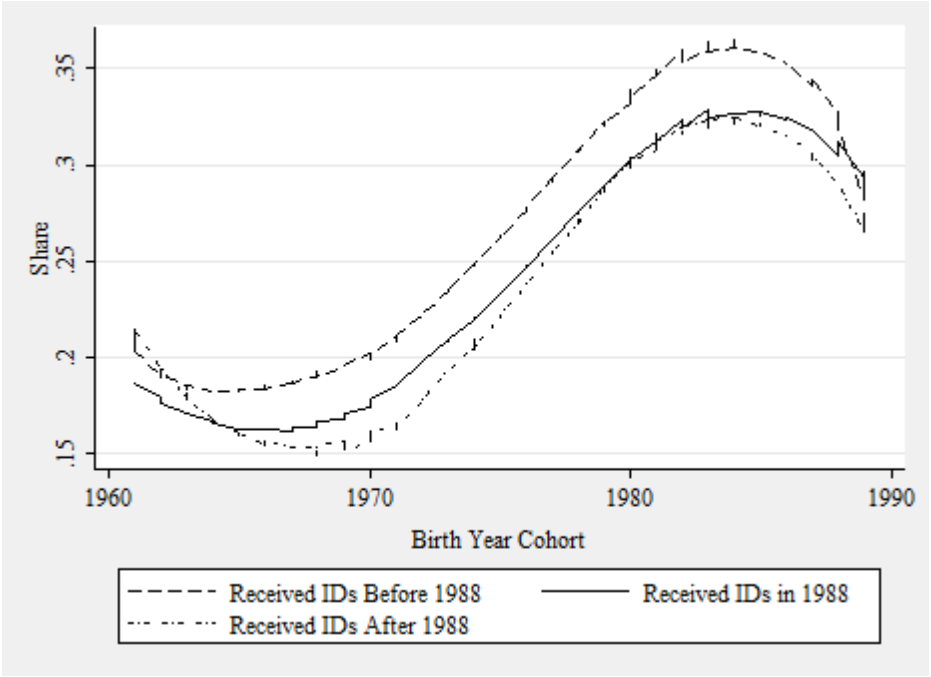
Yang, Dean. 2008. "International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants Exchange Rate Shocks." *The Economic Journal* 118: 591-630.

Figure A.1
Cohort Average Educational Attainment
Lowess Fit



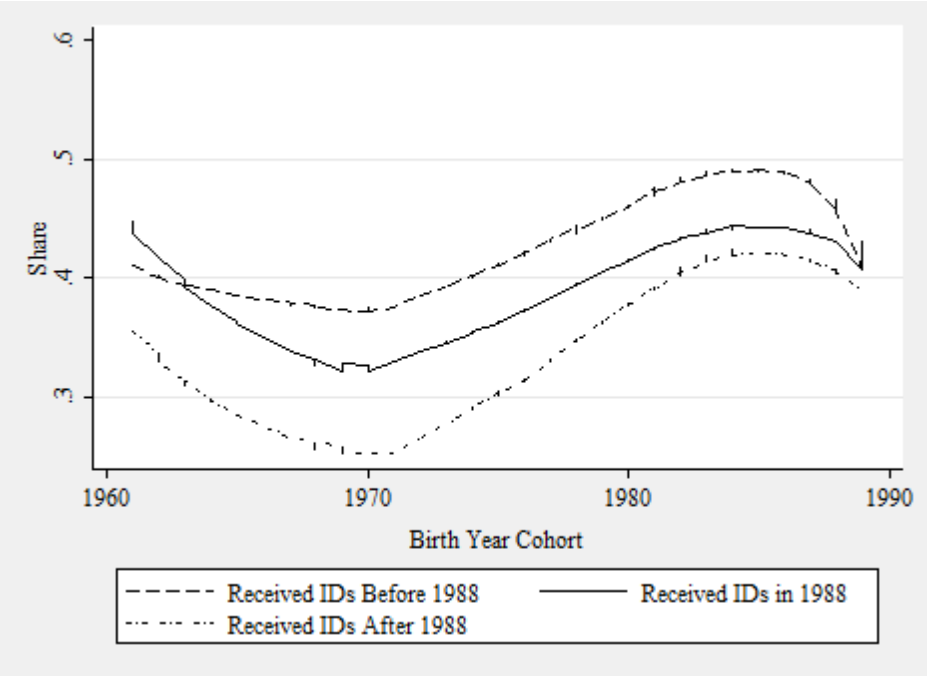
Source: RCRE Supplemental Survey (2004).

Figure A.2
Share of Age Cohort Entering High School
by Timing of ID Card Receipt
Lowess Fit



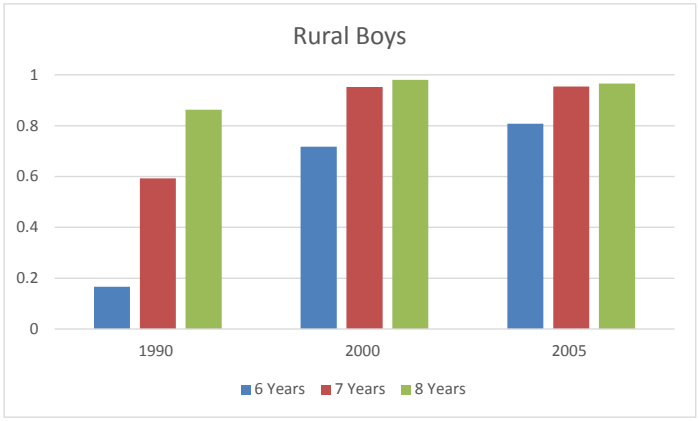
Source: RCRE Supplemental Survey (2004).

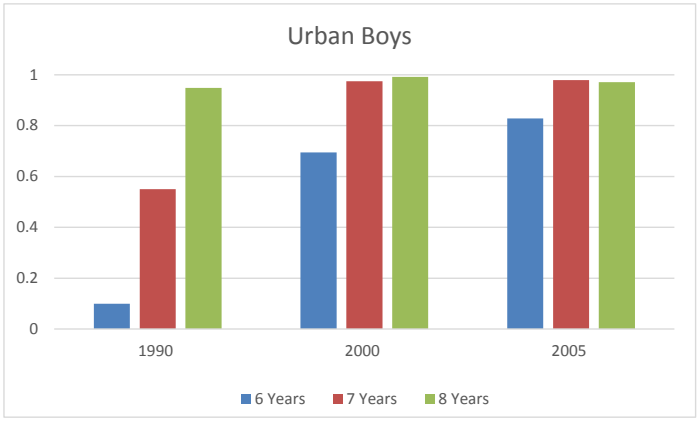
Figure A.3
Share of Middle School Graduates Entering High School
By Timing of ID Card Receipt
Lowess Fit

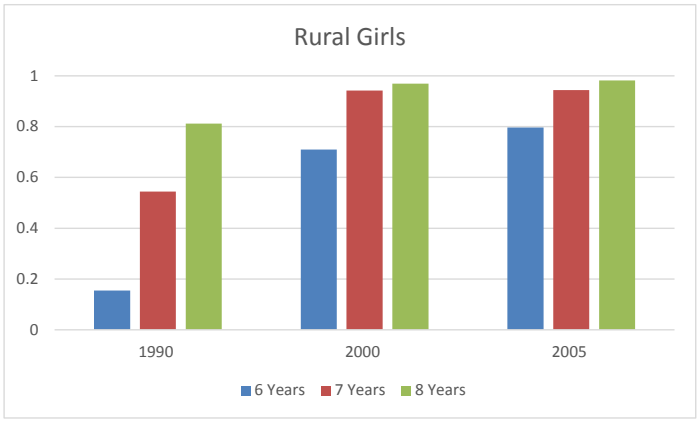


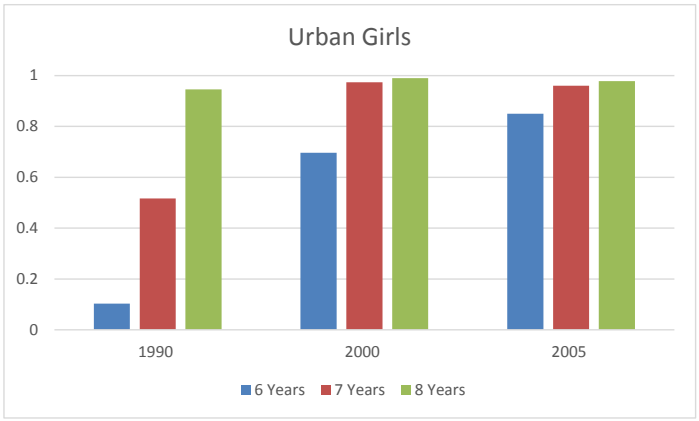
Source: RCRE Supplemental Survey (2004).

Figure A.4
Evidence from the Census on School Enrollment Rates of Young Children Aged 6, 7 and 8







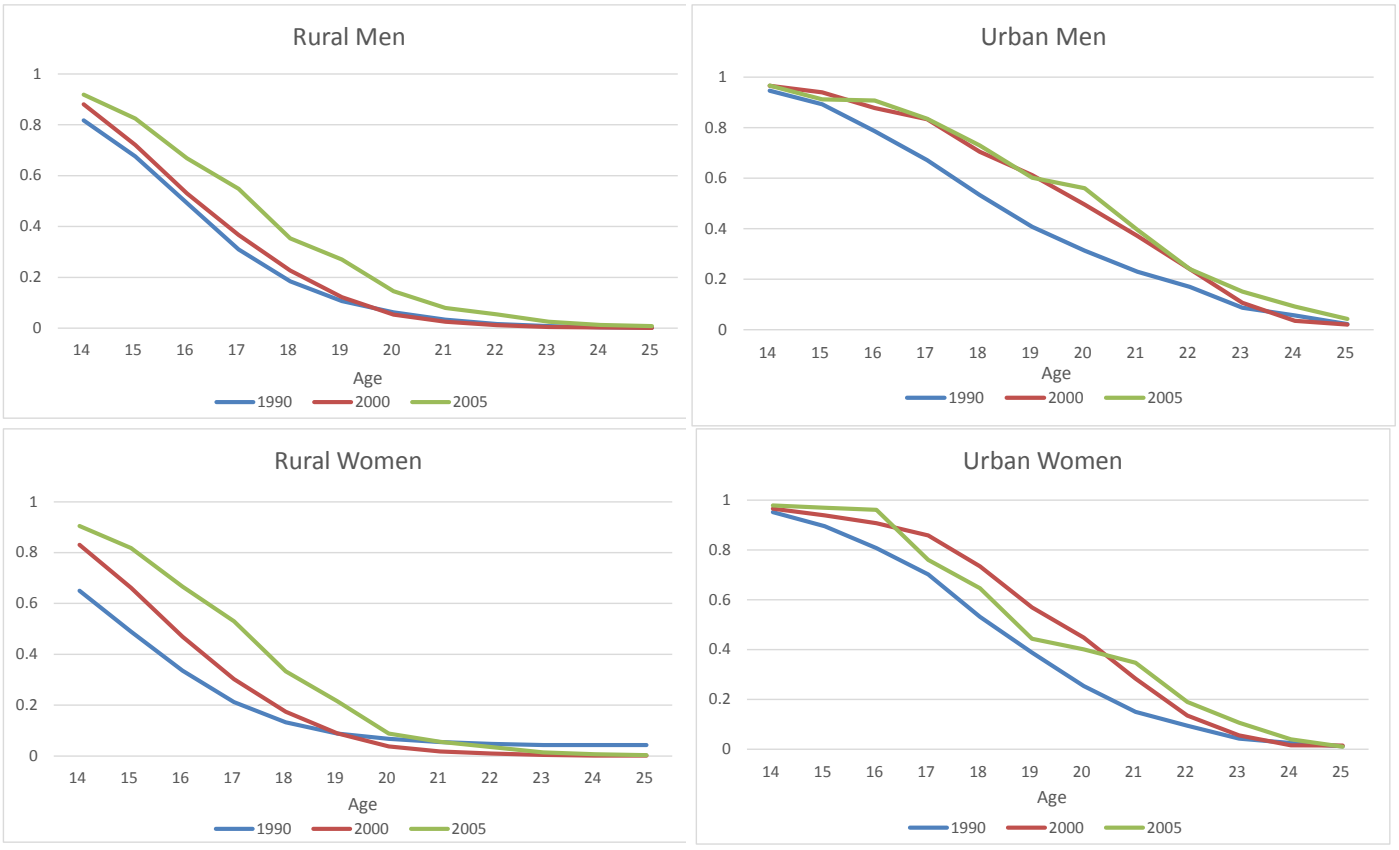


Sources: 1990 and 2000 Population Census and 2005 One Percent Population Sample. National Bureau of Statistics (Beijing).

Figure A.5
Educational Attainment Migrants and Local Urban Workers Aged 22-25
Evidence from Population Census Data on Educational Attainment

Source: Population Census (1990, 2000) and 2005 Population Sample (2005), National Bureau of Statistics (Beijing).

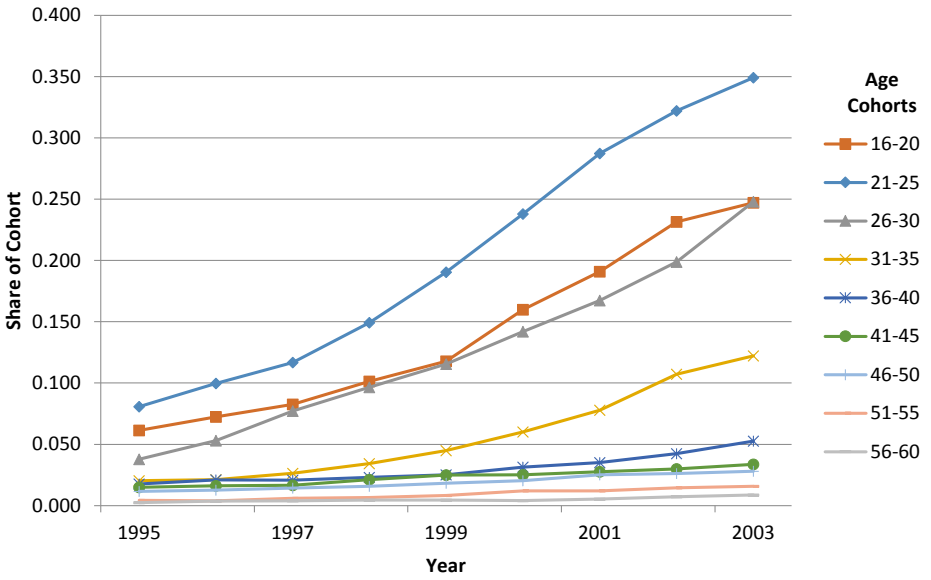
Figure A.6
Educational Enrollment Rates of Teens and Young Adults by Residential Registration (*Hukou*)
Evidence from the Population Census



Note: Enrollment may be in urban or rural areas, these rates are calculated as share of registered rural and urban population enrolled regardless of current residence location.
Source: 1990 and 2000 Population Census and 2005 Population Sample, National Bureau of Statistics (Beijing).

Figure A.7
Evidence on the Age Structure and Working-Age Adults in RCRE Villages

A. Migrant Share of Five-Year Age Cohorts



B. Five-Year Age Group as a Share of the Registered Village Population

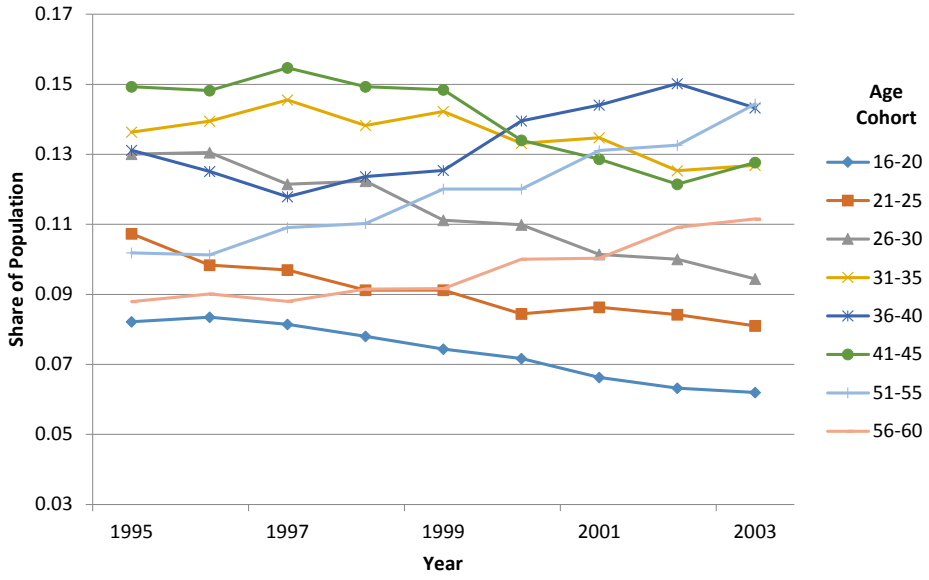


Table A.1
Local Networks of Rural-Urban Migrants at Time of Migration
Five-City China Urban Labor Survey (CULS) Migrant Survey*

	Source Community Location	
	All Provinces	4 RCRE Provinces
Share of Migrants with:		
Job Arranged Before *First* Migration Experience	0.52	0.57
Job Arranged Before *Current* Migration Experience	0.53	0.56
Some Acquaintance from Home Village in City Before Migrating	0.91	0.94
**Close Family Member in City Before Migration	0.35	0.35
**Extended Family Member in City Before Migration	0.52	0.58
**Hometown Acquiantances	0.65	0.67
Five or Fewer Hometown Acquaintances	0.39	0.44
More than Five Hometown Acquaintances	0.27	0.24
At Least One Local Acquaintance	0.09	0.08
Number of Migrants	2,463	481

*Respondents are holders of rural registration (*hukou*). The survey was conducted in Fuzhou, Shanghai, Shenyang, Wuhan and Xian during late 2001. Sample frames were assembled using information on distribution of migrants within cities from the 2000 Population Census. After selecting neighborhoods through a proportional population sampling procedure, sample frames were assembled using residents' committee records of migrant households and registers of migrants living on construction sites and held by local by police stations. Very short-term migrants, who lack a residence that falls under the jurisdiction of either of these authorities, are unlikely to have made it into the sample frame.

**A *close family member* is adult sibling or member of nuclear family (e.g., spouse, child, parent). An *extended family member* refers to cousins or other relatives. *Hometown acquaintances* are unrelated, but known by the respondent. Note that migrants may have acquaintances in several categories, so that subcategories of acquaintances will add to more than 100.

Table A.2
Evidence on Educational Attainment of Migrants
from the China Urban Labor Survey (2001)

	Source Community	
	All Provinces	4 RCRE Provinces
Education		
Elementary or Less	0.247	0.220
Some Middle School	0.086	0.096
Middle School	0.485	0.501
Some High School	0.039	0.045
High School	0.120	0.120
Some Post Secondary	0.009	0.011
College	0.010	0.012
Number of Observations	2,463	481

*Source: China Urban Labor Survey (see discussion on note of Table 1).

Table A.3

Reported Age Starting Primary School

Individuals Age 10 to 34 in 2000

Age	Number	Share
4	6	0.002
5	56	0.021
6	530	0.198
7	1336	0.499
8	639	0.239
9	83	0.031
10	18	0.007
11	6	0.002
12	2	0.001
13	2	0.001
14	1	0.000

Source: China Center for Agricultural Policy (CCAP) Data Set, 2000. See de Brauw et al (2002) for a description of the CCAP survey.

Table A.4
Proportion of Individuals Staying in School
by Grade and Primary School Type

	Six Year Primary Schools		Five Year Primary Schools	
Grade	Proportion	N	Proportion	N
2	1.00	1310	1.00	4193
3	1.00	1310	0.99	4186
4	1.00	1296	0.99	4118
5	0.99	1285	0.98	4019
6	0.98	1257	0.91	3904
7	0.95	1211	0.95	3484
8	0.92	1122	0.87	3238
9	0.90	1011	0.43	2712
10	0.47	877	0.68	1134
11	0.95	388	0.84	729
12	0.91	351	0.62	574
13	0.36	305	0.59	329
14	0.91	100	0.81	183
15	0.83	83	0.58	138
16	0.36	61	0.32	74
17	0.40	20	0.21	24
18	0.80	5	0.50	4
19	0.00	3	0.50	2
20	0.00	0	0.00	1

Notes: Proportions are conditional on school enrollment the previous year.
Assumes children start school at age 7 and do not skip.
Source: RCRE Supplemental Survey (2004).

Table A.5

Average Village Characteristics in 1988

by Timing of ID Card Distribution

		Year ID Cards Were Issued			
		prior to 1988	in 1988	after 1988	p-value[1]
Share of Productive Assets Owned by the Village Collective	mean	0.399	0.260	0.246	0.230
	std. dev	0.278	0.205	0.276	
Mean Consumption Per Capita	mean	414.3	367.7	405.1	0.566
	std. dev	154.6	159.6	86.9	
Mean Income Per Capita	mean	627.2	504.2	558.1	0.278
	std. dev	243.3	213.7	162.4	
Cultivable Share of Total Land Area	mean	0.691	0.546	0.512	0.203
	std. dev	0.277	0.268	0.309	
Share in Mountains	mean	0.14	0.23	0.31	0.589
	std. dev	0.36	0.43	0.48	
Share Near a City	mean	0.21	0.04	0.08	0.341
	std. dev	0.43	0.20	0.28	
Cropped Land Gini Ratio	mean	0.21	0.15	0.17	0.024
	std. dev	0.07	0.06	0.05	
Average Household Size	mean	4.40	4.70	4.68	0.311
	std. dev	0.66	0.48	0.53	
Total Village Land	mean	4508	4633	7676	0.512
	std. dev	4694	4676	9401	
Male Share in Population	mean	0.51	0.51	0.50	0.716
	std. dev	0.02	0.02	0.02	
Share of Labor Force Earning Wage Locally	mean	0.27	0.17	0.16	0.317
	std. dev	0.21	0.19	0.22	
Village Population	mean	1646	1288	1501	0.429
	std. dev	1089	537	925	
Village Consumption Per Capita Gini	mean	0.18	0.16	0.16	0.142
	std. dev	0.03	0.03	0.03	
Village Income Per Capita Gini	mean	0.23	0.22	0.21	0.855
	std. dev	0.07	0.05	0.07	
Average Years of Schooling, aged 18-22	mean	8.25	7.57	7.47	0.186
	std. dev	1.73	1.33	1.21	
Share of 15-18 Year Olds Enrolled in High School	mean	0.35	0.33	0.32	0.861
	std. dev	0.47	0.47	0.47	
Observations		14	25	13	

Notes:

1. p-values in column 4 test the hypothesis that the three means are equal.
2. Consumption and income per capita are reported in 1986 RMB Yuan.
3. Sources: RCRE Household and Village Surveys (1986 to 2003), and RCRE Supplemental Surveys (2004).

Appendix Table A.6

Descriptive Statistics for Children Graduating from Middle School

Selected Variables, for Selected Years

	All Years	Year					
		1987	1990	1993	1996	1999	2002
Individual Level Variables							
Enrolled in High School? (1=yes)	0.43	0.32	0.44	0.41	0.42	0.47	0.47
	(0.50)	(0.47)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)
Gender (1=male)	0.57	0.58	0.60	0.58	0.53	0.53	0.48
	(0.50)	(0.49)	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)
First Born (1=yes)	0.45	0.47	0.49	0.46	0.52	0.50	0.37
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.48)
Birth Order	1.87	1.92	1.94	1.92	1.76	1.71	1.94
	(1.04)	(1.12)	(1.18)	(1.14)	(1.02)	(0.85)	(0.93)
Household Level Variables							
First Born in Household	0.40	0.28	0.44	0.43	0.38	0.42	0.43
was Male (1=yes)	(0.50)	(0.45)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)
Father's Years of Schooling	6.39	5.44	5.52	6.25	6.47	7.27	6.79
	(3.21)	(3.28)	(3.28)	(3.32)	(3.16)	(3.06)	(3.04)
Mother's Years of Schooling	4.22	3.30	3.07	3.67	4.23	5.02	5.14
	(3.30)	(3.03)	(2.99)	(3.22)	(3.36)	(3.17)	(3.32)
Number of Potential Migrants, Household, Male	0.46	0.30	0.47	0.45	0.44	0.45	0.54
	(0.62)	(0.48)	(0.58)	(0.61)	(0.59)	(0.62)	(0.62)
Number of Potential Migrants, Household, Female	0.49	0.25	0.51	0.55	0.52	0.47	0.59
	(0.71)	(0.48)	(0.69)	(0.81)	(0.81)	(0.66)	(0.79)
Village Level Variables							
Share of Migrants, Village Workforce	0.11	0.03	0.02	0.09	0.13	0.14	0.21
	(0.12)	(0.04)	(0.03)	(0.09)	(0.12)	(0.10)	(0.12)
ln(Village Mean Income Per Capita)	6.42	6.20	6.21	6.22	6.52	6.52	6.66
	(0.39)	(0.32)	(0.35)	(0.31)	(0.32)	(0.38)	(0.31)
ln(Village Mean Wealth Per Capita)	8.81	8.62	8.51	8.69	8.89	8.93	9.07
	(0.55)	(0.48)	(0.57)	(0.48)	(0.50)	(0.50)	(0.45)

Appendix Table A.6 Continued on Next Page

Appendix Table A.6 (Continued)

	All Years	Year 1987	1990	1993	1996	1999	2002
Village Total Land (mu)	5090	4820	5080	4870	5200	5100	5760
	(5710)	(5110)	(5190)	(5500)	(5310)	(6240)	(6460)
Village Cropped Land Per Capita Gini	0.21	0.18	0.18	0.20	0.21	0.23	0.26
	(0.08)	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.11)
Village Labor Force	861	780	851	867	899	820	948
	(486)	(352)	(419)	(433)	(470)	(487)	(559)
Years Since IDs Issued	7.37	0.33	2.12	4.86	7.87	11.01	13.86
	(5.11)	(1.02)	(1.65)	(2.07)	(2.20)	(2.23)	(2.28)
Cultivable Share of Village Land	0.58	0.60	0.57	0.62	0.55	0.57	0.54
	(0.28)	(0.26)	(0.28)	(0.28)	(0.28)	(0.29)	(0.31)
Forest Share of Village Land	0.15	0.16	0.16	0.15	0.14	0.16	0.19
	(0.27)	(0.27)	(0.25)	(0.27)	(0.25)	(0.28)	(0.30)
Orchards Share of Village Land	0.04	0.02	0.03	0.04	0.07	0.06	0.06
	(0.07)	(0.03)	(0.05)	(0.06)	(0.10)	(0.10)	(0.09)
Aquaculture Share of Village Land	0.04	0.05	0.05	0.04	0.04	0.04	0.05
	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)
Share of Households with Non-Agricultural Self-Employment Income	0.56	0.66	0.66	0.55	0.57	0.49	0.51
	(0.28)	(0.27)	(0.27)	(0.27)	(0.25)	(0.28)	(0.26)
Quota Share of Grain Produced	0.09	0.12	0.12	0.08	0.10	0.07	0.03
	(0.08)	(0.10)	(0.09)	(0.08)	(0.08)	(0.07)	(0.05)
Scaled Lagged July-November Rainfall Shock, Squared	0.15	0.02	0.02	0.02	0.04	0.02	0.02
	(2.12)	(0.04)	(0.07)	(0.02)	(0.05)	(0.05)	(0.05)
Number of Observations	3167	158	162	237	262	238	187

Notes: The first column includes descriptive statistics for all years; the second through seventh columns include descriptive statistics for selected years.

Sources: RCRE Supplemental Survey (2004), Annual RCRE Household and Village Surveys (1986-1991, 1993, 1995-2003).

Appendix Table A7
What Factors Determine the Share of Village Residents in the Migrant Network?
First-Stage Regression Using the Sample of Individuals Completing Middle School, 1986-2003

	Dependent Variable: Share of Registered Village Residents Working as Migrants				
Model	1	2	3	4	5
Years Since IDs issued	-0.017*	-0.021**	-0.021**	-0.021**	-0.021**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Years Since IDs Issued Squared	0.007***	0.008***	0.008***	0.008***	0.008***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Years Since IDs Issued Cubed	-0.001**	-0.001***	-0.001***	-0.001***	-0.001***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
(Years Since IDs Issued)^4/10	0.002**	0.002***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(Village Average Income Per Capita)		0.020	0.020	0.019	0.019
		(0.015)	(0.015)	(0.015)	(0.015)
Total Land in Village (Mu)		0.0004	0.0004	0.0004	0.0004
		0.0003	0.0003	0.0003	0.0003
Cropped Land Gini Coefficient		-0.013	-0.014	-0.011	-0.011
		(0.087)	(0.086)	(0.087)	(0.086)
Size of Village Workforce(/10)		-0.001***	-0.001***	-0.001***	-0.001***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Cultivable Share of Village Land		-0.029	-0.030	-0.029	-0.029
		(0.038)	(0.038)	(0.038)	(0.038)
Gender (1=male, 0=female)			-0.001	-0.001	-0.001
			(0.002)	(0.002)	(0.002)
First Born? (1=yes, 0=no)			0.002	0.002	0.002
			(0.003)	(0.003)	(0.003)
First Born in Household is Male? (1=yes, 0=no)			-0.003	-0.003	-0.004
			(0.002)	(0.002)	(0.003)
Father's Years of Schooling				-0.001**	-0.001**
				(0.0004)	(0.0004)
Mother's Years of Schooling				0.0002	0.0002
				(0.0005)	(0.0005)
Number of Potential Migrants, Male					-0.001
					(0.002)
Number of Potential Migrants, Female					0.001
					(0.002)
Number of Observations	3160	3157	3157	3157	3157

Notes: Columns 1 through 5 are the first stage of IV regressions shown in models 1 to 5 of Table 11. .

Appendix Table A.8
Relationship Between Various Definitions of GDP Growth Instrument and Share of Migrants from the Village

	No denominator	Linear in Distance	Quadratic in Distance	$(D\text{-}X)^2$
	(1)	(2)	(3)	(4)
GDP Growth	0.019	-0.829	4.786	0.837***
	(0.050)	(0.708)	(3.590)	(0.303)
Lagged GDP	0.021	1.499**	10.27**	0.490**
Growth	(0.049)	(0.738)	(3.95)	(0.199)

Notes: Each cell represents an individual regression, all of which include village fixed effects and province-year fixed effects. Standard errors clustered at city-year level in parentheses. **-indicates significance at the 5 percent level; ***-indicates significance at the 1 percent level. There are 2394 observations in regressions in row 1 and 2173 observations in row 2.

Appendix Table A.9
Determinants of High School Enrollment using Lagged Growth in Nearest City Instrument Conditional on Completing Middle School, 1993-2003

	Dependent Variable: Enroll in High School Next Year = 1				
Model	1	2	3	4	5
	IV-GMM	IV-GMM	IV-GMM	IV-GMM	IV-GMM
Share of Migrants in Village	-2.179**	-2.638**	-2.181	-2.540*	-2.534*
Labor Force	(1.109)	(1.289)	(1.343)	(1.352)	(1.318)
Gender (1=male)			0.036	0.039*	0.038
			(0.023)	(0.023)	(0.023)
First Born (1=yes)			0.052**	0.026	-0.010
			(0.024)	(0.025)	(0.035)
First Born in Household			-0.065***	-0.075***	-0.054**
was Male (1=yes)			(0.022)	(0.022)	(0.024)
Father's Years of Schooling				0.022***	0.022***
				(0.005)	(0.005)
Mother's Years of Schooling				0.023***	0.023***
				(0.005)	(0.005)
Number of Potential Migrants, Household, Male					-0.046*
					(0.024)
Number of Potential Migrants, Household, Female					-0.006
					(0.020)
ln(Village Mean Income Per Capita)		-0.084	-0.080	-0.068	-0.066
		(0.097)	(0.090)	(0.094)	(0.095)
Village Total Land (/100)		0.002	0.002	0.003	0.003
		(0.002)	(0.002)	(0.002)	(0.002)
Village Cultivable Land Per Capita		0.639	0.501	0.514	0.507
Gini		(0.552)	(0.520)	(0.533)	(0.527)
(Village Labor Force)/10		-0.001	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Cultivable Share of Village Land		0.170	0.143	0.110	0.112
		(0.205)	(0.189)	(0.199)	(0.198)
Cluster Corrected F statistic	6.079	4.571	4.533	4.539	4.556
Weak instrument robust Anderson-Rubin F statistic	0.33	0.71	0.71	0.31	0.33
Anderson-Rubin p-value	0.566	0.399	0.399	0.581	0.563
Number of Obs.	2,154	2,154	2,154	2,154	2,154

Notes: In parentheses, we show robust standard errors that allow for arbitrary correlation within city-year observations. All regressions control for factors related to village location with village fixed effects, and macroeconomic shocks using province*year fixed effects. All models are estimated using an instrumental variables-generalized method of moments estimator that is efficient in the presence of presence of heteroskedasticity and arbitrary within city-year cluster correlation (see Wooldridge 2002, page 193). Weak instrument robust Anderson-Rubin F statistics are calculated with a minimum distance estimator suggested by Finlay and Magnusson (2011). *-indicates significance at the 10 percent level; **- indicates significance at the 5 percent level; ***- indicates significance at the 1 percent level.

Appendix Table A.10
Determinants of High School Enrollment using Contemporaneous Growth in Nearest City Instrument
Conditional on Completing Middle School, 1991-2003

	Dependent Variable: Enroll in High School Next Year = 1				
Model	1	2	3	4	5
	IV-GMM	IV-GMM	IV-GMM	IV-GMM	IV-GMM
Share of Migrants in Village	-2.162***	-2.275***	-2.135**	-2.166***	-2.168***
Labor Force	(0.733)	(0.840)	(0.835)	(0.700)	(0.699)
Gender (1=male)			0.026	0.030	0.029
			(0.022)	(0.022)	(0.022)
First Born (1=yes)			0.047**	0.022	-0.011
			(0.023)	(0.023)	(0.033)
First Born in Household			-0.062***	-0.072***	-0.056**
was Male (1=yes)			(0.021)	(0.021)	(0.023)
Father's Years of Schooling				0.020***	0.020***
				(0.005)	(0.005)
Mother's Years of Schooling				0.024***	0.023***
				(0.005)	(0.005)
Number of Potential Migrants, Household, Male					-0.040*
					(0.023)
Number of Potential Migrants, Household, Female					-0.008
					(0.019)
ln(Village Mean Income Per Capita)		0.006	0.006	0.016	0.019
		(0.093)	(0.091)	(0.090)	(0.091)
Village Total Land (/100)		0.001	0.001	0.001	0.001
		(0.002)	(0.002)	(0.002)	(0.002)
Village Cultivable Land Per Capita		0.316	0.272	0.223	0.209
Gini		(0.412)	(0.399)	(0.390)	(0.390)
(Village Labor Force)/10		-0.001	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Cultivable Share of Village Land		0.152	0.147	0.091	0.095
		(0.175)	(0.169)	(0.165)	(0.166)
Cluster Corrected F statistic	7.636	6.626	6.647	6.846	6.840
Weak Instrument Robust Anderson-Rubin F-Statistic	4.55	4.38	3.88	3.89	3.92
Anderson-Rubin p-value	0.033	0.036	0.049	0.049	0.048
Number of Obs.	2,372	2,372	2,372	2,372	2,372

Notes: In parentheses, we show robust standard errors that allow for arbitrary correlation within city-year observations. All regressions control for factors related to village location with village fixed effects, and macroeconomic shocks using province*year fixed effects. All models are estimated using an instrumental variables-generalized method of moments estimator that is efficient in the presence of presence of heteroskedasticity and arbitrary within city-year cluster correlation (see Wooldridge 2002, page 193). Weak instrument robust Anderson-Rubin F statistics are calculated with a minimum distance estimator suggested by Finlay and Magnusson (2011). *-indicates significance at the 10 percent level; **- indicates significance at the 5 percent level; ***- indicates significance at the 1 percent level.

Appendix Table A.11
Returns to Education Among Migrants from RCRE Villages in 2003
Heckman Selection Models

	Model 1		Model 2	
	ln(Daily Migrant Wage)	Migrant? (1= Yes)	ln(Daily Migrant Wage)	Migrant? (1= Yes)
Years of Schooling	0.029	-0.007	--	--
	(0.012)	(0.014)		
0<=Years of Schooling <9	--	--	0.040	0.072
			(0.019)	(0.022)
9<=Years of Schooling <12	--	--	0.003	-0.126
			(0.034)	(0.036)
Years of Schooling>=12	--	--	0.047	-0.029
			(0.059)	(0.068)
Age	0.146	0.285	0.149	0.286
	(0.041)	(0.043)	(0.034)	(0.042)
Age Squared	-0.003	-0.005	-0.002	-0.005
	(0.001)	(0.001)	(0.001)	(0.001)
Male	0.213	0.311	0.217	0.307
	(0.053)	(0.061)	(0.052)	(0.061)
Fathers Years of Education	-0.013	-0.0283	-0.012	-0.027
	(0.010)	(0.011)	(0.009)	(0.011)
Mothers Years of Education	0.008	-0.005	0.008	-0.005
	(0.010)	(0.012)	(0.010)	(0.011)
Household Size	--	-0.010	--	-0.020
		(0.044)		(0.044)
Number of Adult Laborers	--	0.054	--	0.057
		(0.046)		(0.046)
Household Land Per Capita	--	-0.115	--	-0.126
		(0.052)		(0.052)
Number of Elderly in Household	--	-0.054	--	-0.045
		(0.038)		(0.038)
Dependency Ratio	--	-0.079	--	-0.087
		(0.178)		(0.179)
Male/Female Ratio	--	-0.354	--	-0.317
		(0.202)		(0.202)
Number of Children Under 5	--	-0.197	--	-0.203
		(0.068)		(0.069)
Number of Observations	3880		3880	
Censored Observations	3101		2101	
Uncensored Observations	779		779	

Notes: Individual information necessary to estimate daily returns to education from migrant employment are only available for the 2003 survey. We estimate returns to education in migrant employment for children of the household head and spouse who are under 50 and over 15 years of age. Standard errors, clustered at the village, are shown in parentheses.