Paper submitted to the special Issue of Journal of Chinese Economics and Business studies

Are children dropping out during compulsory schooling, and are there differences between ethnic groups? The cases of rural areas in Guizhou and Hunan provinces, China♣

Kelly Labar* and Bente Castro Campos*

** IAMO Leibniz Institute of Agricultural Development in Central and Eastern Europe, Halle (Saale), Germany.*

IAMO
Theodor-Lieser Strasse, 2
06120 Halle (Saale)
Germany
Contact: labar@iamo.de; Castro_Campos@iamo.de

1. Introduction

Improving access to education in China has become one of the government's priorities. As stressed in the World Bank mid-term evaluation of China´s 11th Five Year Plan (2006-2011) "this is the first time that promoting education equity is identified as a key objective" (World Bank 2008). Among the four main measures included in this plan, universalizing the nine years of compulsory education is stated as a goal. Particularly, a more equal access to education for children from rural areas, poor families, for girls and for children of ethnic minorities is promoted. Furthermore, in some provinces, members of ethnic minority groups often cumulate these factors excluding children from the educational system. This is for instance the case in South and South-West China where ethnic minority families are often living in rural areas and among the poorest (World Bank 2009).

Whereas much focus has been concentrated on the analysis of the rural-urban education gap, fewer analyses are dedicated to the comparison in schooling attainment between the Chinese ethnic majority group (the Han) and the ethnic minorities. The available

microeconomics literature until now has mainly emphasized geographically heterogeneous findings. For instance, the ethnic minorities located in Northeast China (mainly Koreans and Manchus) have a higher educational level compared to those in Southern and Western provinces (Gladney 1995, Hannum 1999, 2002, Bhalla and Qui 2006, Gustafsson and Sai 2006). But to our knowledge, few studies have been concentrated on educational levels of ethnic groups from South and South West China, particularly for the years after the implementation of the law on compulsory schooling (1986).

The present article proposes to fill this gap by focusing on the provinces Guizhou and Hunan in Southern China. In order to focus on the poorest tail of the population and for more comparability, we assess the differences between ethnic minorities and the Han in rural areas. In view of the heterogeneous implementation of the nine years of compulsory schooling, we restrict our attention to the first nine years of education and assess whether the probability of dropping out of school before validating nine years significantly depends on the ethnic status.[1]

Indeed, potential differences in the drop out rates between ethnic groups find their origin in various factors. First, access to education for ethnic minorities is restrained due to their mainly rural location and their stronger poverty (Hannum, Wang and Adams forthcoming). Second, higher levels and higher quality schools are often situated in cities and school fees are also a barrier for poorer households among which ethnic minorities in Southern and Western China are often found. Besides this lower access to education, heavier opportunity costs are often observed for ethnic minorities compared to the Han counterparts. Third, members of ethnic minority groups also face a lower access to off-farm jobs (Chapman, Chan and Postiglione 2000, Bhalla and Qui 2006). Returns to higher education may thus be valued by the parents as too low to urge them to invest further in their children's education. Furthermore, households from ethnic minority groups mostly involved in agriculture may benefit from the child's assistance on the farm. Sending them to school would mean to lose a helping hand in doing household obligations, which implies a lower household income in the short run.

In 1986- the Chinese government established the nine years of compulsory school which covers the years from primary school up to junior high school. This decision, however, was taken in parallel with the decision of decentralized financing of education. The provincial governments then employed school fees to finance the educational system in their province during the eighties and nineties. As a consequence children from poor families have been excluded from the educational system and the efficiency of the law of compulsory schooling slowed down. Since September 2006, tuition fees have been removed for primary school in order to overcome this issue. Combined with the new objective of the government to make the nine years of compulsory education universal, equal access to education should be guaranteed whatever children's socio-economic background or ethnic status.

While the effect of this new legislation will appear in the coming years, we can already assess whether the gap between children from ethnic minorities and the Han has decreased in rural areas after the establishment of the nine years of compulsory education despite the decentralization in the financing of education. Using a theoretical model including the major determinants identified as influencing an individual's educational attainment, we show how the ethnic minority status can play a role in the probability to drop out of school. The China Health and Nutrition Survey, which compiles panel structured data on a regular basis between 1991 and 2006 allows us to use survival analysis to study whether children from ethnic minority families are dropping out of school earlier than children from Han families. We focus on rural areas in Guizhou and Hunan provinces. In our sample three major ethnic minorities (the Bouyei, Miao and Tujia) are particularly represented. We will show in section 2 how ethnicity may be linked to the supply factors (e.g. availability of schools) and demand factors (e.g. household income). The role played by preferential policies with regards to these two major groups of factors will also be considered here. In section 3 we integrate the conclusions of the literature review in our theoretical model in order to highlight how ethnic status enters in the school length decision. Section 3 also presents the empirical methodology used to test for the main hypothesis following from the theoretical part. In section 4, we describe the China Health and Nutrition Survey (CHNS) dataset which has the particularity to be longitudinal running from 1991 to 2006 allowing us to implement

survival analysis. Results are provided in section 5 distinguishing between descriptive and econometric elements. Finally section 6 concludes.

2. Ethnic minorities and the Chinese educational system

2.1. China's educational system

The school availability and quality in China depend primarily on the individual's place of residence, independently of ethnic status. Only the primary schools (grades 1 to 6) are located in the villages, while schools for obtaining grades 7 to 9 are only available in the nearby small towns, and schools for obtaining grades 10 to 12 in the closest city. For households living in remote rural areas, what is the case for most of the ethnic minorities, the distance to the next secondary education schools can thus be very long. The establishment of boarding schools in border, pastoral and mountainous regions contributes to the improvement of this situation, but studying and living conditions in these schools are not always optimal (Sautman 1997).[2]

Besides the heterogeneous geographic availability of schools, the financial accessibility has suffered from the decentralized financing of the educational system implemented by the end of the 1980's. Sub-national governments newly responsible for the educational budget have established tuition fees. These costs come above the ones linked to the school stationery (e.g. books, pencils) and travel expenditures. Education has slowly represented a growing part in the household's budget (Gustafsson and Li 2003) excluding children from the poorest families among which we find people from rural areas and from ethnic minorities (World Bank 2009). With the introduction of the K-9 rural education program meant at universalizing access to education until nine years of schooling, all ethnic groups should have equal access to this compulsory education (Education Law of September 2006). Indeed, this law stipulates the elimination of expenses for tuition and books in primary schools. The REAP (2008) stresses, however, that it is very difficult to change immediately what was the rule for more than 50 years. Furthermore, the expenses for higher education have not been reduced by the government.[3] Loans for the best students are now offered by the State and receive an increasing attention from the Chinese government to broaden the access to higher education for the poorer.

The quality of education is another issue in rural areas. Lower quality of education in primary schools has consequences on the final educational attainment impeding children to perform well at junior high school. Language plays also a major role here. Often the ethnic minorities live in compact communities without any Han influence. In school they have the first contact with Mandarin, which gives them a different starting level than the Han as they have to learn a new language. For rural Guizhou it was reported that ethnic minorities often have to give up their own ethnic identity and have to adapt to Han culture in order to be successful in school; speaking Chinese instead of the ethnic language and the unfamiliar majority culture can lead to cultural shocks for the ethnic minorities and increase the school drop-out rates (Castro Campos, forthcoming). This also negatively affects the access to senior high school or university for which the access is restricted by the entrance examination. Some of the ethnic minorities may even not consider to take part in the examination due to their language disadvantages. The expectation of a higher failure at this examination leads thus to a huge drop out of ethnic minorities already during primary education although they may benefit from preferential policies.

Opportunity costs of education, resulting from the child going to school instead of helping with household obligations, represents a negative incentive for parents to send their child to school.[4] This is particularly the case for poorer families and households involved in agriculture who would benefit from the help of the child at home, on the field or as an employee (Basu and Tzannatos 2003). Rural people and members of ethnic minorities represent a huge part of poorer and/or agricultural households (Gustafsson and Sai 2008). Despite the law of banning child labor (Regulations Prohibiting the Use of Child Labor, China 1991, revised 2002), the Research Report n°3 of the China Labor Bulletin stresses that child labor is still a common practice in China.[5] Child's involvement in the household's obligations is also influenced by the availability of schools. If the child has to spend the week or, depending on the boarding school, the whole term at a boarding school, he/she cannot contribute to the household's obligations after school. The households in remote areas face thus higher opportunity costs of sending their children to school than households closer to townships. If parents privilege the present, they could consequently decide to take their children out of school sooner

than they would have done if the school would have been closer. If they, however, consider the long term perspective, they could decide to send the child to the boarding school where he/she could better focus on his/her studies as no household obligations have to be done. On the other hand, low expected future income from education also negatively influences the schooling length. Overall, returns to education appear lower in rural areas even if De Brauw and Rozelle (2007) and Zhang, Huang, and Rozelle (2002) argue that the aggregate rates of returns to education have increased over time in these areas.[6] Besides, comparing the Han majority and the Hui minority in employment in the state sector in Lanzhou (Gansu) in 2001, Zang (2008) finds that ethnic minority status is the main determinant of labor market discrimination, controlling for educational attainment and other key characteristics. Ethnic minority status is even more important than gender in job attainment in the state sector. Parents from ethnic minority groups may thus consider an additional year of education for their children more as a financial burden than as a present investment leading to higher future earnings.

Finally, it is interesting to note that some authors have observed different preferences between ethnic groups with regards to education. For example, Bhalla and Qui (2006) suggest for example that the Miao value education higher than other ethnic groups. Generally speaking, language barriers, religion and the degree of assimilation into Han culture play a central role in explaining the various educational situations of ethnic minorities in China (Chiang 2001; Feng 2005; Wu 2006; Yi 2005).

2.2. Education and preferential policies towards ethnic minorities

From the White Papers of the Chinese Government: (http://www.china.org.cn/e-white/20050301/V.htm), we can read that "In 2003, there were 83,726 schools at all levels and of all kinds in ethnic autonomous areas, with a total enrollment of 29.43 million, an increase of five fold compared with 1952, […] and of 10.6 percent compared with 1994. There were 1.541 million specialized teachers, an increase of 16 percent compared with 1994. […] The fifth national census, conducted in 2000, showed that the years of schooling of 14 ethnic minorities, including the Korean, Manchu, Mongolian and Kazak groups, were higher than the national average."

These observations can partly be explained by different preferential policies in favor of ethnic groups which are meant to influence the children's educational attainment both through relaxed direct educational costs and through an improvement of labor opportunities.

In terms of direct costs, reduced school fees for ethnic minorities have been implemented (Sautman 1997). Besides, while both children from the Han group and from the ethnic minorities are equally facing the lower educational quality in rural areas, children from ethnic minorities benefit from preferential policies which decrease the weight of this disadvantage in their later studies. These policies cover a quota system supporting the entrance of minority children at senior high school and universities. They receive additional points for their entrance examination and additional preparatory courses for this examination. In this case actually the Han could be penalized in accessing better universities.[7] In some cases the entrance examinations can also be obtained in the indigenous language of the ethnic minority group (Sautman 1997). For the greater part of the ethnic minorities it is, however, not the case. Children living in compact ethnic communities may speak their local language at home and get in contact with Mandarin only in school.

Preferential policies can differ between ethnic minority groups, universities and between geographical areas. Bhalla and Qui (2006) argue for instance that the urban bias of the preferential policies might be an explanation for the urban ethnic minorities to have more education than the Han. Positive outcomes of preferential policies have been noticed in some provinces. In Xinjiang the inequalities in employment declined with the preferential policies in education through better access to senior high school and universities for the ethnic minorities (Sautman 1999). In Inner Mongolia, as a result of the preferential policies in education, 48 % of Mongolian students continued to receive additional education after completing high school in the middle of 1990s, which is a higher proportion compared to other ethnic groups (Iredale, Bilik, and Su 2001).

With regards to labor market opportunities, the 1982 constitution gave more rights to minorities. Article 4 of the constitution of PRC guarantees that all nationalities of PRC are equal, and that discrimination against and oppression of any nationality are

prohibited. More specific, preferential policies in support of minorities address economic development and poverty reduction, cultural aspects (family planning, preservation of minority culture, languages and religion) and educational and labor market aspects. (Bhalla and Qui 2006, pp. 119-120). In January 2008 the Employment Promotion Law came into force, promising equal employment opportunities among the ethnic minority and majority groups. In the autonomous regions, prefectures, and counties the governmental positions [not Chinese Communist Party (CCP) positions] should be given to the major ethnic minority groups of this autonomous area. Often issues regarding the allocation of these positions arise as many autonomous prefectures and counties are divided between two ethnic minority groups.

These policies, however, have not yet succeeded to prevent the differentiate access to the labor market in some areas (e.g. Zang 2008 for Gansu and Hillman 2008 for Tibet). Bhalla and Qui (2006) and Chapman, Chan, and Postiglione (2000) argue that the low perceptions of ethnic minorities about job prospects may increase their opportunity costs towards educational attainment and thus have an influence on the lower education of children in remote areas.

3. Theoretical framework

3.1. Schooling length

The focus of the study is on the determination of schooling length as being dependant on the child's ethnic status.[8] Without considering this specific characteristic, Card (2001) and Orazem and King (2008) have developed similar models leading to the optimal choice of schooling length. We will base our analysis on these models, looking more specifically at the link between ethnic status and the different elements influencing the child's schooling length. Furthermore, we remind that the Chinese educational system establishes nine years of compulsory school and in parallel increasingly appeals to the parental budget as the educational level of the children becomes higher (Gustafsson and Li 2003). This may lead to non-linear effects through educational years of certain parental and contextual factors on the choice of school length. After emphasizing these non-linearities in the theoretical model, we will focus on what centrally interests us: whether or not the parents validate the nine years of compulsory education for their

children and whether or not this validation differs between the ethnic majority and minority groups.

We consider that the household is made out of two generations: the parents and the child.[9] In what follows, we use *c* as a superscript for elements linked only to the child, *p* for elements linked only to parents and *h* when elements concern the whole household. The parents decide about the child's schooling length (choice variable) based on the expected costs and benefits. We consider consequently that the parents maximize the present value of the household's utility derived from the consumption subject to the budget constraint. We assume- like Card (2001) that households have an infinite planning horizon divided between the period when the child attends school (*t=0* corresponds to the beginning) and the period after he/she has terminated the chosen educational level. We highlight, moreover, the different factors influencing the decision for an educational attainment higher or lower than 9 years of education. At period *t*, the household with characteristics *z* has a utility *u* that depends on the household's consumption $c^h(t,z)$ and during the period of education also on the disutility $\varphi^h(t,z)$ coming from the fact that the child is at school instead of helping with household obligations. *z* includes the household's ethnic status, which can influence the consumption behavior or the preference for work or school, and other individual or household characteristics influencing the household's consumption or income.[10] With ρ being the subjective discount rate, we can then define the present value for the household conditional on the child's schooling length *E* as:

$$V^h(E, c^h(t,z)) = \int_0^E \left(u\left(c^h(t,z)\right) - \varphi^h(t,z)\right) e^{-\rho t} dt + \int_E^{\infty} u\left(c^h(t,z)\right) e^{-\rho t} dt \qquad (1)$$

We then construct the household's budget constraint as a function of costs and incomes. First, school fees *T(t,z)* enter as costs. They depend on household characteristics as we assume that the overall household income is used to pay for it. As a result of the preferential policies children from poorer households or from ethnic minority groups benefit from a lower level of tuition fees.[11] If one considers, moreover, the law cancelling the tuition fees until nine years of education, we see that the parental decision is influenced in a different manner before and after this nine year threshold. Furthermore, one effect of the law on compulsory education is that a social cost may appear if parents

decide to take their child out of school against the law. We mention here a social cost in the sense that a financial penalty for families, who do not respect this law, has not been established in China. We note this cost *SC(E,z)* as it appears only when the drop out of school occurs. It is characterized by:

$$\begin{aligned} &MaxSC(E,z) = SC(0,z) \\ &SC(E \geq 9, z) = 0 \\ &\frac{\partial SC(E,z)}{\partial E} < 0 \end{aligned} \qquad (2)$$

The social cost is considered as a function of the level of education which is decreasing as the level of education comes closer to the nine years compulsory level. It is also considered to depend on parental characteristics in a way that peer effects linked to the household's status may also influence this cost. For instance, other members of a similar social group may particularly promote education and would thus blame parents for taking the child out of school before completing the compulsory education. In terms of income, we define $y^c(E,t,z)$ as being the individual earnings at period t linked to the final level of education reached E and on the household characteristics z.[12] Parents' work also provides an income which we assume is used for the overall household $y^p(t,z)$ during the child's education and after the completion of the expected level of education. This income is considered exogenous to the model despite the limiting power of this hypothesis (parents can decide to work more hours in order to increase their income and to be able to invest more in the child's education). We thus consider that parental income does not depend on the child's level of education in period t. As will be clarified later, this hypothesis is linked to data limitations. We thus consider it as such already in the model. Parental income depends on the period t. Under the hypothesis that the individual can borrow and lend freely, the costs and overall household earnings are discounted by an exogenous interest rate R, leading to define the intertemporal budget constraint as:

$$\int_0^{\infty} c^h(t,z)e^{-Rt}dt = \int_0^{E}\left(y^p(t,z) - T(t,z)\right)e^{-Rt}dt + \int_E^{\infty}\left(y^p(E,t,z) + y^c(E,t,z)\right)e^{-Rt}dt - \frac{SC(E,z)}{(1+R)^E} \qquad (3)$$

In equation (3), if E^*, which is the optimal chosen educational level of the child given the budget constraint of the household, is ≥ 9 , the last term equals zero as specified in

equation (2). If $E^* < 9$, the income is decreased by the social cost linked to the violation of the law of nine years of compulsory school. Tuition fees, moreover, differ between the educational levels.

The schooling length choice maximizes equation (1) subject to the budget constraint (3). This leads to the maximization of:

$$\Omega(E, c(t,z), \lambda) = V^h(E, c^h(t,z)) + \lambda \left\{ \begin{array}{l} \int_0^{\infty} c^h(t,z) e^{-Rt} dt - \int_0^E \left(y^p(t,z) - T(t,z) \right) e^{-Rt} \\ - \int_E^{\infty} y^p(E,t,z) + y^c(E,t,z) e^{-Rt} dt + \dfrac{SC(E,z)}{(1+R)^E} \end{array} \right\} \quad (4)$$

with regards to *E*. The optimal choice *E** is obtained thanks to the following first order condition:

$$\Omega_E(E, c(t,z), \lambda) = \lambda e^{-RE} \{ MB(E,z) - MC(E,z) \} \quad (5)$$

Where

$$MB(E,z) = \int_0^{\infty} \left(\partial y^p(E, E+\tau, z) / \partial E + \partial y^c(E, E+\tau, z) / \partial E \right) e^{-R\tau} d\tau$$

represents the marginal benefit from investing in year E^{th} education,

and

$$MC(E,z) = y^c(E,E,z) + y^p(E,E,z) - y^p(E,z) + T(E,z) + \frac{1}{(1+R)^E} \frac{\partial SC(E,z)}{\partial E} + \left(1 / \lambda e^{-(\rho+R)E} \right) \varphi^{hh}(E,z) \quad (6)$$

represents the marginal cost linked to the investment in year E^{th} education.

This first order condition provides the following testable hypothesis: 1. the higher the parental income is, the longer the child will stay at school; 2. the higher the expected parental income is, the longer the child will stay at school; 3. the higher the school fees are, the sooner the child will drop out of school; 4. the higher the disutility of the household linked to school is, the sooner the child will drop out of school. We have, moreover, demonstrated that non-linear costs of education can be observed. For our purpose, we will focus on what happens before entering in senior high school. We want

to find out about the amount of children dropping out of school before finishing the nine years of compulsory education and about differences in the drop out rates between ethnic minorities and the majority group. Besides the hypotheses mentioned above, secondary hypotheses linked to the role played by the ethnic status can thus be highlighted. From the literature review, we can deduce that ethnic status is a potential determinant of the child's probability to drop out of school as they are often among the poorest households living in remote rural areas. Considering these facts we can hypothesize that i. the costs of education may depend negatively either directly (tuition fees) or indirectly (distance to school) on the ethnic minority status; ii. the present parental income and the child´s expected returns to education depend negatively on the ethnic minority status; iii. the time preference may differ between ethnic groups.

Our objective is consequently to test the hypotheses i., ii. and iii. looking at the channels through which ethnic status may influence the child's drop out of school.

3.2. Survival analysis applied to schooling length

In general we are interested in the probability of children to drop out of school before having reached nine years of education, and in particular in differences in these drop out rates among ethnic groups. We assume that each year parents must decide whether or not the child will continue his/her education given that he/she has already reached a certain level. Survival analysis is thus the appropriate empirical method to implement. It studies the time to the occurrence of an event that is in our case, dropping out of school. Dropping out of school corresponds to the failure event given that the individual has attended school until this event. The hazard function provides the probability that the failure event occurs at a given time, conditional on the survival of the individual until this date. Transposed to the current analysis, the hazard function describes the time when parents decide not to send their child to school anymore. The hazard function h is defined as:

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > E > t \mid E > t)}{\Delta t} = \frac{f(t)}{S(t)} \qquad (7)$$

where f is the density function and S is the survivor function mirroring the probability to survive beyond t.

Apart from the work of Arulampalam, Naylor and Smith (2004), most studies using survival analysis to investigate school length, consider a continuous proportional hazard duration model (Brown and Park 2002, Raymond and Sadoulet 2003, Garcia-Aracil and Winter 2006). We choose in the present study, however, to make use of a discrete model for two main reasons. First, the length of the maximum spell (duration of the state of "being at school") is with only nine years very short. Second, the unit of time duration is relatively long compared to the maximum spell (the year). It is worth stressing that similar results have been obtained using equivalent estimation methods in the continuous frame.

We consequently consider a parametric discrete hazard model which is the discrete version of a Weibull specification. It is the so-called complementary log-log specification. As a proportional hazard model, the hazard function can be written as:

$$h(E \mid X) = h_0(E)\exp(\beta' X) \qquad (8)$$

X represents the matrix of covariates and β the matrix of coefficients associated to each element of X which quantifies its effect on the hazard rate and is estimated from the data. $h_0(E)$ is the baseline hazard referring to the overall hazard when all covariates equal to zero. We choose as a functional form for the baseline hazard the discrete equivalent of the Weibull function. The duration dependence is consequently of the form $h_0(E) = r\log(E)$ where r is a parameter to be estimated. If $r>0$, the hazard increases monotonically, decreases if $r<0$ and is constant if $r=0$. We assume r to be positive given that the probability to drop out of school increases with the number of school years. It may, however, also be argued that the closer the obtained educational level is to the compulsory 9 years, the higher the motivation of the child to finish compulsory schooling may be. Regarding the significance of the coefficients β, the Wald test used in the present case is based on the null hypothesis that $\beta = 0$. Then, if one coefficient is higher than zero, this means that the covariate increases the risk to drop out of school while a coefficient lower than zero means that the covariate decreases the probability to drop out of school. According to the theoretical model, the covariates will include variables proxying for ethnic status, parental income, educational costs, labor market opportunities and preferences. Aside from these co-variates and given the usual problem of

endogeneity due to unobserved individual characteristics when analyzing education (e.g. individual learning abilities or personality), we consider a model controlling for unobserved heterogeneity called in survival analysis frailty.[13] In the case of discrete survival analysis, the proportional hazard model with unobserved heterogeneity is thus:

$$c\log\log\left[h\left(E,X|\nu\right)\right]=h_0(E)+\beta' X+u \qquad (9)$$

Where $u\equiv\log(\nu)$ characterizes the unobserved heterogeneity. We consider the Gamma distribution proposed by Meyer (1990) for ν.[14]

As already specified in the theoretical part, we want to focus on the child's school attendance until nine years of education without following them after this threshold. In view of the data structure, notably regarding the place of residence of the considered ethnic groups, we will estimate equation (9) for Guizhou and Hunan separately. Further details regarding the dataset construction are provided in the following section.

4. The CHNS longitudinal database

The database uses stems from the China Health and Nutrition Survey (CHNS).[15] The CHNS began in 1989 and provides databases for the years 1989, 1991, 1993, 1997, 2000, 2004 and 2006. We do not consider the year 1989 due to some weaknesses in the dataset for this year. Our database thus covers the years from 1991 onwards. As a micro survey, the sample size is relatively large with an average number of 14,000 individuals for each considered year. Our topic of interest and the chosen methodology, however, reduce the sample; first to individuals who either are enrolled in school or have finished school during the period of observation; second to individuals who are observed at least twice. We focus, moreover, on rural areas only. The final number of observations is consequently smaller (cf. table 1 below). Concerning the sample design, a multistage random-cluster sampling procedure was used to draw the sample from each of the provinces. Counties in the nine provinces were stratified by income (low-, middle- and high-income tertiles) with per capita income figures from the State Statistical Office. A weighted sampling scheme was used to randomly select four counties in each province (one low income, two middle income, and one high income). Probability-proportional-to-size sampling was used to select the sample from these units. The survey initially

includes 9 provinces (Guangxi, Guizhou, Henan, Heilongjiang, Hubei, Hunan, Jiangsu, Liaoning and Shandong), mostly situated in the east and center of the country. As specified in the introduction, we focus on the southwestern provinces of the CHNS dataset where many ethnic minorities are living, i.e. Guizhou and Hunan.[16] In Guizhou around 37 % of the population is from an ethnic minority group (Zhang 2003). For this province, our analysis focuses on the Bouyei, the Miao and the Tujia. Table 2 shows that in Guizhou there are altogether three autonomous prefectures and eleven autonomous counties. Zhang (2003) points out that after the settlement of the Han Chinese in the lowlands, the ethnic minorities were pulled back to the mountainous areas. In Hunan around 8 % of the population is from an ethnic minority group (Wang, 2003). Table 2 shows that in Hunan there is one autonomous prefecture and seven autonomous counties. Two-thirds of them are living in the autonomous counties of the Miao and the Tujia in western Hunan (Wang 2003). For this province our analysis focuses on the Miao ethnic minority group.

With regards to the dependent variable, educational status of individuals is given by the answers to three questions: (1) "How many years of formal education have you completed in a regular school?"; (2) "What is the highest level of education you have completed?"; (3) "Are you currently at school?". In order to approximate the year in which the child left school, we follow each individual through time. Then, we identify the periods of observation where the answer of question (3) changed from "Yes" to "No". We determine the level of education at the period when the answer is "No", which defines the failure, and thus provides the individual's final schooling attainment.

According to the theoretical model, the following covariates explaining the schooling length are considered. First, the child's ethnic status represents the main variable of interest. It is defined in two ways depending on the province of interest. In Guizhou, four dummy variables are defined equal to one for each ethnic group (Han, Miao, Bouyei and Tujia) and zero if another ethnic group is observed. In Hunan, we construct a dummy variable equal to one if the child is Han and zero if the child belongs to the Miao group. We use household income for each considered year deflated at the 2006 prices. Despite the suspicion regarding the exogeneity of this variable, no robust instruments have been identified in the dataset impeding us to control for the potential endogeneity bias.[17] To

proxy for the educational costs, we consider the distance to junior high school in kilometers.

Unfortunately, no information about tuition fees is available in the dataset. Regarding labor market opportunities, we consider the percentage of people involved in agriculture in the village as well as the daily wage for an ordinary male/female worker in the village.[18] Besides being village specific, this last variable is also gender specific: The expected wage for ordinary work of girls and boys is assumed to be the female and male daily wages of the village, respectively. In terms of parental preferences, we include mother's and father's education as well as the number of cigarettes smoked per day by the father. This has been stressed in the literature to be a relatively good proxy for time preference (Becker and Mulligan 1997).[19]

Control variables include the age and the gender of the child as well as the observation period. We remind that we carry out separate estimations for Guizhou and Hunan provinces. In Guizhou besides the Han, our sample includes information about the Bouyei, the Miao and the Tujia. Our sample for Hunan considers Han and Miao ethnic groups. We include for each province dummy variables for clusters of communities with specific ethnic structures. These dummies are summed up in table 1.

Table 1 to be included here.

For Guizhou province both dummies for Han and Bouyei are excluded. This is due to the fact that 83% of the members of the Bouyei ethnic minorities live in communities where only Bouyei are living. There would consequently be a multicolinearity problem if both the dummy variable for Bouyei ethnicity and the dummy variable for villages, where only Bouyei are living, would be included. We have thus to keep in mind while interpreting the results that in the case of the Bouyei ethnic group, we cannot clearly disentangle the community effect from the ethnic effect.

When using survival analysis, different issues must be stressed concerning the definition of the sample. The ideal data configuration is to observe the child from entering school to dropping out of school. The problem is that we do not always observe or follow individuals from the beginning to the end of the survival period. Four situations may occur. 1) The child is already enrolled in school when coming under observation, but

finishes during the period of observation. This situation is called left truncation. 2) The child enrolls in school during the period of observation but the drop out (failure) occurs after the end of the period of observation. This is called right censoring. 3) The child has finished his/her education before the period of observation. This is called left censoring. 4) The child suffers from both left truncation and right censoring when entering school before coming under observation and is not observed anymore before finishing school. In the third case, individuals cannot be included in the sample of analysis. Moreover, as described above, the structure of our dataset leads data to be collected at different points in time but not for each year. For instance, the child *k* enters in the sample while he/she has 3 years of education in 1991. Then, he/she is observed further in 1993 and 1997 with 5 and 8 years of education, respectively. In 1997, the variable, characterizing whether the child is still at school or not, informs us that the child is not at school anymore. Three conclusions can be drawn from this example which leads to various hypotheses afterwards. First, given the initial level of 3 years of education, we know that the child went to school before 1991. He/she began in 1989 and in 1990 the educational level was 2 years. Second, we know that he/she is not at school anymore in 1997. Third, given the final level of education observed in 1997, we can deduce that the child dropped out of school before 1997, most probably in 1996 (where he/she has reached 8 years of education).[20] Despite the gaps in the data, we can fully characterize the education duration profile.

For the time-varying co-variates we, however, have to follow a different procedure. To implement survival analysis, we need one observation per year for all co-variates. This requires to expand the dataset. Regarding the ethnic status, the parents' level of education (which is stable through time) and the gender, the gaps can be easily filled. The value of these variables is the same for each level of education. Missing age data of the child are adjusted yearly by considering the birthday as the basis. Assumptions, however, have to be made regarding the evolution between the periods of observation of the co-variates like household income, village characteristics (average wage, percentage of people working in agriculture, distance to school). First, regarding the above mentioned example, for left-truncated childrenwe know that the child had 3 years of education in 1991 and thus had begun his/her education in 1989. In this case, we make the assumption

that the levels of varying co-variates in 1989 and 1990 were the same as in 1991. Second, for the following time periods, we consider that the value at time *t* remains stable until the period of observation *t+1*. In the above mentioned example, this assumption means that the value of the variable x_1 in 1991, which is associated with three years of education for the individual *k*, remains stable until 1993. Our dataset is consequently expanded with stage values for the time-varying variables.[21]

5. Results

5.1. Descriptive statistics

Let us first have a look at the evolution of the average level of education for various age cohorts considering the Han, Bouyei, Miao and Tujia separately. The individuals are grouped according to their birthday in 10 year interval cohorts for the years 1920-1970. Altogether six cohorts are observed.[22]

Figure 1 to be included here.

From Figure 1, we note that despite the fact that the average level of education for the Han is mostly above the ones for any other ethnic group; curves are lying closely together. Particularly, we see that the average educational level of the Tujia is close or even higher than the average educational level of the Han for cohorts born in 1920-1930. A clearly lower average level of education is, however, observed for the Bouyei for all generations considered. Altogether, it is visible that the average level of education increased for younger cohorts among all ethnic groups. For those born between 1970-1980 we observe a convergence towards the nine-years compulsory school level for all ethnic groups.

Tables 3 and 4 provide descriptive statistics for the two provinces, distinguishing between the different ethnic groups in each province.

Table 3 and 4 to be included here.

In Guizhou province, we note a significant difference in the gender balance between the Han and other ethnic groups. Men are over-represented in the Miao and Tujia ethnic groups compared to the Han group while they are under-represented in the Bouyei ethnic

group. The sample definition linked to the construction of the dataset for our empirical implementation could be potentially responsible for such difference. Before selecting individuals with the requested characteristics such gender imbalance, however, was already observed in a larger sample of children under 16. This phenomenon is consequently either linked to a representative gender disequilibrium between the Han and the Miao in Hunan province, or a sample selection problem linked to the CHNS sample design. On average, the Han fathers have a higher level of education than those from ethnic minorities. [23] The educational level of Han fathers is, however, not larger than the educational level of Bouyei fathers. With regards to the education levels of mothers, no significant difference is observed between the Han, the Bouyei and the Miao. Tujia mothers, however, have a significantly lower level of education compared to the other ethnic groups. The income gap between the Han and the members of other ethnic groups appear also in Guizhou, particularly for the Miao and the Tujia. We observe that the lower average income level is not always related to the percentage of people involved in agriculture in the communities where different ethnic groups are living. For instance, the difference in income between the Han and the Bouyei is the lowest but the percentage of people involved in agriculture is the highest for the Bouyei. We note, moreover, that the distance to school is significantly longer for the Bouyei communities compared to other communities. Finally, differences in the fathers' smoking habits or in the average community wages are not strong between the majority and the ethnic minority groups.

Similar observations can be made in Hunan province regarding the comparison of gender, income, parents' educational levels and fathers' smoking habit between the Han and the Miao. The gender balance shows that a higher proportion of men belongs to the Miao ethnic group. This is in accordance with the observations in Guizhou province, and is also not related to the sample selection process. We observe, moreover, that the age structure of the Miao in Hunan province is similar to that in Guizhou, supporting the random structure of the sample. With regards to parents' characteristics, we observe that both Miao fathers and mothers experience a lower level of education than their Han counterparts. The difference in averages is even two years between Miao and Han mothers. The other significant difference for the Miao ethnic group relates to the average level of household income. No significant difference in the smoking behavior of the

fathers from both ethnic groups is observed. . Communities' characteristics, distances to school as well as average wage levels are similar in communities where Miao and Han are living. We note, however, a small difference in the percentage of people involved in agriculture, highlighting that the Miao are living in villages where agriculture is more present.

Generally speaking, we mainly emphasize through these descriptive statistics that on average members from ethnic minority groups have lower levels of education and income. These first observations are in accordance with the hypotheses ii. "the present parental income and the child´s expected returns to education depend negatively on the ethnic minority status" and iii. "the time preference may differ between ethnic groups" regarding the channels through which the ethnic status may influence the parental choice of their child's education.

5.2. Survival analysis

In this section, we present the results for the estimations of equation (9), distinguishing between Guizhou and Hunan provinces. We also provide the estimation of the predicted discrete hazard as a graphical analysis in order to have a look at the profile of the probability to drop-out of school as a function of the number of years of education. In order to analyze the robustness of the results of different specifications, we distinguish between five models: 1) the estimation of the complementary log-log specification without frailty; 2) the estimation of the complementary log-log specification with frailty, without community dummies. This is done to distinguish between the ethnic and the community effect taking into account the ethnic structure of the communities. 3) As in the second model plus community controls. Models 4) and 5) are the same as the second and the third one including household income additionally. We have distinguished these specific cases to observe whether or not results are changing once household income is included given that this variable is suspected of being endogenous.

The results for Guizhou province are presented in table 5. Results are stable for all models. In the models with frailty, coefficients are larger than in the model without frailty and the coefficient of the Miao dummy is even significant, emphasizing the fact that the Miao seem to face a higher probability to drop out of school than the Han in rural

Guizhou. The test for the hypothesis of null variance of frailty is rejected, leading us to focus our interpretation on the models 2 to 5.

Regarding the main variables, the Miao dummy is only significant in the second model (without considering household income and community dummies). It is, however, important to underline that the coefficient of the Miao dummy is relatively stable and high compared to the coefficients of the Bouyei and the Tujia. From our estimations, we consequently cannot conclude that members from ethnic minority groups have higher drop out rates in compulsory schooling compared to the Han in rural Guizhou. Parental education appears to be a much stronger determinant. Both the mother's and the father's education positively and significantly influence the child's school drop out before nine years of education. The coefficient of mother's education is, however, larger than the one of the father. This tends to support the hypothesis that mothers play an important role in enhancing the well-being of their children. For China, Chen and Li (2009) and Liu (2008) demonstrated the central role played by mothers with regards to the child's health. A good health status has a positive effect on learning capabilities which in return is crucial for the child's schooling performance. To our knowledge, no study, however, has yet focused on the mothers' influence on the child's level of education in China.

In most cases, including the dummy variables controlling for the community ethnic composition does not influence the results. The effect of the inclusion of household income is stronger. Particularly, once introduced in the model, the variable describing the percentage of people involved in agriculture becomes significant. Even if the sign tends to validate our theoretical model (a higher percentage of people involved in agriculture tends to discourage parents to send their children longer to school), this change associated with the inclusion of the income variable casts doubt on its exogeneity. We remind that we are not able, given the information available in the dataset, to control for the endogeneity of income. The percentage of people involved in agriculture however, is the only coefficient which is modified by the inclusion of income in models 4 and 5. The coefficient of income is significantly positive, what is not in accordance with our theoretical hypothesis 1) "the higher the parental income is, the longer the child will stay at school".

Table 5 to be included here.

It is worth noting the significantly positive coefficient of the time dependency variable (ln of education). The graphical representation (see figure 2) of the discrete estimated hazard linked to model 5 confirms this observation. The hazard increases with the number of years of education. Furthermore, the hazard profile for the Miao is below those for the Bouyei or the Tujia. Our estimations show, however, that no significant differences exist between the groups.

Figure 2 to be included here.

A last striking result observed in Guizhou province for all models is the significantly positive gender coefficient. All models indicate that boys have higher drop out rates than girls during compulsory schooling. This is a surprising result particularly in rural areas where girls are mentioned to still be dropping out of school sooner than boys (Hannum 2002). A lower probability to drop out of school for girls in comparison to boys was unexpected. It would be interesting to find out whether or not the higher drop out rates of boys are particularly linked to the nine years of compulsory education, to the geographic focus, or to the general equalizing trend in educational attainment between boys and girls.

Table 6 presents the results for Hunan province. Again, most conclusions are stable in all models.. Comparing models 1 and 2, which are the closest in terms of specification, two main differences appear: mother's education is not significant in the simple complementary log-log specification while it is with frailty; the gender dummy is significantly positive in model 1 while not anymore in model 2. We note, however, that the time dependency, which is specified as logarithm of years of education in our models, is positive and strongly significant in both cases. This means that the hazard increases with the number of years of education. The coefficient of the Han dummy, which equals 1 if the individual is classified as Han and 0 if the individual is classified as Miao, is significantly negative in the two models, supporting the idea that the probability to drop out before nine years of education is lower for the Han population compared to the Miao. The coefficient is even twice as high when frailty is taken into account. This last specification is again privileged in the following interpretations given that the test for the null of the variance of frailty is rejected, leading to consider unobserved heterogeneity as

a significant phenomena. Focusing now on models 2 to 5, three main conclusions arise. First, comparing the models 2 and 4 and the models 3 and 5, we see that results are quite stable. Variables which were significant without including household income remain so with similar coefficient ranges. Even if the income variable is endogenous, it does not seem to bias coefficients of the other variables. Second, keeping the focus on the income variable, we note a significantly positive coefficient on the logarithm of household income. Our hypothesis 1 seems consequently challenged for Hunan province as well, casting doubt on the exogenous character of this variable.

The third conclusion concerns the role of community dummies. While the dummy variable describing the ethnic status was significant when the community variables were not included, its significance vanishes once these variables are considered in the model. This occurs despite the fact that other variables are controlling for village effects; such as the percentage of people working in agriculture, average wage level in the village, and the distance to the nearest junior high school. It consequently seems that differences at the individual level between members of different ethnic groups are not emphasized in our sample. Even if the community dummies are also not significant after including both the dummies for ethnicity and for communities into the model, the results tend to support the fact that geographical location is more important than the ethnic status for the choice of additional years of schooling. When Han and Miao are living in mixed communities, lower educational levels are also observed for the Han compared to Han living in other communities. The coefficient of mother's education is significantly negative in all models 2-5, being stable when community dummies and household incomes are introduced. Contrarily to rural Guizhou, the father's education is not significant in estimations for rural Hunan.

Table 6 to be included here.

As for Guizhou province, the coefficient on the time dependence variable (ln of education) is significantly positive. Figure 3 provides the graphical representation of the discrete hazard rate obtained on the basis of model 5.

Figure 3 to be included here.

This means that with an increase in the number of school years an increasing hazard is observed. The equal drop-out rates of the Han and the Miao are also visible in figure 2.

The coefficient of the number of cigarettes smoked by the father is significantly positive. The hypothesis that a stronger smoking habit of the father indicates a stronger preference for the present and thus a lower schooling attainment of the child is thus validated. It is also interesting to note that the dummy variable characterizing the individuals' gender is never significant in Hunan province.

6. Conclusions

The Chinese educational system was subject to several policy changes during the last decades. On the one hand, the decentralization of the financing of education has led to the development of school fees which has contributed to decrease the access to education for the poorest. On the other hand, in the 11th Five year Plan (2006-2011) education equity is stated as being the key objective of the central government. Among others a more equal access to education for children from rural areas, poor families, for girls and for children of ethnic minorities is promoted. The cancellation of school fees for the nine years of compulsory education since 2006 should have contributed to reach this objective.

Without being able to observe yet the effects of the current five year plan on the educational attainment of children with different ethnical statuses, it is interesting to study whether or not the access to nine years of education was different between the ethnic majority group, the Han, and the ethnic minority groups before this strengthening. This has not, to our knowledge, been investigated until now.

To pursue this goal, we have considered the China Health and Nutrition Survey panel datasets available for periods between 1991 and 2006 to implement a survival analysis describing whether or not the ethnic status has an influence on the probability to drop out of school earlier than nine years of education. Our assessment has focused on Guizhou and Hunan provinces where in our sample the Bouyei, the Miao, and the Tujia are represented besides the Han.

The findings emphasize no significant individual differences between children from an ethnic minority group and from the Han once controls for community specific effects are

taken into consideration. This mirrors the fact that in remote and often poorer communities all ethnic groups face the same inferior schooling conditions. We can consequently conclude that inequalities in the access to the compulsory threshold of nine years of education are not mostly related to the ethnic status but more to the local conditions. Our results, moreover, emphasize the role of parental education in the validation of the nine years of education. Children of parents with higher educational levels are more probable to terminate the nine years of compulsory schooling than children of parents with lower educational levels. From our results, we can consequently conclude that the K-9 law of compulsory education can have positive effects in reducing inequalities of access to education between urban and rural areas or between mountainous and plain regions in the future. Nevertheless, this was at least not for our sample in rural Guizhou and Hunan, not a vital measure to reduce the gap between children from ethnic minorities and the majority given that such gap was filled in during the previous years. The preferential policies should more focus on poorer areas and households rather than on ethnic differences.

♣ Acknowledgment: This research uses data from the China Health and Nutrition Survey (CHNS). We thank the National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention, Carolina Population Center, the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, and R01-HD38700) and the Fogarty International Center, NIH for financial support for the CHNS data collection and analysis files from 1989 to 2006 and both parties plus the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009 and future surveys.
The authors would also like to thank Stephan Brosig, Heinrich Hockmann, Martin Petrick and the members of the International China Research Group at IAMO for useful comments and suggestions.

[1] It will later be explained that since 2006, school fees for the first nine years of primary school have been cancelled by the government. Due to the sample considered in the present study, the children included have, however, not benefited from this change.

[2] School rooms and dormitories in mountainous regions do not always benefit from the necessary heating during winter. Children are often working with their coat, bonnet and scarves. Moreover, during informal conversations with scholars and local people in rural Guizhou (Castro-Campos, forthcoming), it was reported that teachers are not motivated as in many cases they are sent to villages or townships where they lack family and community ties.

[3] Modest tuition fees were introduced in 1989 and substantial tuition fees in 1994 (Sautman, 1997, p. 11) depending on the place and university.

[4] Such obligations cover work on the field in the case of a farming household, as well as, housekeeping, child caring or paid work bringing additional income.

[5] For example, (Castro Campos, forthcoming) finds during informal conversations with foreign experts in the field of language and cultural research in Southwest China that owners of small stores and restaurants are allowed to receive help from their children. Often young girls working in restaurants are thus officially considered as family members, while they actually are not a family member but engaged in child labor. This means that they drop out during the years of compulsory school and at that time without the minimum working age level cannot apply for a legal position in a danwei. As soon as they are old enough they can receive a legal position. Illegal children, which are children without registration, also play a crucial role here.

[6] Moreover, they argue that private returns to education may be very low for some individuals based on the fact that some employers still use non-market factors (e.g. guanxi) to assign jobs to workers instead of giving the position to the most skilled and qualified worker.

[7] For example, (Castro Campos, forthcoming) was informed during informal conversations with a Han girl from Hubei province that she would have liked to study English language and literature in Hubei, but that she was not accepted there. She said that ethnic minorities who receive higher scores by law got study places instead. She then received a study place at the ethnic minority university in Duyun, Guizhou province. The tuition fees in Duyun are lower which reduced the financial burden for her parents. For this reason she finally considered studying at Duyun University to be her best choice. Another example (Castro Campos, forthcoming) comes from a couple living in Kaili. The man belongs to the Dong group and the woman belongs to the Han group. They decided that their daughter also receives the Dong status so that one day, given the additional points in the entrance examination, her chance to access university will be higher.

[8] It is worth noting that many inter-ethnic marriages exist in China. Consequently, the child's ethnic status is chosen to be either the one from the mother or the father depending on the parents' preferences. This also sometimes depends on advantages parents could obtain for their child relying on his/her declared ethnic status.

[9] This paper makes the assumption that each household only has one child. The consideration of households with more than one child is left to further studies.

[10] Inter-ethnic marriages are not considered in the model as influencing in a specific way the household's preferences.

[11] This may also include the different levels of tuition fees depending on whether the school chosen by the household is public or private.

[12] We do not consider the situation where the child works and studies at the same time.

[13] Education is intrinsically linked to the child's ability which is often unobserved. This characteristic is consequently omitted in the estimated model, leading to an omitted variable problem. In the case of the usual estimations using the number of years of education as the dependant variable, once the child has finished his/her education, controlling for this individual heterogeneity cannot be made using fixed effects as education itself is a constant variable. This problem does not occur in the present empirical implementation.

[14] We used STATA to estimate this model and particularly benefited from the program pgmhaz8 written by Stephen Jenkins (Jenkins 1997).

[15] This survey is a collaborative project between the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute of Nutrition and Food Safety at the Chinese Center for Disease Control and Prevention. For a detailed description of the dataset and overall number of observations, please refer to the web-site: www.cpc.unc.edu/projects/china.

[16] Guangxi province could have been introduced in the sample as an autonomous Zhuang minority province. The selection of the sample, however, underrepresented the Zhuang minority group in this province. We have consequently excluded this province from our sample as the selection appeared non random.

[17] Tansel (1997) and Tansel (2002) use non monetary family income, the number of adults in the household and family assets as instruments. Assets registered in the CHNS dataset concern the investment made during the year preceding the study. They can consequently be closely linked to financial income itself. The same can be said for the number of Mu used by agricultural households. The more money they have, the more able they are to rent land. Assets and area cultivated would thus be improper instrumental variables in our case.

[18] These wages are provided in the dataset at the community level. We thank the persons responsible for the CHNS dataset for having accepted to communicate us this data. The variable relates to the answer to the question "what is the daily wage for an ordinary male worker in this village/neighborhood? (yuan)".

[19] We do not consider mother's number of smoked cigarettes as most mothers in our sample do not smoke. The hypothesis behind the smoking habit proxying for time preference is that a person who smokes has a preference for the present. In our case, this would mean that a larger amount of smoked cigarettes indicates a higher preference for the present. A higher preference for the present will be visible if the child drops out of school earlier.

[20] It is worth stressing that few kids in China are repeating their school years. Due to the rarity of this phenomenon, we did not consider this as an option while working on the expansion of the dataset.

[21] Another possibility would have been to consider the interval censoring estimations. However, two issues would have arisen. First, the intervals are not of the same length through time. Some techniques allow, however, to take this point into consideration. Nevertheless, it is worth pointing out that the intervals between two observations are not the same for all individuals. Some left the sample during two periods of observation and re-entered in the sample afterwards. It was consequently not possible to create individual specific intervals.

[22] The number of observation per cohort and per ethnic group is at least 40. The highest number of observations is found for the 1940, 1950, 1960 and 1970 cohorts.

[23] Results for the test of equal average fort the different groups are available upon request.

References:

Arulampalam W., Naylor R. A. and J. Smith. 2004. A Hazard Model of the probability of Medical School Drop-out in the UK. *Journal of the Royal Statistical Society*, 167: 157-178.

Basu, K. and Z. Tzannatos. 2003. The Global Child Labor Problem: What do we know and what can we do?, *The World Bank Economic Review*, 17: 147-173.

Becker, G. S. and C. B. Mulligan. 1997. The Endogenous Determination of Time Preference, *The Quarterly Journal of economics*, 112: 729-758.

Bhalla, A. S. And S. Qui. 2006. Poverty and Inequality among Chinese Minorities. Cambridge: University of Cambridge.

Brown, P. H. and A. Park . 2002. Education and Poverty in Rural China. *Economics of Education Review*, 21: 523-541.

Card, D. 2001. Estimating the Return to Schooling: Progress on some Persistent Econometric Problems. *Econometrica*, 69: 1127-1160.

Castro Campos, B. . Forthcoming. Ethnic Minorities and Rural Employment in Guizhou Province, China (working title). Monograph.

Chapman, D., Chan, X., and G. Postiglione. 2000. Is Preservice Teacher Training Worth the Money? A Study of Teachers in Ethnic Minority Regions of the People's Republic of China. *Comparative Education Review*, 44: 300-328.

Chen, Y. and H. Li. 2009. Mother's Education and Child Health: Is There a Nurturing Effect? *Journal of Health Economics*, 28: 413-426.

Chiang, L. H. 2001. Education of Chinese Muslims: Changes in the 20th century. Paper presented at the Annual Meeting of the American Association of Chinese Studies. http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/9b/93.pdf.

De Brauw, A., and S. Rozelle. 2007. Returns to education in rural China. NY: Routledge.

Feng, A. 2005. Bilingualism for the minor or the major? An evaluative analysis of parallel conceptions in China. *International Journal of Bilingual Education and Biculturalism*, 8: 1-23.

Garcia-Aracil, A. and C. Winter. 2006. Gender and Ethnicity Differentials in School Attainment and Labor Market Earnings in Ecuador, *World Development*, 34: 289-307.

Gladney, D. C. 1995. China´s ethnic reawakening. Honolulu: East-West Center, AsiaPacific Issues, 18.

Gustafsson, B., and D. Sai. 2006. Villages where China's Ethnic Minorities Live. IZA DP No. 2418, Institute for the Study of Labor.

Gustafsson, B., and D. Sai. 2008. Temporary and Persistent Poverty among Ethnic minorities and the Majority in Rural China. IZA DP No. 3791, Institute for the Study of Labor.

Gustafsson, B., and S. Li. 2003. The Ethnic Minority-Majority Income Gap in Rural China during Transition. *Economic Development and Cultural Change*, 51: 806-822.

Hannum, E. 1999. Poverty and basic-level schooling in China: Equity issues in the 1990s. *Prospects*, 29: 561-577.

Hannum, E. 2002. Educational Stratification by Ethnicity in China: Enrollment and Attainment in the Early Reform Years. *Demography*, 39: 95-117.

Hannum, E., Wang, M. and J. Adams. Forthcoming. Urban-Rural Disparities in Access to Primary and Secondary Education under Market Reforms. In M. King Whyte, One Country, Two Societies? Rural-Urban Inequality in Contemporary China. Harvard University Press.

Hillman, B. 2008. Money can´t buy Tibetans´ Love. Far Eastern Economic Review, www.feer.com, 8-12.

Iredale, R., Bilik, N., and W. Su. 2001. Contemporary minority migration, education and ethnicity in China. Cheltenham Elgar.

Jenkins, S. 1997. Discrete Time proportional Hazards Regression. *Stat. Tech. Bull.*, 39: 22-32.

Liu, H. (2008). The Impact of Women's power on Child Quality in Rural China, *China Economic Review*, 19: 101-115.

Meyer, B. D. 1990. Unemployment insurance and Unemployment Spells, *Econometrica*, 58: 757-782.

Orazem, P. F. and E.M. King. 2007. Schooling in Developing countries: The Roles of Supply, Demand and Government Policy. In T. Schultz and J. Strauss (Ed.), Handbook of Development Economics, Vol.4, Chapter 55 (pp.3475-3559). Amsterdam, North Holland Press.

Raymond, M. and E. Sadoulet. 2003. Educational Grants Closing the Gap in Schooling Attainment between Poor and Non-Poor, UC Berkeley, WP Series n°986.

REAP. 2008. Rural Education Action Programme, http://www.reapchina.org.

Sautman, B. 1997. Affirmative action, ethnic minorities and China's universities. Paper presented at the fifth conference of the Chinese Studies Association of Australia. Hong Kong University of Science and Technology, working papers in the Social Sciences ; No. 13.

Sautman, B. 1999. Expanding Access to Higher Education for China´s National Minorities, Policies and Preferential Admissions. In G. A. Postiglione (Ed.), China´s National Minorities Education, Culture, Schooling and Devlopment (pp. 173-210). NY: Falmer Press.

Tansel, A.. 1997. Schooling Attainment, Parental education, and Gender in Cote d'Ivoire and Ghana. *Economic Development and Cultural Change,* 45: 825-856.

Tansel, A. 2002. Determinants of School Attainment of Boys and Girls in Turkey: Individual, Household and Community Factors. *Economics of Education Review,* 21: 455-470.

Wang, N. 2003. Hunan. In I. f. A. Hamburg (Ed.), Das Grosse China Lexikon (pp. 319-320). Darmstadt: Wissenschaftliche Buchgesellschaft, Darmstadt.

White Papers of the Chinese Government . 2010. http://www.china.org.cn/e-white/20050301/V.htm,http://www.china.org.cn/e-white/20050301/Appendix.htm

World Bank. 2008. Mid-Term Evaluation of China's 11th Five Year Plan, Discussion Paper, World Bank, Beijing.

World Bank. 2009. From Poor Areas to Poor people: China's Evolving Poverty Reduction Agenda, Poverty Reduction and Economic Management Department, East Asia and Pacific Report.

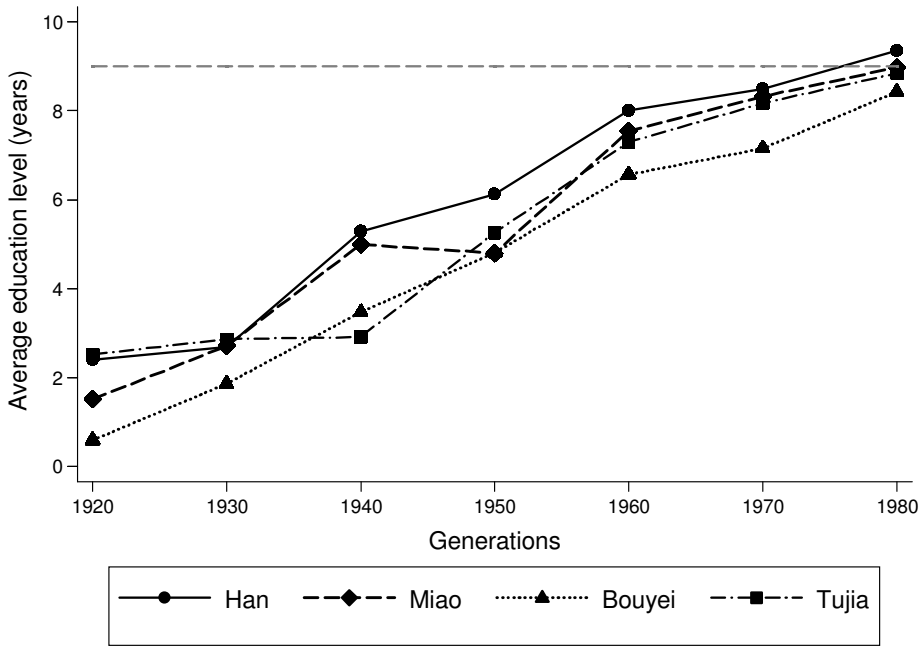
Wu, X. 2006. The impact of ethnic identity on student achievement in China: A meta analysis. Brigham Young University, Brigham, UK.

Yi, L. 2005. Muslim Narratives of Schooling, Social Mobility and Cultural Difference: A Case Study in Multi-ethnic Northwest China. *Japanese Journal of Political Science*, 6: 1-28.

Zang, X. 2008. Market Reforms and Han-Muslim Variation in Employment in the Chinese State Sector in a Chinese City. *World Development*, 36: 2341-2352.

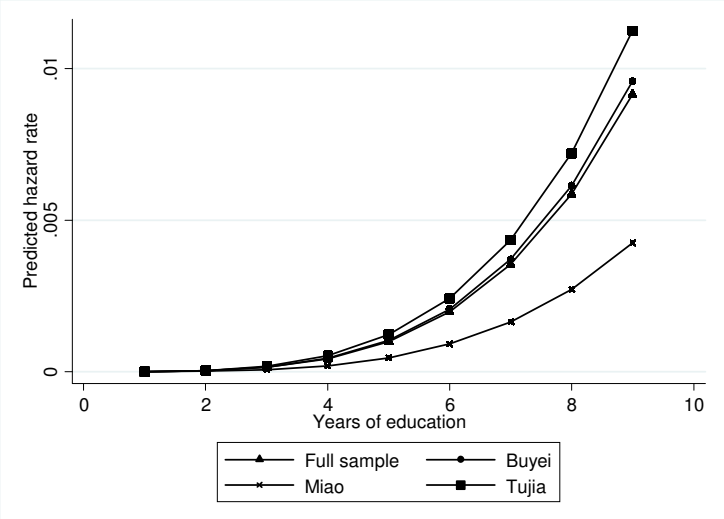
Zhang, D. D. 2003. Guizhou. In I. f. A. Hamburg (Ed.), Das Grosse China Lexikon (pp. 282-283). Darmstadt: Wissenschaftliche Buchgesellschaft, Darmstadt.

Zhang, L. & Huang, J. and S. Rozelle. 2002. Employment, Emerging Labor Market, and the Role of Education in Rural China, China Economic Review, 13: 313-328.



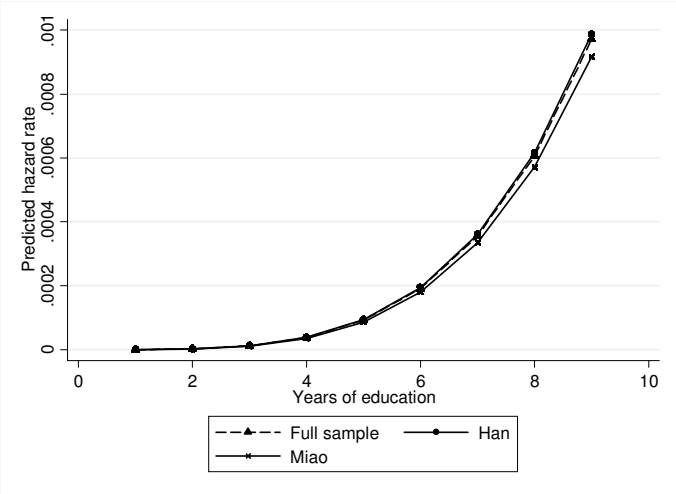
Source: Authors calculation based on CHNS sample

Figure 1: Evolution of the average level of education by ethnic groups.



Source: Authors calculation based on CHNS sample

Figure 2: Predicted discrete hazard rates for Guizhou province. Comparison between the Bouyei, the Miao and the Tujia.



Source: Authors calculation based on CHNS sample

Figure 3: Predicted discrete hazard rates for Hunan province. Comparison between the Han and the Miao.

Table 1: Ethnic groups per community dummy variables.

	Number of communities, Number of subjects per ethnic group	Percentage of total ethnic groups' populations
Guizhou		
Excluded dummies: Communities including only Han	5 communities; 103 Han	67 % of total Han
Excluded dummies: Communities including only Bouyei	3 communities; 86 Bouyei	83 % of total Bouyei
Communities only with Tujia	2 community; 35 Tujia	69 % of total Tujia
Communities mixed including Han	4 communities; 26 Han, 20 Miao, 18 Bouyei, 16 Tujia	17 % of total Han; 54 % of total Miao; 17 % of total Bouyei; 31% of total Tujia
Communities Han and Miao	2 communities; 25 Han; 17 Miao	16% of total Han; 46% of total Miao
Hunan		
Excluded dummy: Communities including only Han	13 communities; 230 Han	91.6 % of total Han
Communities Miao	1 community, 28 Miao	48.3 % of total Miao
Communities Mixed	3 communities; 21 Han, 30 Miao, 1 other	8.4 % Han; 51.7 % Miao

Source: Authors calculation based on CHNS sample

Table 2: Ethnic Autonomous Areas in Guizhou and Hunan provinces.

Name of Ethnic Autonomous Area	Time of Founding	Capital	Area (square km)	Population (thousand)	Proportion of Ethnic Minority Population in the Total (%)
Autonomous prefectures					
Hunan Province					
Xiangxi Tujia-Miao	Sept. 20, 1957	Jishou City	15,461	2,655.5	74.59
Guizhou Province					
Qiandongnan Miao-Dong	July 23, 1956	Kaili City	30,337	4,193.8	77.10
Qiannan Bouyei-Miao	Aug. 8, 1956	Duyun City	26,193	3,790.1	55.28
Qianxinan Bouyei-Miao	May 1, 1982	Xingyi City	16,804	3,016.2	42.94
Autonomous counties					
Hunan Province					
Chengbu Miao	Nov. 30, 1956	Rulin Town	2,620	257.2	57.59
Tongdao Dong	May 7, 1954	Shuangjiang Town	2,225	221.1	88.50
Jianghua Yao	Nov. 25, 1955	Tuojiang Town	3,216	458.3	63.97
Xinhuang Dong	Dec. 5, 1956	Xinhuang Town	1,511	250.5	87.56
Zhijiang Dong	Sept. 24, 1987	Zhijiang Town	2,096	356.9	61.25
Jingzhou Miao-Dong	Sept. 27, 1987	Quyang Town	2,211	260.1	73.00
Mayang Miao	April 1, 1990	Gaocun Town	1,561	359.8	77.97
Guizhou province					
Songtao Miao	Dec. 31, 1956	Liaogao Town	2,861	639.9	42.49
Zhenning Bouyei-Miao	Sept. 11, 1963	Chengguan Town	1,721	334.6	58.61
Ziyun Miao-Bouyei	Feb. 11, 1966	Songshan Town	2,284	322.4	68.44
Weining Yi-Hui-Miao	Nov. 11, 1954	Caohai Town	6,296	1,095.9	25.37
Guanling Bouyei-Miao	Dec. 31, 1981	Guansuo Town	1,468	320.0	58.99
Sandu Shui	Jan. 2, 1957	Sanhe Town	2,383	314.7	96.85
Yuping Dong	Nov. 7, 1984	Pingxi Town	516	136.8	82.70
Daozhen Gelao-Miao	Nov. 29, 1987	Yuxi Town	2,156	336.6	79.18
Wuchuan Gelao-Miao	Nov. 26, 1987	Duru Town	2,773	419.3	96.25
Yinjiang Tujia-Miao	Nov. 20, 1987	Yinjiang Town	1,961	399.4	71.36
Yanhe Tujia	Nov. 23, 1987	Heping Town	2,469	558.7	55.74

Source: http://www.china.org.cn/e-white/20050301/Appendix.htm

Table 3: Descriptive statistics Guizhou province.

Samples	Full sample	Han	Bouyei	Miao	Tujia
Subjects	346	154	103	38	51
Age	12.3795	12.6151	12.2171	11.9384	12.2682
	(2.7401)	(2.9312)	(2.5356)	(2.6864)	(2.4894)
Gender (%, Male=1)	53.55	52.27	50.24	58.22	60.45
	(49.89)	(49.99)	(50.06)	(49.49)	(49.01)
Father's level of education (years)	6.6170	6.9015	6.8488	6.1986	5.6091
	(2.9560)	(2.8194)	(2.8192)	(3.6191)	(2.8878)
Mother's level of education (years)	3.9875	4.0606	4.3732	4.8288	2.4909
	(3.6692)	(3.6504)	(3.7338)	(3.7972)	(3.0855)
Father's Number of cigarettes per day	13.4659	15.3970	11.5146	10.7945	13.0818
	(9.7587)	(9.7157)	(9.0002)	(8.8738)	(10.6429)
Household's yearly income (yuan)	7,254.03	8,335.51	7,181.89	6,585.21	4,602.66
	(8,234.93)	(8,092.41)	(10,405.72)	(5,716.78)	(3,538.93)
Average distance to lower middle schools (km)	2.3892	1.4717	4.9732	1.3734	1.0004
	(5.7741)	(1.8846)	(9.9893)	(2.0323)	(1.0116)
Average wage in the village (yuan per day)	9.6667	9.3470	10.4951	9.8931	8.9318
	(10.4573)	(11.7135)	(11.1851)	(5.2801)	(7.0290)
Percentage of people involved in agriculture in the village (%)	57.70	50.77	67.12	66.14	55.37
	(27.75)	(28.69)	(20.87)	(13.80)	(35.55)

Source: Authors calculation based on CHNS sample, Incomes are deflated based on the 2006 prices. Figures in parentheses are the standard deviations.

Table 4: Descriptive statistics Hunan Province.

Samples	Full sample	Han	Miao
Subjects	310	251	58
Age	12.1496	12.1353	12.1781
	(2.4595)	(2.6232)	(2.6232)
Gender (%, Male=1))	50.83	48.49	63.47
	(50.01)	(50.00)	(48.26)
Father's level of education (years)	7.5187	7.6972	6.6758
	(2.5787)	(2.4620)	(2.9816)
Mother's level of education (years)	6.3206	6.6991	4.6849
	(3.1933)	(3.0640)	(3.2785)
Father's Number of cigarettes per day	14.6189	14.7770	14.0639
	(11.4703)	(11.8299)	(9.6567)
Household's yearly income (yuan)	8,034.164	8,647.785	5,452.74
	(9,144.846)	(9,700.97)	(5,762.80)
Average distance to lower middle schools (km)	0.8968	0.8748	0.8776
	(1.2084)	(1.0955)	(1.5328)
Average wage in the village (yuan per day)	9.5807	9.6387	9.5479
	(7.9536)	(8.5240)	(4.5897)
Percentage of people involved in agriculture in the village (%)	58.89	57.85	62.68
	(30.87)	(31.03)	(29.97)

Source: Authors calculation based on CHNS sample, Incomes are deflated based on the 2006 prices. Figures in parentheses are the standard deviations.

Table 5: Factors influencing the probability to drop out before nine years of compulsory education. Estimations for Guizhou province.

	Without control for frailty	With control for frailty			
	(1)	(2)	(3)	(4)	(5)
Ln(education)	1.7090 ***	3.9640 ***	3.9450 ***	3.8463 ***	3.7924 ***
	(0.4939)	(0.9327)	(0.9475)	(0.8853)	(0.8964)
Bouyei	0.0169	-0.2180	-0.1471	-0.1006	-0.0119
	(0.2359)	(0.3849)	(0.4088)	(0.3430)	(0.3654)
Miao	-0.6523	-1.0451 *	-1.0698	-0.7555	-0.8285
	(0.3997)	(0.6231)	(0.7121)	(0.5650)	(0.6274)
Tujia	0.3414	-0.0345	0.0204	0.0749	0.1491
	(0.2556)	(0.4177)	(0.4382)	(0.3820)	(0.3990)
Labor market characteristics					
% people working in agriculture	0.0064	0.0101	0.0090	0.0135 **	0.0130 **
	(0.0042)	(0.0064)	(0.0069)	(0.0061)	(0.0065)
Village average wage (yuan per day)	-0.0099	0.0060	0.0075	0.0043	0.0057
	(0.0141)	(0.0138)	(0.0140)	(0.0141)	(0.0142)
Distance to junior high school	0.0161	-0.0121	-0.0118	-0.0137	-0.0127
	(0.0143)	(0.0231)	(0.0228)	(0.0241)	(0.0234)
Household characteristics					
Mother's education (years)	-0.1627 ***	-0.2480 ***	-0.2582 ***	-0.2297 ***	-0.2414 ***
	(0.0356)	(0.0753)	(0.0860)	(0.0656)	(0.0705)
Father's education (years)	-0.0837 **	-0.1684 ***	-0.1669 **	-0.1589 ***	-0.1551 ***
	(0.0369)	(0.0646)	(0.0682)	(0.0607)	(0.0631)
Father's cigarette consumption (unit)	0.0143	0.0220	0.0211	0.0216 **	0.0207
	(0.0095)	(0.0140)	(0.0148)	(0.0131)	(0.0137)
Log Household's income (yuan per year)				0.3971 **	0.4150 ***
				(0.1895)	(0.1960)
Controls					
Age	0.3604 ***	0.4401 ***	0.4566 **	0.3529 ***	0.3683 ***
	(0.0564)	(0.1495)	(0.1840)	(0.1218)	(0.1367)
Gender (0 if female)	0.6983 ***	0.5813 *	0.5930 *	0.5520 **	0.5578 **
	(0.1901)	(0.3085)	(0.3143)	(0.2817)	(0.2864)
Communities mixed with Han			-0.0930		-0.0036
			(0.4106)		(0.3826)
Communities Han Miao			0.2854		0.3844
			(0.5901)		(0.5137)
Model diagnostic					
Ho: all parms except const=0					
Likelihood ratio		-206.76	-206-56	-202.40	-202.09
P-value		0.000	0.000	0.000	0.000
Ho: Variance of frailty=0					
LR-statistic		173.71	171.10	158.69	158.01
P-value		0.000	0.000	0.000	0.000
Observations/state changes	1436/140	1436/140	1436/140	1426/137	1426/137

Source: Authors calculation based on CHNS sample. Period 1991-2006. Children living with their parents. Coefficients reported, standard errors in parenthesis. *** significant at p<0.01; ** p<0.05; *p<0.1. Dummies for time periods are included. The year is taken as reference. Income inflated to 2006 prices.

Table 6: Factors influencing the probability to drop out before nine years of compulsory education. Estimations for Hunan province.

	Without control for frailty	With control for frailty			
	Model 1	Model 2	Model 3	Model 4	Model 5
Ln(education)	2.5949***	3.8603***	4.2969 ***	3.4763 ***	4.0030 ***
	(0.8125)	(1.3584)	(1.5130)	(1.3294)	(1.4892)
Han (Han=1; Miao=0)	-0.5149 *	-1.0913 **	-0.0363	-1.3064 **	0.0748
	(0.2938)	(0.5510)	(1.0411)	(0.5459)	(1.0343)
Labor market characteristics					
% people working in agriculture	-0.0033	-0.0127	-0.0143	-0.0101	-0.0114
	(0.0046)	(0.0084)	(0.090)	(0.0081)	(0.0087)
Village average wage (yuan per day)	-0.0054	-0.0099	-0.0127	-0.0438	-0.0522
	(0.0226)	(0.0366)	(0.0381)	(0.0391)	(0.0417)
Distance to junior high school	0.0570	-0.0027	-0.0173	0.0710	0.0684
	(0.0964)	(0.1558)	(0.1772)	(0.1501)	(0.1701)
Household characteristics					
Mother's education (years)	-0.0688	-0.2204 ***	-0.2212 ***	-0.2175 ***	-0.2149 ***
	(0.0428)	(0.0818)	(0.0857)	(0.0784)	(0.0826)
Father's education (years)	-0.0586	-0.0870	-0.0943	-0.0540	-0.0558
	(0.0532)	(0.0984)	(0.1017)	(0.0948)	(0.0982)
Father's cigarette consumption (unit)	0.0429 ***	0.0433 **	0.0438 **	0.0419 **	0.0433 **
	0.0104	0.0175	0.0180	0.0172	0.0179
Log Household's income (yuan per year)				0.5116 **	0.5827 **
				(0.2830)	(0.3011)
Controls					
Age	0.3423 ***	0.7166 ***	0.7007 ***	0.7539 ***	0.7470 ***
	(0.0843)	(0.1786)	(0.1816)	(0.1776)	(0.1834)
Gender (0 if female)	0.4434 *	0.6322	0.6483	0.6973	0.7161
	(0.2515)	(0.4405)	(0.4601)	(0.4353)	(0.4603)
Communities Mixed			1.2403		1.6130
			(0.9980)		(1.0288)
Communities Miao			1.2940		1.7746
			(1.2998)		(1.3189)
Model diagnostic					
Ho: all parms except const=0					
Likelihood ratio	-208.37	-159.67	-158.79	-155.44	-153.99
P-value	0.000	0.000	0.000	0.000	0.000
Ho: Variance of frailty=0		97.40	94.58	93.10	88.99
LR-statistic		0.000	0.000	0.000	0.000
P-value					
Observations/state changes	1257/77	1257/77	1257/77	1226/77	1226/77

Source: Authors calculation based on CHNS sample. Period 1991-2006. Children living with their parents. Coefficients reported, standard errors in parenthesis. *** significant at p<0.01; ** p<0.05; *p<0.1. Dummies for time periods are included. The year is taken as reference. Income inflated to 2006 prices.