Community Poverty, Industrialization, and Educational Gender Gaps in Rural China

Hua-Yu Sebastian Cherng
*University of Pennsylvania, cherng@gse.upenn.edu*

Emily C. Hannum
*University of Pennsylvania, hannumem@soc.upenn.edu*

Follow this and additional works at: [http://repository.upenn.edu/gansu_papers](http://repository.upenn.edu/gansu_papers)

Part of the *Education Commons*, and the *Sociology Commons*

Reprinted from *Social forces* (2013): sot084.
Publisher URL: [http://sf.oxfordjournals.org/content/early/2013/07/31/sf.sot084.short](http://sf.oxfordjournals.org/content/early/2013/07/31/sf.sot084.short)

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/gansu_papers/41](http://repository.upenn.edu/gansu_papers/41)

For more information, please contact libraryrepository@pobox.upenn.edu.
Community Poverty, Industrialization, and Educational Gender Gaps in Rural China

Abstract
This paper investigates community impoverishment and industrialization as explanations for educational gender gaps in rural China with analysis of a multi-province household survey and a longitudinal study of youth in one impoverished province. We consider attributes of poor communities that might shape gaps and the related roles of household and community poverty. Three major results emerge from this paper: community impoverishment, not industrialization, correlates with gaps; poverty and isolation shape gaps differently at different educational levels; and girls in relatively wealthy households fare better than boys at the transition to high school. Results suggest the importance of theorizing differences by educational stage and the need for research that conceptualizes the non-local dimensions of industrialization as potential considerations in educational decisions.

Disciplines
Education | Social and Behavioral Sciences | Sociology

Comments
Reprinted from Social forces (2013): sot084.

Publisher URL:http://sf.oxfordjournals.org/content/early/2013/07/31/sf.sot084.short

This journal article is available at ScholarlyCommons: http://repository.upenn.edu/gansu_papers/41
Community Poverty, Industrialization, and Educational Gender Gaps in Rural China

May 1, 2013

This paper investigates community impoverishment and industrialization as explanations for educational gender gaps in rural China with analysis of a multi-province household survey and a longitudinal study of youth in one impoverished province. We consider attributes of poor communities that might shape gaps and the related roles of household and community poverty.

Three major results emerge from this paper: community impoverishment, not industrialization, correlates with gaps; poverty and isolation shape gaps differently at different educational levels; and girls in relatively wealthy households fare better than boys at the transition to high school. Results suggest the importance of theorizing differences by educational stage and the need for research that conceptualizes the non-local dimensions of industrialization as potential considerations in educational decisions.
INTRODUCTION

Across the world, one of the most consistent findings in educational research has been the declining disadvantage for girls, and, indeed, that a girls’ advantage has emerged in many settings (e.g., Blossfeld and Shavit 1993; Dorius and Firebaugh 2010; Fiske 2012; Grant and Behrman 2010). Globally, gender gaps are concentrated in certain regions, and among the poorest households in those regions (Filmer 2005)¹. A secular trend of declining gender inequality in education is also apparent in China (Filmer 2005; Hannum and Xie 1994; Hannum 2005). Some recent evidence raises questions about the degree to which gaps favoring boys persist at all in some areas of China, and at some educational transitions, including the transition to higher education (Hannum et al. 2010; Tsui and Rich 2002; Wu and Zhang 2010). Also paralleling research on the global concentration of gender disparities in certain regions, research in China indicates considerable geographic variability in the degree of gender disparity, with girls’ disadvantage concentrated in rural areas, which are much poorer than urban areas, and in poor households (Connelly and Zheng 2007a; Connelly and Zheng 2007b; Hannum 2005; Hannum et al. 2009; Hannum et al. 2010; Wu and Zhang 2010).

We know little, however, about why gender gaps persist in some poor communities: is the source rooted more in labor market explanations and incentives these contexts provide to parents, or in dimensions of community poverty²? Do the effects of community poverty operate through household economic conditions, and do attributes of community disadvantage associated with community poverty matter?

¹The regions cited are South Asia and North, Western, and Central Africa (Filmer 2005). This study was based on analysis of the Demographic and Health Surveys, which do not cover China.
²For the sake of brevity, we will refer to the “traditional” gender gap – girls’ disadvantage – as “gender gap” in the rest of this paper.
This paper investigates the state of gender disparities in education in rural areas. First, to test the community correlates of gender disparities, we analyze a multi-province rural household income survey from 2002, the Chinese Household Income Project (CHIP) (Shi, Chinese Household Income Project, 2002). Second, to investigate how particular dimensions of poor communities relate to gender gaps, we analyze a single-province case study of one of China’s poorest provinces and one that has persistent gender disparities. For this purpose, we use the Gansu Survey of Children and Families (GSCF), a longitudinal survey of rural children that contains extensive measures of the household and community conditions of the childhood home of respondents (University of Pennsylvania 2010).

**Community Poverty and Gender Gaps in Education**

*Cross-National Perspectives*

A substantial body of research has examined community effects on education in contexts across the world (Ainsworth 2002; Crane 1991; Garner and Raudenbush 1991; Harding et al. 2010; Kling, Liebman, and Katz 2005). One line of this research investigates how dimensions of community context are associated with gender gaps in educational outcomes. For example, one mixed-method study argues that the origins of gender gaps in math among high-scoring middle schoolers in the United States may lie in access to neighborhood academic programs (Entwisle, Alexander, and Olson 1994). Another line of research investigates how community poverty influences educational outcomes, more broadly (Brown and Park 2002; Connell 1994; Brooks-Gunn, Duncan, and Aber 1997; Hannum 2003).

---

3 Using a case study of a US metropolitan area, the authors found that girls spent less time in their community environment than boys, and as a consequence, were less able to access community resources that could foster academic success. As a result, the authors argue, high-scoring boys out-scored their female counterparts.
However, few sociological studies have investigated how community poverty, particularly in middle and low income countries, is linked to educational gender gaps.

One dimension of communities that has been linked theoretically to gender gaps in developing country contexts is the degree of industrialization of labor markets. Empirical work on trends in educational gender gaps has explicitly or implicitly explained trends with a modernization, industrialization, or development framework, which suggests that the shift of the work force out of agriculture and into industry, or the development of a “modern” economy, decreases girls’ disadvantage in education. This framework suggests that as patriarchal societies move out of agriculture and into industry, old cultural or family-based incentives for prioritizing boys over girls will decline (Forsythe, Korzeniewicz, Durrant 2000; Treiman 1970, but see Hannum 2005 for inconsistencies with the Chinese case).

Of course, many scholars have contested modernization theory over the years, starting most notably with the work of Ester Boserup, who argued that men were best positioned initially to take advantage of new opportunities for work posed by industrialization and development (Boserup 1970; 1989). Emerging labor markets under industrialization and development might reinforce incentives to families for investments in boys, at least in the short term (Boserup 1970; 1989; Boserup et al. 2007; Forsythe et al. 2000a; Forsythe et al. 2000b).

Although communities that experienced movement out of agriculture and into industry are likely to be less poor than those that remain primarily agricultural, a modernization or industrialization framework does not directly address the issue of whether dimensions of community poverty, beyond effects via industrialization, matter
for gender inequality. This omission is potentially important, as economists studying developing countries have found, in many national settings, that community poverty is associated with educational gender gaps (Colclough, Rose, and Tembon 1999; Filmer 2000; World Bank 1999).

Community poverty is commonly assumed to shape gender disparities through household decision-making processes. The household economy framework, or household decision-making framework, common in economics, suggests that parents make decisions about human capital investments in their children based on perceived monetary costs and benefits associated with educating children (for recent articulations of this framework, see Filmer 2005; Huisman and Smits 2009). These rational calculations are assumed to differ for sons and daughters, for example, if parents expect daughters to be less likely to earn returns on their education, or if those returns are deemed less likely to flow to the natal family due to traditions of co-residence with sons (Filmer 2005; Huisman and Smits 2009).

Assuming that parents with sufficient resources attempt to provide education for all of their children, one might expect this calculation to disadvantage girls most notably among poor families, for whom resources to invest in children are constrained. To the extent that poor communities have higher numbers of the poorest households than other communities, the effect of community poverty may simply reflect a preponderance of the poorest households, where girls are educationally disadvantaged.

However, it is also possible that girls in general, not just girls in the poorest households, are disadvantaged relative to boys in poor communities. This pattern could emerge if households are responding not just to their own resource constraints, but also to
elements of the broader community context that shape incentives and risks for investing in boys and girls. For example, as implied by the industrialization framework, the presence or absence of local employment opportunities outside of agriculture may create or impede incentives to educate girls, relative to boys.

Research has highlighted a number of other characteristics of poor communities that matter for educational access, and may matter for gender disparities. For example, scholars have argued that community isolation influences overall enrollment for boys and girls, but may disadvantage girls more than boys if parents are unwilling to send daughters to far away schools due to safety concerns (Colclough, Rose, Tembon 1999; Jacobs 1996, UNESCO 1998). An investigation of determinants of primary school enrollment in 30 developing countries found that when children have to travel longer distances to school, they are less likely to go to school (Huisman and Smits 2008). In Turkey, Tansel (2002) found that the distance between a community and a regional metro center was associated with lower probabilities of school enrollment for boys and girls, but that the disadvantage for girls was greater.

Community infrastructure has also been shown to be associated with enrollment, and could be linked to gender gaps in some contexts. In one study of rural Bangladesh, rural electrification was associated with increased school participation for all rural children (Khandker 1996). One study using household surveys found that Pakistani women who resided in villages with convenient access to water spent less time gathering water each day and more time employed in these villages than women in villages with weaker infrastructure (Ilahi and Grimard 2000). To the extent that girls are also
responsible for domestic tasks, it may be the case that girls who reside in villages with stronger infrastructure spend less time on household chores and more time in school.

Community support for education could also benefit students and potentially narrow gender gaps. Development research has highlighted the importance of community involvement for improving school enrollment rates. One review of World Bank projects designed to improve access to quality primary education in sub-Saharan Africa argued that community support is vital for successful interventions, and effective programs often incorporate the viewpoints of community members (Heneveld and Craig 1996). Collaborations between project and community leaders also strengthen the community's commitment to education, which in turn increases enrollment rates.

Community leaders and members who support the education of youth in their communities may also share progressive views of girls’ education. Evaluations of USAID projects in Mali found that efforts to increase community participation in education, both by financial contributions and involvement in school planning, led to marked increases in girls’ enrollment and entry into primary school (Rugh 2000; Tietjen 1999). Although this study investigated girls’ enrollment and not gender gaps, per se, it is likely that villages that share norms that encourage girls’ education may also have smaller gender gaps.

Research in China

Many studies attest to secular declines in girls’ educational disadvantage in China (e.g., Hannum and Xie 1994; Hannum 2005; Hannum et al. 2010; Wu and Zhang 2010; Zhou et al. 1998). However, temporal patterns in the decline are not very consistent with
a story of development or industrialization: much of the narrowing of the gap emerged during the Cultural Revolution (Hannum and Xie 1994; Hannum 2005).

In rural areas, where the majority of the population resides, incentives persist for son preference in educational investments. In most rural areas of China, sons are the primary source of old age support, and for this reason, parents face strong incentives to invest in sons as long-term insurance. Research in a rural county in Yunnan province indicates that expectations of support from sons are more pronounced among mothers in poorer, more remote rural areas (Li and Lavelly 2003). In rural Gansu Province, which is one of China’s poorest provinces, only about 15 percent of mothers did not expect to get support from children, and the vast majority of those who did expected to get support from sons (Hannum et al. 2009). Further, expectations of future support were tied to aspirations for children’s education, which, in turn, were predictive of children’s subsequent school persistence (Zhang et al. 2007).

However, recent research among families of primary and junior high school age children in Gansu Province has suggested few overt signs that families were substantially biased toward boys in their educational attitudes or practices, though modest gender gaps did exist in parental aspirations for children and years of schooling attained by children (Hannum et al. 2009). The same study found no evidence of performance differences between boys and girls.

There is some evidence linking poverty to gender inequality. Song et al. (2006) report that female schooling is a “luxury good;” it is more income elastic than male schooling. Consistent with this viewpoint, other work with older data has suggested that
higher income households make more egalitarian decisions about enrolling sons and daughters in school (Hannum 2003; 2005).

A small body of literature has also explored the links between community poverty on the education of boys and girls in rural China. Prior scholarship investigates how village resources condition educational opportunities for both boys and girls in rural China (Adams 2006; Adams and Hannum 2005; Knight, Li, Deng 2009). Knight et al. (2009) and Adams and Hannum (2005) find that county and village economic resources are strongly associated with school enrollment, and Adams (2006) finds that average village income is also associated with higher test scores.

However, little research focuses on community poverty and gender gaps. One study using older data found that girls' enrollment in schools benefited more than boys’ in terms of enrollment from living in villages with higher average income (Hannum 2003). Scant research explores the implications of particular characteristics of poor communities, such as labor market opportunities, isolation, infrastructure, or support of education, for gender gaps in China.

Further, little research has considered whether gender disparities in access associated with community and household wealth operate differently at different educational stages. In China, compulsory education encompasses nine years of education, which includes primary and middle schools (Ministry of Education 1986). High schools in rural China serve more villages and charge much higher fees than middle schools, and scholars have argued that higher fees influence rural girls’ disadvantage at
this transition (Wu and Zhang 2010). Therefore, household economic constraints may matter more at the transition to high school than at compulsory stages.

Summary and Research Questions

Overall, research on gender gaps points to a long-term decline in female disadvantage in many global contexts. Prior work has shown that there are still persistent gender gaps – for example, in parts of rural China – and there is some evidence that pockets of female disadvantage are associated with poverty. However, little research has systematically explored the link between female disadvantage with degree of industrialization—emerging non-agricultural labor markets—or with other aspects of community poverty. It also remains unclear whether the link between community poverty and gender gaps simply reflects the aggregated effect of girls’ disadvantage in poorer households, and whether the story may differ by educational stage.

In this paper, using the case of educational gender gaps in rural China, we address four questions:

1. *Is the gender gap smaller in rural communities that have more industrialized (greater non-agricultural) labor market structures?*

2. *Is the gender gap greater in resource-deprived rural communities?*

3. *How do community and household poverty together affect gender inequality?*
   
   3a. *Does the effect of community poverty on gender inequality persist once we control for gender interactions with household poverty?*

4. *What other attributes of poor communities are linked to gender inequality?*
   
   4a. *Do other attributes of poor communities explain the effect of community poverty on gender inequality?*

---

4 In both datasets used in the analyses of this paper, there is no clear correlation between village wealth and the percent of labor force in non-agriculture: the correlations in the CHIP and GSCF datasets are 0.07 and 0.04, respectively. Please see Appendix Table D for more information.
DATA AND MEASURES

2002 Rural Chinese Household Income Project (CHIP) Data

We employ data from the rural sample of the 2002 Chinese Household Income Project (CHIP) to provide descriptive context on the scale of gender inequality in education in rural China, and to address our first two research questions. The 2002 CHIP rural sample covers 22 provincial level administrative units of China: Beijing, Hebei, Shanxi, Liaoning, Jilin, Jiangsu, Zhejiang, Anhui, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu and Xinjiang. Sampled households are located in 961 villages located in 120 different counties (Gustafsson and Ding 2004, p. 5). In addition to household questionnaires, village questionnaires were administered to cadres.

Analytic Samples

We employ two samples from the CHIP dataset: one sample includes individuals ages 15 and older, and the second sample is a subsample of individual ages 15 to 17. The sample of individuals ages 15 and older includes 29,686 respondents, and analyses using

---

5 Gustafsson and Ding (2008, p. 7) describe the sample for the 2002 rural CHIP as follows “The sample was drawn from the large sample used by [the National Bureau of Statistics] in its annual household survey covering around 67,000 households. This sample is selected in a multi-stage procedure to be representative at the province level and each province statistical bureau is responsible for samples at the village level. At the village level a probability sample of ten households is selected. The rural households are asked to keep detailed records of their expenditures as well as provide information on their income. A large number of assistant enumerators aid the households in keeping good accounts and in checking the information.” Detailed sampling documentation for the annual household survey and the original 1988 CHIP subsample drawn from that source can be found in documentation provided with the data created by Griffin and Zhao (1988). There are some differences in the sampling procedures reported for the 1988 survey by Griffin and Zhao and the 2002 data, as described above.
this sample provide contextual information on rural China’s past educational expansion and change in educational gender gaps. Analyses using the subsample of individuals ages 15 to 17 investigate whether youth reach a middle school level of education. We use this sample for two reasons. First, it is reasonable to expect that most individuals who would eventually obtain a middle school education would have at least begun middle school by age 15. Moreover, individuals in this age range are still residing in homes with their parents and have not left to form their own households\(^6\). For this reason, we are able to include in our analyses information on the childhood household conditions that shape respondents’ educational experiences. The total number of respondents ages 15 to 17 is 2,826\(^7\).

However, with this sample, we are only able to analyze middle school access, and not middle school completion or later educational transitions. In the CHIP dataset, 67 percent of 15 to 17 year olds were still enrolled in middle school, and it is reasonable to expect that many of these students might continue beyond middle school. Working with an older sample, for whom we could expect transitions to upper secondary to be more complete, would mean facing bias associated with youth no longer residing in natal households and not being linkable to natal household socioeconomic characteristics.

**Dependent Variables**\(^8\)

---

\(^6\) 96.5 percent of 15 to 17 year olds in the CHIP dataset are children of the head of household.

\(^7\) 1.2 percent of 15 to 17 year olds in the CHIP dataset have missing values on their own educational level.

\(^8\) Appendix Table A provides descriptive statistics for each of the variables using the analytic sample for each model.
We use two educational outcome variables. The first measure is *years of education* completed. The second variable is *reaching middle school*, which measures whether an individual reports a middle school level of education or higher (coded 0 = has not reached middle school, 1 = has reached middle school). Individuals who report reaching middle school include respondents who have attained a middle school level of education and have continued onto higher levels of education, completed middle school and did not continue in their education, are currently enrolled in middle school, or attended middle school at some point but are not currently enrolled in school.

**Independent Variables**

A key analytic variable in our analyses is *female* (coded 0=male, 1=female).

As a proxy for *household wealth*, we use the estimated total value of durable goods and furniture in the household\(^9\). From this variable, we create a categorical variable for *wealth quintile* to account for potential non-linearities and to offer more intuitive interpretations of findings. For the categorical variable, we contrast the middle categories as a group with the top and bottom categories (0 = household is in the middle three wealth quintiles, 1 = household is in the lowest quintile of household wealth, 2 = household is in the highest quintile of household wealth).

To represent the degree to which the local labor market is industrialized, we use a measure of the percent of the village workforce involved in non-agricultural industries:

---

\(^9\)Durable goods include color televisions, black and white televisions, bicycles, motorcycles, refrigerators, washing machines, stereo sets/recorders, VCRs / VCD / DVD, air-conditioners, automobiles.
percent non-agricultural labor force in village. Another village variable represents the prosperity of the village. Household wealth is aggregated to the village level to create a measure of village wealth: village is in lowest quintile of village wealth (coded = 0 village is not in the lowest village wealth quintile, 1 = village is in the lowest village wealth quintile).

All models control for age. One specification represents age group (grouped into 10-year age categories: 25-34 year olds, 35-44 year olds, 55-64 year olds, 65-74 year olds, with 15 – 24 year olds serving as the reference category). We also incorporate an age variable in years into some models.

*Gansu Survey of Children and Families (GSCF) Data*

Per Capital Annual Income”). Gansu has one of the highest incidences of rural poverty among provinces in China, and Gansu’s population lags significantly by educational indicators (China Yearly Provincial Macro-economy Statistics 2011: Region: Gansu, Year: 2009, “Basic Statistics for Education”; Wang 2004). Moreover, in 2000, analyses of census data showed that Gansu was one of three provinces in China where the difference in rural boys’ and girls’ percent attending middle school was 10 percent or more (Connelly and Zheng 2007, Table 4).

The project began with 2000 children ages 9 to 12 in the year 2000, and followed them in 2004, 2007, and 2009. The sample was drawn using a multi-stage, clustered design with random selection procedures employed at each stage (county, township, village, and child). At the final stage, children were sampled from birth records for the full cohort of 9 to 12 year-old children in 100 selected villages. Of these children, 98.9 percent were currently enrolled in school in 2000. The sample included only rural villages, and did not include urban cities or townships.

The GSCF data is well suited to complement our analyses of the CHIP data because our last two research questions focus on community poverty and the GSCF contains data on poor rural communities in one of China’s poorest provinces. Moreover, the GSCF contains household information on the childhood residence of the sample individuals, and follows those children as they grow up. Therefore, we can include older children who may have left their homes because they can be linked with natal household conditions, which can ameliorate selection bias. Older children are also more likely to have completed more educational transitions, and we are able to analyze the transition to
high school using the data from the GSCF. Here, we make use of the 2004 and 2009
data, with variables taken from the 2009 data except where noted.

**Dependent Variables**

To measure educational attainment using the GSCF dataset, we employ two
variables. The first educational outcome variable is whether students *graduate from*
*middle school* (coded 0 = the student did not graduate from middle school, 1 = the student
graduated from middle school). The second dependent variable is whether students
*transition to high school*, among middle school graduates (coded 0 = student did not
transition to high school, 1 = student transitioned to high school).

**Independent Variables**

The main analytic variable in our analyses is *female* (coded 0 = male, 1 = female).

The *household wealth* variable, measured in yuan, is taken from the 2004 survey,
and represents the total value of a household’s house, fixed assets, and durable goods\(^\text{10}\). The 2004 measure is appropriate because later measurements reflect household wealth
after educational transitions occurred. In modeling, we created a categorical variable for
*wealth quintile*. We contrast the middle categories as a group with the top and bottom
categories, in these recoded variables (0 = household is in the middle three wealth

---

\(^{10}\) Fixed assets include vehicles, tractors, push carts, animal-drawn carts, windmills, sprayers, threshers,
hullers, pumps, electronic motors, diesel motors, processing equipment, boats, stone mills, construction
material, and stone rollers. Durable goods include sewing machines, radios, tape recorders, televisions,
bicycles, cameras, clocks and watches, washing machines, refrigerators, electric fans, motorcycles, jewelry,
beds, furniture, telephones, VCD or DVD players, and stereo component systems.
quintiles, 1 = household is in the lowest quintile of household wealth, 2 = household is in the highest quintile of household wealth).

To measure the degree of industrialization of local labor markets, we employ a variable that represents the percentage of households engaged in non-agricultural economic activities: *percent nonagricultural households in village*. Household wealth was aggregated to the village level to create a measure of village wealth: *village is in lowest quintile of village wealth* (coded 0 = village is in the four highest village wealth quintiles, 1 = village is in the lowest village wealth quintile).

We also investigate the role of attributes of poor communities in shaping gender gaps. We use measures of four community attributes, which include village economic opportunity, isolation, infrastructure, and community support\(^{11}\). *Economic opportunity* is a binary variable that indicates whether a village has a county, township, or private household enterprise (0 = no enterprise, 1 = one more enterprises). *Isolation* is an average of the distance, measured in kilometers, the village is from the town seat. *Infrastructure* is the percentage of households with access to water and electricity. *Community support for education* is represented with two binary measures: one is whether the village donates money to schools, and the other is whether villagers donate money to schools. All village variables are taken from the 2004 survey.

*Age* is measured in years.

\(^{11}\) Variables representing dimensions of community poverty are not highly correlated with each other, and correlation coefficients between any two variables are below 0.18.
Analytic Strategy

To illustrate the context of this study, which is one of educational expansion and decline in female disadvantage, we use linear regression models estimating years of education among individuals of all ages to show change in educational patterns across rural cohorts using CHIP data. We include a main effects model to show the effects of educational expansion over the past four decades, and a model that includes interaction terms between gender and cohort to show the decrease in gender gaps across cohorts. We also include a figure that shows rural gender gaps in years of education for young adult and all adult cohorts by province.

To address our first two research questions, which ask whether gender gaps are linked to industrialization or community poverty, we use logit models with random effects at the village level for both the CHIP and GSCF datasets to investigate the main effects of gender, community industrialization, and community poverty on educational outcomes. To examine the association of community industrialization and poverty with gender gaps, we use interaction terms between each community measure and gender.

To address our final two research questions, which examine how community poverty, household poverty, and attributes of poor communities affect gender inequality, we utilize the GSCF data. We employ two-level logit models of educational outcomes and include interaction terms between gender and, in turn, household poverty, community poverty, and other community attributes associated with community poverty. To investigate whether greater gender gaps in the poorest communities can be explained by accounting for specific dimensions of community poverty, we estimate models that include main effects and gender interaction effects of community poverty, employment
opportunity, isolation, infrastructure, and support of education, with controls for the
effect of household wealth on gaps.

ANALYSES

Rural China Context

[Table 1 here]

As noted in the introduction, the context of this study is one of substantial
expansion in education in China, and narrowing of gender gaps. Analysis of the CHIP
data show patterns of educational change across rural cohorts that are consistent with
these trends. Table 1 shows models of years of schooling attained among youth and
young adults, with coefficients for gender and age cohort (model 1) and for all gender-
by-cohort interactions (model 2). Model 1 speaks to the dramatic expansion of education
over the past forty years. Those in the oldest cohort (ages 65 and higher) have 5.9 fewer
years of education than the youngest cohort of 15 to 24 year-olds. Model 1 also shows,
overall, that there is a 1.3-year disadvantage for women in years of education. Model 2,
with interactions that allow the gender penalty to vary by age cohort, shows that the
expansion in education has been especially strong for women. Women aged 65 to 74
have about a 2.8-year penalty in years of schooling, compared to their male
counterparts\textsuperscript{12}. The gap is progressively

\textsuperscript{12} -2.466 years + -0.376 years
smaller, though still statistically significant, with each younger cohort, to a penalty of less than half of a year for those in the youngest (reference) cohort.

[Figure 1 here]

The gender gap varies across China. Figure 1 shows rural gender gaps in years of education attained for all adults and for young adult cohorts by province, in decreasing order of provincial per capita GDP, in 2002 (China Yearly Provincial Macro-economy Statistics 2002). Like Table 1, Figure 1 speaks to the dramatic narrowing of the gender gap with younger cohorts across all provinces included in the CHIP. However, among young adult cohorts, there persists considerable variability in the scope of gender disparity. A number of provinces display trivial gender differences. In some cases – Beijing, Chongqing, and Xinjiang – female young adults in the CHIP sample actually have more average years of schooling than males, although in none of these cases of a reverse gender gap is the difference statistically significant.

At the same time, a number of provinces still have substantial female disadvantages: Gansu, Guizhou, Jiangxi and Anhui still have gender gaps of greater than or close to one year among young adult cohorts. In 2009, these four provinces also have lower than average per capita annual rural household income, with Gansu being the second poorest province (China Yearly Provincial Macro-economy Statistics 2011).

**Industrialization, Community Poverty, and Gender Gaps**

Are gender gaps in education more closely associated with movement out of agriculture or with community poverty? Table 2 addresses this question by presenting models of reaching middle school attainment in rural China among 15 to 17 year-olds,
with a focus on how the gender gap differs in more and less agricultural communities and richer and poorer communities.

In Table 2, models in Panel A focus on the percentage of laborers in the community employed in the non-agricultural labor force and models in Panel B focus on whether the village is in the lowest wealth quintile. Models 1 to 5 include age (all models), interactions with gender (Models 3 to 5), and controls for household poverty (Models 2, 4 and 5) and the interaction between gender and household poverty (Model 5). Model 6 includes all the variables in specifications (A) and (B).

Two main results emerge. First, while results suggest that the degree of transition from agriculture to industry is associated with slightly higher probabilities of reaching middle school, there is no evidence that residence in more industrialized communities is associated with less gender disparity. Specifically, in Models 3a through 5a, no interaction terms with gender and percent non-agricultural labor force in village are significant. These findings stand after accounting for household poverty in Models 4a and 5a. Second, community poverty is associated with greater gender inequality. Models 3b through 6 show significant interactions between gender and residing in a village in the lowest wealth quintile, and these community effects of poverty persist with the main effects of household wealth and the interaction between household wealth and gender, which are not significant in any model. In short, results are more consistent with a community resource deprivation story than an industrialization one.

[Table 2 here]

13 Anonymous reviewers suggested we include controls for the number of male siblings in these models, as well as all models found in this paper. Results are robust to the inclusion of this variable.
Community Poverty, Household Poverty, and Attributes of Poor Communities

In the previous analysis, we find evidence of a relationship between community poverty and exacerbated gender gaps in rural China. We next investigate the dimensions of community poverty that matter using a case study of one of China’s poorest provinces: Gansu Province. As shown in Figure 1, rural Gansu had a persistent gender gap in years of education among young adults in the rural CHIP data.

We begin by confirming in rural Gansu that community poverty, but not the degree of labor market industrialization, correlates with gender disparity. Table 3 presents coefficients from logistic regression models estimating two stages of education: middle school graduation and transition to high school among middle school graduates. The table has two panels: Specifications in Panel A contain an independent variable representing the percentage of laborers in the community employed in non-agricultural labor force, and specifications in Panel B contain a variable representing whether the village is in the lowest village wealth quintile. Each panel contains six model specifications. Models 1 through 3 estimate middle school graduation. Model 1 includes main effects of gender, household wealth (whether the household is in the lowest or highest wealth quintile, with middle wealth quintiles as the reference category), and the percent of households in the village engaging in non-agricultural industry. Model 2 introduces an interaction term between gender and the percent of non-agricultural households in village. Model 3 includes an additional interaction between gender and household wealth. Models 4 through 6 are analogous in modeling strategy to the previous three models, but estimate transition to high school. All models include controls for age.
Similar to findings in Table 2 based on the multi-province rural CHIP data, results from rural Gansu in Table 3 offer no evidence consistent with an association between village industrialization and narrowing gender gaps, but do suggest exacerbated female disadvantage in the poorest villages. Girls in poor villages have disproportionately lower odds of graduating from middle school (Panel B, Models 2 and 3). The main effects of residing in the lowest quintile of households are significant, indicating a disadvantage in poor households for all children (Panels A and B, Models 1 through 3). However, the gender interaction with household poverty is not significant at this stage.

At the transition to high school, girls are again disproportionately disadvantaged in poor communities (Panel B, Model 5). However, the interaction between gender and community poverty is only marginally significant, and the effect dissipates after household wealth interactions with gender are included, which suggests that the relationship between community poverty and female disadvantage reflects the aggregate effects of household wealth on gender gaps within a village (Panel B, Model 6). Household wealth is also directly linked to patterns in the transition to high school by gender, although the main effect of wealth is not significant\textsuperscript{14}. In Panels A and B, Model 6, girls from relatively wealthy households experience an advantage over boys from similar households at the transition to high school\textsuperscript{15}.

\textsuperscript{14} The statistical insignificance of the main effects of residing in the poorest of households at the transition to high school is unsurprising because the models already control for village measures of poverty.

\textsuperscript{15} Evidence of female advantage is also found in similar models that include community attributes (Table 5b, Panels A – D, Models 3).
To summarize, the preceding table confirms that community poverty is associated with exacerbated gender gaps at both the compulsory and post-compulsory stage. However, female disadvantage associated with community poverty at the transition to high school is only marginally significant, and is no longer significant once household wealth interactions are considered. Moreover, there is a surprising advantage for girls in wealthier households. We turn next to an investigation of the relationship between attributes of poor communities and gender gaps. To begin by providing some context, Table 4 shows the means and proportions of village dimensions of poverty – economic opportunity, isolation, infrastructure, and community support – by the top four and lowest village wealth quintiles. On all dimensions, the wealthiest villages fare better than the poorest villages. For example, the poorest villages are located 3.5 kilometers further away from the nearest town seat than wealthier villages, and less than a fifth of the poorest villages donate money to schools, compared to almost half of wealthier villages.

[Table 4 here]

Tables 5a and 5b investigate the relationship between attributes of poor communities and gender gaps at compulsory and post-compulsory stages. Both tables follow a parallel modeling strategy. Each table has four panels, each containing specifications that include an independent variable representing one community attribute: Panel A is economic opportunity, Panel B is isolation, Panel C is infrastructure, and Panels Da and Db are community support for education by villages and villagers, respectively. In each panel, Model 1 estimates completion or transition as a function of the main effects of the community attribute, household wealth, gender, and age. Model 2 introduces an interaction term between gender and the community characteristic. Model
3 includes additional interaction terms between gender and residing in a household in the lowest or highest wealth quintile.

[Table 5a here]

[Table 5b here]

Turning first to Panel A in Tables 5a and 5b, there is no evidence of a relationship between increased village economic opportunity – meaning the presence of village enterprises – and decreased gender gaps. There are no statistically significant interaction terms between gender and the measure of village economic opportunity in models estimating middle school graduation and transition to high school. This finding is consistent with results from Tables 2 and 3 that show no association between degree of industrialization of the labor market and gender gaps.

In Panel B, village isolation is associated with female disadvantage, but only at the post-compulsory stage (Table 5b, Panel B, Model 3). The coefficients for the main effects of village isolation are also statistically significant and negative at both educational stages, which suggest that children in isolated villages experience a disadvantage compared to youth residing in villages that are closer to the town seat (Tables 5a and b, Panel B, Model 1).

Similar to patterns for village isolation, village infrastructure is associated with gender gaps only at the transition to high school, though the interaction term is only marginally significant once household wealth-gender interactions are included (Table 5b, Panel C, Model 3). However, the coefficient for the main effect of community infrastructure is statistically significant and positive at the compulsory stage, which
suggests that boys and girls in villages with strong infrastructure fare better than their counterparts in villages with weaker infrastructure (Table 5a, Panel C, Model 1).

From the Panel D, community support for education is also associated with gender gaps at the post-compulsory level. Girls in villages that support education fare better than their female counterparts in less supportive villages (Table 5b, Panels Da and Db, Model 2); however, the effect of community support interaction with gender disappears once interactions between household wealth and gender are included in Model 3. The coefficients for the main effects of community support are also statistically significant and positive at the compulsory stage, which indicate that both boys and girls in villages that support education experience an advantage over children in less supportive villages (Table 5a, Panels Da and Db, Models 1).

The previous two tables established an association between gender gaps and community isolation, and, in some specifications, infrastructure and support for education. We next explore whether the link between gender gaps and community poverty can be explained by girls vulnerability to other community attributes. In Table 6, we focus on middle school graduation, which is the educational stage that has a significant gender interaction with community poverty, net of gender interactions with household wealth\textsuperscript{16}. Model 1 estimates completion as a function of the main effects of the community poverty,

\textsuperscript{16}The model that estimates middle school graduation, and includes interaction terms between community poverty and gender and household wealth and gender is found in Table 3, Panel B, Model 3. At the transition to high school, the interaction between gender and community poverty is not significant in models that include gender interactions with household wealth (Table 3, Panel B, Model 6).
household wealth, gender, and age. Model 2 only includes the main effects of community attributes, household wealth, gender, and age, and Model 3 includes main effects for community poverty and all community attributes. Model 4 introduces an interaction term between gender and community poverty. Model 5 includes an additional interaction between gender and household wealth. Model 6 includes interaction terms between gender and community poverty, household wealth, and attributes of poor communities.

[Table 6 here]

Even after considering the effects of other community characteristics on gaps, the association between community poverty and female disadvantage at the compulsory stage persists. Model 1, which is reproduced from Table 3 for ease of comparison, shows that individuals residing in the poorest villages have 56 percent lower odds of graduating than individuals in wealthier villages, net of household wealth\footnote{e^{[-0.83]}}. The main effects of all the attributes of poor communities are shown in Model 2, and two are statistically significant: village isolation is associated with lower odds of graduating from middle school and village infrastructure is associated with higher odds of graduating middle school. Model 3 shows that individuals residing in poor villages have 33 percent lower odds of graduating from middle school than individuals residing in wealthier villages\footnote{e^{[-0.40]}. This coefficient is marginally significant.}. The effect of village isolation is statistically significant, though that of infrastructure is marginally significant. Model 4 indicates that girls residing in poor villages experience
disproportionate disadvantages in graduating from middle school, compared to boys. The interaction between gender and village poverty is also statistically significant in Models 5 and 6, which add controls for the interaction between gender and household wealth (Model 5) and interactions between gender and community attributes (Model 6). Additionally, in Model 6, the main effect of village isolation is still significant and negative, indicating that village isolation is still associated with disadvantage for all children, net of other dimensions of community disadvantage.

CONCLUSION

Overall, three major results emerge from this paper. First, results show that community impoverishment, more than industrialization per se, correlates with gender gaps. Second, poverty and isolation shape gaps differently at different educational stages. Third, girls in the wealthiest households fare better than boys at the high school transition. In the following section, we discuss the implications of our three major findings.

First, the finding that gender gaps are not well explained by degree of industrialization or economic opportunities in local labor markets casts doubt on a straightforward industrialization framework for explaining educational gender inequality patterns. Two potential explanations can help illuminate the lack of relationship we find between industrialization and gender gaps. First, the nature of industrialization in contemporary China is likely different from notions of industrialization on which classical frameworks are based. For example, Boserup stated that “factory work by women is frowned upon in all parts of the under-developed world” (Boserup 1970: 115);
however, it is clear that industrialization in China has brought many women to the workforce. In 2005, almost half of female workers ages 16 - 34 (49 percent) were employed in non-agricultural industries and 20 percent of young women were employed in the manufacturing industry (China 2005 1% Population Survey Data Assembly 2005). Qualitative work has revealed that foreign demand for manufacturing has led to an increasing need for “compliant” young female labor in urban factories, as well as service sectors (Ngai 2004; Ngai and Smith 2007; Otis 2008; Zheng 2008). Moreover, women also report desiring work in industrial sectors for economic gain, as well as for personal independence and to fulfill familial responsibilities (Lee 1998).

The draw of industrialization farther afield may also explain the lack of relationship we find between local industrialization and gender gaps. The targeting of migrant women for industrial occupations located in cities suggest that local measures of labor markets may not be the only relevant forms of industrialization. Vast changes to China’s economy and society in recent decades have attracted an unprecedented number of rural migrants to large urban centers, including a large number of women. Analysis using data from the 1 percent sample of the 2005 mid-censal survey reveals that young migrant women comprise 52 percent of all young rural migrants. Moreover, we find that among young migrant workers, women make up the majority of six of the ten largest non-agricultural occupations, including electronic devices manufacturing and service jobs (Appendix Figure 1)\(^\text{19}\). Future work should consider how industries that prefer female labor and how the possibility of migration to urban industrial centers impact considerations about education for girls and boys.

\(^{19}\) Descriptive statistics is among female migrants ages 15 – 34.
Our second finding shows that the associations between gender gaps and poverty and isolation vary by stage of education. At the compulsory stage of education, girls face greater relative disadvantages in poor communities, and the association of community poverty and gender inequality persists net of both household poverty and other measured attributes of community disadvantage. One possible explanation for the robust effect of community poverty at this stage could be that poor communities share gender norms that disadvantage girls, such as the perception that girls are less capable of doing well in school than boys. These norms may affect more adversely girls in earlier stages of education, where educational aspirations are still forming, than at higher levels of education\textsuperscript{20}. However, in a secondary analysis using GSCF data, we tabulated gender norms, as reflected by village-level aggregates of parental gender norms, by village wealth quintiles. There was no evidence of an association between village wealth and gender egalitarian norms\textsuperscript{21}.

\textsuperscript{20} Non-egalitarian gender norms may stem from structural characteristics of villages, such as a lack of quality schools that girls can access, or cultural factors.

\textsuperscript{21} We investigated this possibility in the GSCF dataset using measures of gender attitudes of mothers and fathers of sample children, aggregated to the community level. Analyses indicate that mothers and fathers in the poorest villages do have less egalitarian gender norms than their counterparts in wealthier villages, and that residing in villages with inequitable gender norms is associated with lower educational attainment among youth. A factor variable was created representing the gender norms of a village. The measures were created by aggregating within a village the responses to questions asking mothers and fathers of child respondents about gender norms, such as whether “girls should enjoy the same education opportunities as boys” and “parents should encourage girls to have their own opinions just like boys.” The factor variable was coded 1, representing very female-biased norms, to 3, representing very gender egalitarian views. The average gender norm for the poorest fifth of villages was 2.59 and 2.66 for the other fourth-fifth of villages. The gender norms of the poorest villages are statistically different (lower) than the other villages, $p < 0.001$. However, further investigation showed that the interaction of community poverty and gender is not substantially changed in models that incorporate community gender norms. Thus, we find no evidence that the disproportionate impact of community poverty on girls at the compulsory stage is accounted for by household poverty, community attributes such as isolation, infrastructure, support for education, or community gender norms. In short, community poverty—signifying material resource deprivation rather than other attributes of community disadvantage—continues to be the most reliable correlate of girls’ disadvantage, at this stage.
Like the relationship between community poverty and gender inequality, the association of village isolation and gender gaps in education also varies by educational level. We find that village isolation is not associated with disproportionate female disadvantage net of interactions between household wealth and gender at the compulsory stage, but is at the transition to high school. At the middle school stage, schools tend to be located within villages or town seats, but at the transition to high school, an educational stage in which schools are often distant from natal villages, students likely have to travel further distances. The importance of isolation at the post-compulsory, but not compulsory stage, is consistent with findings from elsewhere that have suggested that parents are wary of sending daughters to distant schools due to safety concerns (Huisman and Smits 2008; Tansel 2002). These complex relationships between community characteristics and gender gaps at different levels highlight the need for future research to examine more closely why patterns differ between compulsory and post-compulsory stages.

Third, our finding that at the post-compulsory stage, girls benefit disproportionately from residence in relatively wealthy households, compared to boys, is surprising in light of the non-trivial family financial burden associated with the transition to post-compulsory schooling. The female advantage suggests that girls in households that are able to afford substantial tuition fees are supported in continuing school at a greater level than boys in these kinds of households. An emerging female advantage at some educational stages has been found in some other datasets. For example, a study using 2005 China mid-censal data found that college enrollment rates among individuals ages 19 – 22 who had completed senior high school in 2005 were 14.5 percent higher for
women than for men (Wu and Zhang 2010). Results are striking in the rural Gansu sample, where even the wealthiest households are not wealthy by national standards.

In summary, the current study finds a relationship between community poverty, but not measures of local industrialization, and gender gaps at certain stages of education. In order to gain a fuller understanding of how gender shapes progress through educational stages, a different kind of work is needed that focuses on specific educational processes and pathways and how those may differ for girls and boys. Scholars have highlighted the importance of investigating gender disparities in horizontal dimensions of education, such as the quality of education received, in other contexts (Buchmann, DiPrete, and McDaniel 2008). In China, it is not well established whether girls and boys receive similar extracurricular tutoring support, perform at similar levels, or take the same pathways through the system. Different pathways may be particularly important at the post-compulsory transition, when students move into different school types and elect different curricula. For example, less than half of girls in the rural Gansu sample who graduated from middle school transitioned to academic high school, while 57 percent of eligible boys made the transition (our calculations, not shown). In addition, the majority of girls who graduated from any high school and sat the college entrance examination took the arts and letters (humanities) college entrance exam (57 percent), while a much larger majority of boys took the college entrance exam examination in science and engineering (76 percent) (our calculations, not shown). For the rising numbers of those who go onto college, these different paths undoubtedly impact subsequent employment outcomes, such as access to generally higher paying jobs in science or technology fields. Further work on educational support, performance, pathways and content is needed to
fully assess the state of educational gender inequality, and gender inequality trends, in China.
REFERENCES


ABEL Projection Consortium. USAID.


Table 1. Coefficients from Linear Regression Models Estimating Years of Education, All Ages, Rural China, 2002

<table>
<thead>
<tr>
<th>Age group</th>
<th>Main effects model (1)</th>
<th>Interaction model (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 – 34 year olds</td>
<td>-0.878***</td>
<td>-0.607***</td>
</tr>
<tr>
<td>(reference: 15 – 24)</td>
<td>(-0.041)</td>
<td>(-0.055)</td>
</tr>
<tr>
<td>35 – 44 year olds</td>
<td>-1.355***</td>
<td>-0.834***</td>
</tr>
<tr>
<td></td>
<td>(-0.039)</td>
<td>(-0.054)</td>
</tr>
<tr>
<td>45 – 54 year olds</td>
<td>-2.738***</td>
<td>-1.925***</td>
</tr>
<tr>
<td></td>
<td>(-0.039)</td>
<td>(-0.052)</td>
</tr>
<tr>
<td>55 – 64 year olds</td>
<td>-3.392***</td>
<td>-2.555***</td>
</tr>
<tr>
<td></td>
<td>(-0.051)</td>
<td>(-0.065)</td>
</tr>
<tr>
<td>65 – 74 year olds</td>
<td>-5.910***</td>
<td>-4.799***</td>
</tr>
<tr>
<td></td>
<td>(-0.066)</td>
<td>(-0.087)</td>
</tr>
<tr>
<td>Female</td>
<td>-1.342***</td>
<td>-0.376***</td>
</tr>
<tr>
<td></td>
<td>(-0.026)</td>
<td>(-0.050)</td>
</tr>
<tr>
<td>Female X 25 – 34 year olds</td>
<td>-0.559***</td>
<td>(-0.079)</td>
</tr>
<tr>
<td>Female X 35 – 44 year olds</td>
<td>-1.081***</td>
<td>(-0.076)</td>
</tr>
<tr>
<td>Female X 45 – 54 year olds</td>
<td>-1.717***</td>
<td>(-0.076)</td>
</tr>
<tr>
<td>Female X 55 – 64 year olds</td>
<td>-1.917***</td>
<td>(-0.099)</td>
</tr>
<tr>
<td>Female X 65 – 74 year olds</td>
<td>-2.466***</td>
<td>(-0.129)</td>
</tr>
<tr>
<td>Constant</td>
<td>9.309***</td>
<td>8.847***</td>
</tr>
<tr>
<td></td>
<td>(-0.043)</td>
<td>(-0.047)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>29,686</td>
<td>29,686</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.  
Source: CHIP 2002  
*** p<0.01, ** p<0.05, * p<0.1
Figure 1. Male-Female Gaps in Years of Education, By Province/City, Rural China, 2002

Note: Provinces/cities ordered by decreasing provincial per capita GDP.

Source: CHIP 2002
<table>
<thead>
<tr>
<th>Specification</th>
<th>Panel A: Percent nonagricultural households in village</th>
<th>Panel B: Village is in lowest quintile of village wealth</th>
<th>A+B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-0.56*** (0.148)</td>
<td>-0.55*** (0.147)</td>
<td>-0.49 (0.746)</td>
</tr>
<tr>
<td>Percentage nonagricultural labor force in village</td>
<td>0.01** (0.004)</td>
<td>0.08 (0.218)</td>
<td>0.01 (0.006)</td>
</tr>
<tr>
<td>Female X Percentage nonagricultural labor force in village</td>
<td>-0.00 (0.007)</td>
<td>0.08 (0.306)</td>
<td>0.00 (0.07)</td>
</tr>
<tr>
<td>Village is in lowest quintile of village wealth</td>
<td>-0.51** (0.218)</td>
<td>0.08 (0.306)</td>
<td>0.06 (0.324)</td>
</tr>
<tr>
<td>Female X Village is in lowest quintile of village wealth</td>
<td>-1.06*** (0.353)</td>
<td>-0.86** (0.370)</td>
<td>-0.86** (0.379)</td>
</tr>
<tr>
<td>Household wealth (base: middle quintiles)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest quintile of household wealth</td>
<td>-0.03 (0.369)</td>
<td>0.14 (0.361)</td>
<td>0.02 (0.573)</td>
</tr>
<tr>
<td>Highest quintile of household wealth</td>
<td>0.12 (0.400)</td>
<td>0.20 (0.390)</td>
<td>-0.39 (0.603)</td>
</tr>
<tr>
<td>Female X Low HH wealth</td>
<td>-0.22 (0.728)</td>
<td>0.17 (0.724)</td>
<td>0.01 (0.737)</td>
</tr>
<tr>
<td>Female X High HH wealth</td>
<td>0.95</td>
<td>1.08</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 2. Coefficients from Logistic Regression Models Estimating Reaching Middle School, with Measures of Community Labor Market Structure and Poverty
<table>
<thead>
<tr>
<th></th>
<th>(0.786)</th>
<th></th>
<th>(0.775)</th>
<th></th>
<th>(0.787)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.29***</td>
<td>0.29***</td>
<td>0.29***</td>
<td>0.29***</td>
<td>0.30***</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.092)</td>
<td>(0.090)</td>
</tr>
<tr>
<td></td>
<td>0.28***</td>
<td>0.28***</td>
<td>0.28***</td>
<td>0.28***</td>
<td>0.28***</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.091)</td>
<td>(0.092)</td>
<td>(0.091)</td>
<td>(0.092)</td>
</tr>
<tr>
<td></td>
<td>0.29***</td>
<td>0.29***</td>
<td>0.29***</td>
<td>0.29***</td>
<td>0.29***</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.092)</td>
<td>(0.092)</td>
<td>(0.092)</td>
<td>(0.092)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.76</td>
<td>-1.76</td>
<td>-1.71</td>
<td>-1.71</td>
<td>-1.84</td>
</tr>
<tr>
<td></td>
<td>(1.443)</td>
<td>(1.474)</td>
<td>(1.446)</td>
<td>(1.476)</td>
<td>(1.538)</td>
</tr>
<tr>
<td></td>
<td>-1.21</td>
<td>-1.33</td>
<td>-1.28</td>
<td>-1.40</td>
<td>-1.26</td>
</tr>
<tr>
<td></td>
<td>(1.428)</td>
<td>(1.455)</td>
<td>(1.439)</td>
<td>(1.466)</td>
<td>(1.524)</td>
</tr>
<tr>
<td></td>
<td>-1.26</td>
<td>-1.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.544)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2,812</td>
<td>2,812</td>
<td>2,812</td>
<td>2,812</td>
<td>2,812</td>
</tr>
<tr>
<td></td>
<td>2,826</td>
<td>2,826</td>
<td>2,826</td>
<td>2,826</td>
<td>2,826</td>
</tr>
<tr>
<td></td>
<td>2,826</td>
<td>2,826</td>
<td>2,826</td>
<td>2,826</td>
<td>2,826</td>
</tr>
<tr>
<td></td>
<td>2,812</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard errors in parentheses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source: CHIP 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*** p&lt;0.01, ** p&lt;0.05, * p&lt;0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3. Coefficients from Logistic Regression Models Estimating Graduation from Middle School and Transition to High School, with Measures of Community Labor Market Structure and Poverty

<table>
<thead>
<tr>
<th>Specification</th>
<th>Panel A:</th>
<th></th>
<th>Panel B:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Graduation from middle school</td>
<td>Transition to high school</td>
<td>Graduation from middle school</td>
<td>Transition to high school</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.24**</td>
<td>-0.20</td>
<td>-0.30</td>
<td>-0.33**</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.156)</td>
<td>(0.193)</td>
<td>(0.133)</td>
</tr>
<tr>
<td>Household wealth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(base: middle quintiles)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest quintile of household wealth</td>
<td>-0.64***</td>
<td>-0.63***</td>
<td>-0.71***</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.164)</td>
<td>(0.237)</td>
<td>(0.181)</td>
</tr>
<tr>
<td>Highest quintile of household wealth</td>
<td>0.55**</td>
<td>0.55**</td>
<td>0.34</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(0.218)</td>
<td>(0.218)</td>
<td>(0.287)</td>
<td>(0.170)</td>
</tr>
<tr>
<td>Female X Low HH wealth</td>
<td>0.16</td>
<td>-0.28</td>
<td>0.42</td>
<td>-0.83***</td>
</tr>
<tr>
<td></td>
<td>(0.315)</td>
<td>(0.352)</td>
<td>(0.323)</td>
<td>(0.239)</td>
</tr>
<tr>
<td>Female X High HH wealth</td>
<td>0.45</td>
<td>0.86**</td>
<td>0.33</td>
<td>0.87**</td>
</tr>
<tr>
<td></td>
<td>(0.430)</td>
<td>(0.342)</td>
<td>(0.435)</td>
<td></td>
</tr>
<tr>
<td>Community attribute</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Female X Community Attribute</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Age</td>
<td>0.14**</td>
<td>0.14**</td>
<td>0.13**</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.054)</td>
<td>(0.054)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.83</td>
<td>-0.86</td>
<td>-0.77</td>
<td>2.89**</td>
</tr>
<tr>
<td></td>
<td>(1.081)</td>
<td>(1.082)</td>
<td>(1.085)</td>
<td>(1.134)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
<td>1,293</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
Source: GSCF
*** p<0.01; ** p<0.05, * p<0.1
Table 4. Proportions and Means of Village Attributes by Village Wealth Quintiles

<table>
<thead>
<tr>
<th></th>
<th>Top four village wealth quintiles</th>
<th>Lowest village wealth quintile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic opportunity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has village enterprise</td>
<td>0.45</td>
<td>0.15</td>
</tr>
<tr>
<td>Isolation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remoteness (distance in km from town seat)</td>
<td>4.35 (3.98)</td>
<td>7.88 (4.67)</td>
</tr>
<tr>
<td>Infrastructure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. households with water and electricity</td>
<td>0.92</td>
<td>0.74</td>
</tr>
<tr>
<td>Community support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Village donates money to schools</td>
<td>0.47</td>
<td>0.18</td>
</tr>
<tr>
<td>Villagers donates money to schools</td>
<td>0.52</td>
<td>0.44</td>
</tr>
<tr>
<td>Number of villages</td>
<td>79</td>
<td>21</td>
</tr>
</tbody>
</table>

Note: Using two-sample tests of proportion and t-tests, all differences between the top four and bottom villages, in terms of village wealth, are statistically significance at the p<0.01 level. Standard deviations in parentheses.

Source: GSCF
Table 5a. Coefficients from Logistic Regression Models Estimating Graduation from Middle School with Community Attributes

<table>
<thead>
<tr>
<th>Specification</th>
<th>Panel A: Economic opportunity</th>
<th>Panel B: Isolation</th>
<th>Panel C: Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3)</td>
<td>(1) (2) (3)</td>
<td>(1) (2) (3)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.26* (0.137)</td>
<td>-0.31** (0.136)</td>
<td>-0.26* (0.137)</td>
</tr>
<tr>
<td></td>
<td>-0.35** (0.170)</td>
<td>-0.25 (0.226)</td>
<td>-0.85 (0.644)</td>
</tr>
<tr>
<td></td>
<td>-0.45** (0.204)</td>
<td>-0.36 (0.257)</td>
<td>-1.00 (0.667)</td>
</tr>
<tr>
<td>Household wealth (base: middle quintiles)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest quintile of household wealth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.64*** (0.163)</td>
<td>-0.64*** (0.162)</td>
<td>-0.60*** (0.162)</td>
</tr>
<tr>
<td></td>
<td>-0.64*** (0.163)</td>
<td>-0.64*** (0.162)</td>
<td>-0.64*** (0.162)</td>
</tr>
<tr>
<td></td>
<td>-0.73*** (0.235)</td>
<td>-0.71*** (0.234)</td>
<td>-0.71*** (0.235)</td>
</tr>
<tr>
<td>Highest quintile of household wealth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.55** (0.217)</td>
<td>0.34 (0.286)</td>
<td>0.55** (0.217)</td>
</tr>
<tr>
<td></td>
<td>0.55** (0.217)</td>
<td>0.50** (0.217)</td>
<td>0.56** (0.217)</td>
</tr>
<tr>
<td></td>
<td>0.34 (0.286)</td>
<td>0.30 (0.287)</td>
<td>0.33 (0.286)</td>
</tr>
<tr>
<td>Female X Low HH wealth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.17 (0.315)</td>
<td>0.15 (0.314)</td>
<td>0.23 (0.315)</td>
</tr>
<tr>
<td>Female X High HH wealth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.45 (0.430)</td>
<td>0.44 (0.431)</td>
<td>0.49 (0.430)</td>
</tr>
<tr>
<td>Community attribute</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.31 (0.216)</td>
<td>0.17 (0.258)</td>
<td>0.03*** (0.010)</td>
</tr>
<tr>
<td></td>
<td>0.17 (0.258)</td>
<td>0.17 (0.258)</td>
<td>0.02* (0.011)</td>
</tr>
<tr>
<td></td>
<td>-0.10*** (0.023)</td>
<td>-0.09*** (0.027)</td>
<td>0.02* (0.011)</td>
</tr>
<tr>
<td>Female X Community Attribute</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.26 (0.286)</td>
<td>-0.00 (0.029)</td>
<td>0.01 (0.013)</td>
</tr>
<tr>
<td></td>
<td>0.26 (0.286)</td>
<td>-0.00 (0.030)</td>
<td>0.01 (0.013)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.85 (1.074)</td>
<td>0.13** (0.053)</td>
<td>0.13** (0.054)</td>
</tr>
<tr>
<td></td>
<td>-0.79 (1.076)</td>
<td>0.13** (0.054)</td>
<td>0.14** (0.054)</td>
</tr>
<tr>
<td></td>
<td>-0.71 (1.079)</td>
<td>0.12** (0.054)</td>
<td>0.14** (0.054)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.85 (1.074)</td>
<td>-0.23 (1.078)</td>
<td>-2.12* (1.166)</td>
</tr>
<tr>
<td></td>
<td>-0.79 (1.076)</td>
<td>-0.10 (1.085)</td>
<td>-1.89 (1.191)</td>
</tr>
<tr>
<td></td>
<td>-0.71 (1.079)</td>
<td>-0.01 (1.088)</td>
<td>-1.81 (1.194)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
</tr>
<tr>
<td></td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
</tr>
<tr>
<td></td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
</tr>
</tbody>
</table>
Table 5a. Coefficients from Logistic Regression Models Estimating Graduation from Middle School with Community Attributes, continued

<table>
<thead>
<tr>
<th>Specification</th>
<th>Panel Da: Community support, villages</th>
<th>Panel Db: Community support, villagers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.26*</td>
<td>-0.32*</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.166)</td>
</tr>
<tr>
<td>Household wealth (base: middle quintiles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest quintile of household wealth</td>
<td>-0.62***</td>
<td>-0.62***</td>
</tr>
<tr>
<td></td>
<td>(0.162)</td>
<td>(0.163)</td>
</tr>
<tr>
<td>Highest quintile of household wealth</td>
<td>0.54**</td>
<td>0.54**</td>
</tr>
<tr>
<td></td>
<td>(0.217)</td>
<td>(0.217)</td>
</tr>
<tr>
<td>Female X Low HH wealth</td>
<td>0.21</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(0.314)</td>
<td>(0.315)</td>
</tr>
<tr>
<td>Female X High HH wealth</td>
<td>0.47</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>(0.430)</td>
<td>(0.430)</td>
</tr>
<tr>
<td>Community attribute</td>
<td>0.56***</td>
<td>0.46*</td>
</tr>
<tr>
<td></td>
<td>(0.212)</td>
<td>(0.257)</td>
</tr>
<tr>
<td>Female X Community Attribute</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.291)</td>
<td>(0.292)</td>
</tr>
<tr>
<td>Age</td>
<td>0.14**</td>
<td>0.14**</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.99</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td>(1.072)</td>
<td>(1.074)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,735</td>
<td>1,735</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
Source: GSCF
*** p<0.01, ** p<0.05, * p<0.1
Table 5b. Coefficients from Logistic Regression Models Estimating Transition to High School with Community Attributes

<table>
<thead>
<tr>
<th>Specification</th>
<th>Panel A: Economic opportunity</th>
<th>Panel B: Isolation</th>
<th>Panel C: Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.36***</td>
<td>-0.25</td>
<td>-0.39*</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.166)</td>
<td>(0.200)</td>
</tr>
<tr>
<td>Household wealth (base: middle quintiles)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest quintile of household wealth</td>
<td>-0.18</td>
<td>-0.17</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.180)</td>
<td>(0.180)</td>
<td>(0.260)</td>
</tr>
<tr>
<td>Highest quintile of household wealth</td>
<td>0.07</td>
<td>0.07</td>
<td>-0.30</td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td>(0.170)</td>
<td>(0.221)</td>
</tr>
<tr>
<td>Female X Low HH wealth</td>
<td>-0.28</td>
<td>-0.27</td>
<td>-0.07**</td>
</tr>
<tr>
<td></td>
<td>(0.350)</td>
<td>(0.350)</td>
<td>(0.354)</td>
</tr>
<tr>
<td>Female X High HH wealth</td>
<td>0.88**</td>
<td>0.81**</td>
<td>0.88***</td>
</tr>
<tr>
<td>Community attribute</td>
<td>0.28</td>
<td>0.41*</td>
<td>0.40*</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.232)</td>
<td>(0.234)</td>
</tr>
<tr>
<td>Female X Community Attribute</td>
<td>-0.28</td>
<td>-0.26</td>
<td>-0.07**</td>
</tr>
<tr>
<td></td>
<td>(0.273)</td>
<td>(0.275)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Age</td>
<td>2.79**</td>
<td>2.70**</td>
<td>2.97***</td>
</tr>
<tr>
<td></td>
<td>(1.120)</td>
<td>(1.124)</td>
<td>(1.135)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.79**</td>
<td>2.70**</td>
<td>2.97***</td>
</tr>
<tr>
<td></td>
<td>(1.120)</td>
<td>(1.124)</td>
<td>(1.135)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,293</td>
<td>1,293</td>
<td>1,293</td>
</tr>
</tbody>
</table>
Table 5b. Coefficients from Logistic Regression Models Estimating Transition to High School with Community Attributes, continued

<table>
<thead>
<tr>
<th>Specification</th>
<th>Panel Da: Community support, villages</th>
<th>Panel Db: Community support, villagers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.36***</td>
<td>-0.58***</td>
</tr>
<tr>
<td>(0.132)</td>
<td>(0.179)</td>
<td>(0.210)</td>
</tr>
<tr>
<td>Household wealth (base: middle quintiles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest quintile of household wealth</td>
<td>-0.18</td>
<td>-0.16</td>
</tr>
<tr>
<td>(0.180)</td>
<td>(0.181)</td>
<td>(0.260)</td>
</tr>
<tr>
<td>Highest quintile of household wealth</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>(0.170)</td>
<td>(0.171)</td>
<td>(0.221)</td>
</tr>
<tr>
<td>Female X Low HH wealth</td>
<td>-0.22</td>
<td>-0.28</td>
</tr>
<tr>
<td>(0.352)</td>
<td></td>
<td>(0.351)</td>
</tr>
<tr>
<td>Female X High HH wealth</td>
<td>0.86**</td>
<td></td>
</tr>
<tr>
<td>(0.342)</td>
<td></td>
<td>(0.341)</td>
</tr>
<tr>
<td>Community attribute</td>
<td>0.06</td>
<td>-0.17</td>
</tr>
<tr>
<td>(0.190)</td>
<td>(0.229)</td>
<td>(0.231)</td>
</tr>
<tr>
<td>Female X Community Attribute</td>
<td>0.49*</td>
<td>0.41</td>
</tr>
<tr>
<td>(0.269)</td>
<td>(0.272)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.09</td>
<td>-0.09</td>
</tr>
<tr>
<td>(0.055)</td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.87**</td>
<td>2.98***</td>
</tr>
<tr>
<td>(1.121)</td>
<td>(1.129)</td>
<td>(1.138)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,293</td>
<td>1,293</td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses.
Source: GSCF
*** p<0.01, ** p<0.05, * p<0.1
Table 6. Coefficients from Logistic Regression Models Estimating Graduation from Middle School with Community Poverty and other Community Attributes

<table>
<thead>
<tr>
<th>Specification</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-0.27*</td>
<td>-0.27**</td>
<td>-0.28**</td>
<td>-0.02</td>
<td>-0.14</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.136)</td>
<td>(0.136)</td>
<td>(0.164)</td>
<td>(0.196)</td>
<td>(0.760)</td>
</tr>
<tr>
<td>Household wealth (base: middle quintiles)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowest quintile of household wealth</td>
<td>-0.56***</td>
<td>-0.59***</td>
<td>-0.54***</td>
<td>-0.52***</td>
<td>-0.70***</td>
<td>-0.70***</td>
</tr>
<tr>
<td></td>
<td>(0.164)</td>
<td>(0.161)</td>
<td>(0.163)</td>
<td>(0.164)</td>
<td>(0.234)</td>
<td>(0.235)</td>
</tr>
<tr>
<td>Highest quintile of household wealth</td>
<td>0.48**</td>
<td>0.50**</td>
<td>0.46**</td>
<td>0.49**</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>(0.218)</td>
<td>(0.216)</td>
<td>(0.217)</td>
<td>(0.218)</td>
<td>(0.287)</td>
<td>(0.288)</td>
</tr>
<tr>
<td>Female X Low HH wealth</td>
<td>0.36</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.321)</td>
<td>(0.324)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female X High HH wealth</td>
<td>0.30</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.435)</td>
<td>(0.436)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community attribute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Village is lowest quintile of village wealth</td>
<td>-0.83***</td>
<td>-0.40*</td>
<td>0.06</td>
<td>0.06</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.239)</td>
<td>(0.237)</td>
<td>(0.292)</td>
<td>(0.296)</td>
<td>(0.309)</td>
<td></td>
</tr>
<tr>
<td>Female X Village is lowest quintile of village wealth</td>
<td>-0.87***</td>
<td>-0.91***</td>
<td>-0.96***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.303)</td>
<td>(0.316)</td>
<td>(0.352)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic opportunity</td>
<td>0.09</td>
<td>0.00</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.202)</td>
<td>(0.208)</td>
<td>(0.211)</td>
<td>(0.212)</td>
<td>(0.265)</td>
<td></td>
</tr>
<tr>
<td>Female X Economic opportunity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.316)</td>
<td></td>
</tr>
<tr>
<td>Isolation</td>
<td>-0.08***</td>
<td>-0.07***</td>
<td>-0.07***</td>
<td>-0.07***</td>
<td>-0.08***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.022)</td>
<td>(0.026)</td>
<td></td>
</tr>
<tr>
<td>Female X Isolation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>0.02**</td>
<td>0.02*</td>
<td>0.02*</td>
<td>0.02*</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.011)</td>
<td></td>
</tr>
<tr>
<td>Female X Infrastructure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>Community support: villages</td>
<td>0.29</td>
<td>0.24</td>
<td>0.25</td>
<td>0.26</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.204)</td>
<td>(0.205)</td>
<td>(0.207)</td>
<td>(0.208)</td>
<td>(0.258)</td>
<td></td>
</tr>
<tr>
<td>Female X Community support: villages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.317)</td>
<td></td>
</tr>
<tr>
<td>Community support: villagers</td>
<td>0.24</td>
<td>0.26</td>
<td>0.26</td>
<td>0.25</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
<td>(0.193)</td>
<td>(0.196)</td>
<td>(0.197)</td>
<td>(0.247)</td>
<td></td>
</tr>
<tr>
<td>Female X Community support: villagers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.13**</td>
<td>0.13**</td>
<td>0.12**</td>
<td>0.13**</td>
<td>0.13**</td>
<td>0.13**</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Age</td>
<td>(0.054)</td>
<td>(0.053)</td>
<td>(0.053)</td>
<td>(0.054)</td>
<td>(0.054)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.45</td>
<td>-1.42</td>
<td>-1.17</td>
<td>-1.33</td>
<td>-1.27</td>
<td>-1.10</td>
</tr>
<tr>
<td></td>
<td>(1.072)</td>
<td>(1.166)</td>
<td>(1.175)</td>
<td>(1.187)</td>
<td>(1.190)</td>
<td>(1.223)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
<td>1,735</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

Source: GSCF

*** p<0.01, ** p<0.05, * p<0.1
## Appendix Table. Descriptive Statistics for CHIP and GSCF Samples

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean / Percent</th>
<th>Standard Deviation</th>
<th>Analytic Sample Size (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Statistics for CHIP Sample</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of Education</td>
<td>6.61</td>
<td>3.08</td>
<td>35,230</td>
</tr>
<tr>
<td>Female</td>
<td>47.98</td>
<td>---</td>
<td>37,969</td>
</tr>
<tr>
<td>Age</td>
<td>33.04</td>
<td>18.52</td>
<td>37,969</td>
</tr>
<tr>
<td>Province</td>
<td>100</td>
<td>---</td>
<td>37,969</td>
</tr>
<tr>
<td>Beijing</td>
<td>1.48</td>
<td>---</td>
<td>563</td>
</tr>
<tr>
<td>Hebei</td>
<td>3.98</td>
<td>---</td>
<td>1513</td>
</tr>
<tr>
<td>Shanxi</td>
<td>4.27</td>
<td>---</td>
<td>1622</td>
</tr>
<tr>
<td>Liaoning</td>
<td>4.17</td>
<td>---</td>
<td>1583</td>
</tr>
<tr>
<td>Jilin</td>
<td>4.64</td>
<td>---</td>
<td>1796</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>4.2</td>
<td>---</td>
<td>1594</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>5.09</td>
<td>---</td>
<td>1932</td>
</tr>
<tr>
<td>Anhui</td>
<td>4.84</td>
<td>---</td>
<td>1837</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>5.08</td>
<td>---</td>
<td>1927</td>
</tr>
<tr>
<td>Shandong</td>
<td>6.17</td>
<td>---</td>
<td>2343</td>
</tr>
<tr>
<td>Henan</td>
<td>5.84</td>
<td>---</td>
<td>2219</td>
</tr>
<tr>
<td>Hubei</td>
<td>5.51</td>
<td>---</td>
<td>2093</td>
</tr>
<tr>
<td>Hunan</td>
<td>4.87</td>
<td>---</td>
<td>1848</td>
</tr>
<tr>
<td>Guangdong</td>
<td>6.54</td>
<td>---</td>
<td>2483</td>
</tr>
<tr>
<td>Guangxi</td>
<td>5.33</td>
<td>---</td>
<td>2025</td>
</tr>
<tr>
<td>Chongqing</td>
<td>1.78</td>
<td>---</td>
<td>677</td>
</tr>
<tr>
<td>Sichuan</td>
<td>4.82</td>
<td>---</td>
<td>1832</td>
</tr>
<tr>
<td>Guizhou</td>
<td>4.81</td>
<td>---</td>
<td>1825</td>
</tr>
<tr>
<td>Yunnan</td>
<td>3.16</td>
<td>---</td>
<td>1199</td>
</tr>
<tr>
<td>Shaanxi</td>
<td>4.32</td>
<td>---</td>
<td>1641</td>
</tr>
<tr>
<td>Gansu</td>
<td>3.82</td>
<td>---</td>
<td>1449</td>
</tr>
<tr>
<td>Xinjiang</td>
<td>5.27</td>
<td>---</td>
<td>2001</td>
</tr>
<tr>
<td>Village measure: percent non-agricultural labor force in village</td>
<td>31.25</td>
<td>21.32</td>
<td>2860</td>
</tr>
<tr>
<td>Village measure: wealth of villages in lowest wealth quintile</td>
<td>0.04</td>
<td>0.20</td>
<td>2879</td>
</tr>
<tr>
<td>Household wealth</td>
<td>93.91</td>
<td>378.34</td>
<td>2879</td>
</tr>
<tr>
<td>Has reached middle school (overall) among individuals ages 15 to 17</td>
<td>89.38</td>
<td>30.81</td>
<td>2826</td>
</tr>
<tr>
<td>Has reached middle school (girls) among individuals ages 15 to 17</td>
<td>87.25</td>
<td>33.37</td>
<td>1333</td>
</tr>
<tr>
<td>Has reached middle school (boys) among individuals ages 15 to 17</td>
<td>91.29</td>
<td>28.20</td>
<td>1493</td>
</tr>
</tbody>
</table>

**Descriptive Statistics for Variables used in the GSCF Analysis**
<table>
<thead>
<tr>
<th></th>
<th>Value 1</th>
<th>Value 2</th>
<th>Year 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>46.27</td>
<td>49.87</td>
<td>1742</td>
</tr>
<tr>
<td>Age (in 2009)</td>
<td>19.94</td>
<td>1.26</td>
<td>1744</td>
</tr>
<tr>
<td>Household wealth</td>
<td>22,134.96</td>
<td>30,572.40</td>
<td>1744</td>
</tr>
<tr>
<td>Village measure: percent non-agricultural households in village</td>
<td>9.34</td>
<td>15.96</td>
<td>1728</td>
</tr>
<tr>
<td>Village measure: wealth of villages in lowest wealth quintile</td>
<td>10,566.64</td>
<td>8702.92</td>
<td>1744</td>
</tr>
<tr>
<td>Village measure: economic opportunity, village has enterprise(s)</td>
<td>39.11</td>
<td>48.81</td>
<td>1744</td>
</tr>
<tr>
<td>Village measure: isolation, distance of village from town seat</td>
<td>5.06</td>
<td>4.36</td>
<td>1744</td>
</tr>
<tr>
<td>Village measure: infrastructure, percent houses with access to electricity and water</td>
<td>49.78</td>
<td>11.21</td>
<td>1744</td>
</tr>
<tr>
<td>Village measure: community support, village donates money to schools</td>
<td>41.28</td>
<td>49.25</td>
<td>1744</td>
</tr>
<tr>
<td>Village measure: community support, villagers donates money to schools</td>
<td>50.17</td>
<td>50.01</td>
<td>1744</td>
</tr>
<tr>
<td>Graduated from middle school (overall)</td>
<td>82.37</td>
<td>38.12</td>
<td>1735</td>
</tr>
<tr>
<td>Graduated from middle school (girls)</td>
<td>79.58</td>
<td>40.34</td>
<td>801</td>
</tr>
<tr>
<td>Graduated from middle school (boys)</td>
<td>84.76</td>
<td>35.96</td>
<td>932</td>
</tr>
<tr>
<td>Transitioned to any upper secondary school among middle school graduates (overall)</td>
<td>70.79</td>
<td>45.49</td>
<td>1293</td>
</tr>
<tr>
<td>Transitioned to any upper secondary school among middle school graduates (girls)</td>
<td>66.54</td>
<td>47.22</td>
<td>562</td>
</tr>
<tr>
<td>Transitioned to any upper secondary school among middle school graduates (boys)</td>
<td>74.01</td>
<td>43.89</td>
<td>731</td>
</tr>
</tbody>
</table>

*Note: Village measures and household wealth in the CHIP dataset are among individuals ages 15 to 17.*
<table>
<thead>
<tr>
<th>Village Wealth Quintiles</th>
<th>Migration out of village</th>
<th>Village migrant income</th>
<th>Non-agricultural labor force in village</th>
<th>Number of Villages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest</td>
<td>Percent</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>25.98</td>
<td>20.99</td>
<td>192.90</td>
<td>173.07</td>
</tr>
<tr>
<td>2</td>
<td>27.03</td>
<td>19.32</td>
<td>128.44</td>
<td>147.74</td>
</tr>
<tr>
<td>3</td>
<td>25.26</td>
<td>17.13</td>
<td>108.82</td>
<td>84.76</td>
</tr>
<tr>
<td>4</td>
<td>11.08</td>
<td>14.47</td>
<td>73.47</td>
<td>97.96</td>
</tr>
<tr>
<td>Highest</td>
<td>8.83</td>
<td>9.49</td>
<td>63.60</td>
<td>149.93</td>
</tr>
<tr>
<td>Overall</td>
<td>19.75</td>
<td>18.55</td>
<td>113.83</td>
<td>142.54</td>
</tr>
<tr>
<td>Correlation with village wealth</td>
<td>-0.33</td>
<td>-0.28</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

Data source: GSCF wave 2

Note: Village wealth is an aggregated measure of household wealth. The percent migration out of village is defined as the percentage of individuals who went out of the village to work for at least three months in 2003, among all laborers in the village. Village migrant income is defined as the average income received by households from family members working outside of the village in 2003, aggregated to the village level. Non-agricultural labor force in village is defined as the percent of the village workforce involved in non-agricultural industries.

Note: Correlation of village wealth and percent non-agricultural labor force in village using CHIP 2002 data is 0.07.
Appendix Figure 1. Distribution of the Migrant Workforce across Occupational Categories, 15-34 Year Olds

Data source: Our calculations using 1 Percent Sample of 2005 China Mid-Censal Survey.
Note: Migrants are defined as individuals with rural *hukou* who are not residing in this area time of survey.
Note: Women comprise 48 percent of all migrant workers, ages 15 - 34 years.