AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Studies on the Agricultural and Food Sector
in Central and Eastern Europe

Bente Castro Campos

Human capital differences or labor market discrimination?

The occupational outcomes of ethnic minorities in rural Guizhou (China)

IAMO
LEIBNIZ-INSTITUT FÜR AGRARENTWICKLUNG
IN MITTEL- UND OSTEUROPA

Human capital differences or labor market discrimination?
The occupational outcomes of ethnic minorities in rural Guizhou (China)

Studies on the Agricultural and Food Sector
in Central and Eastern Europe

Edited by
Leibniz Institute of Agricultural Development
in Central and Eastern Europe
IAMO

Volume 73

Human capital differences or labor market discrimination?

The occupational outcomes of ethnic minorities in rural Guizhou (China)

by
Bente Castro Campos

IAMO
2013

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists the publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the internet at: http://dnb.ddb.de.

This thesis was accepted as a doctoral dissertation in fulfillment of the requirements for the degree "doctor agriculturarum" by the Faculty of Natural Sciences III at Martin-Luther-University Halle-Wittenberg on 14.11.2012.

Date of oral examination:	29.04.2013
Supervisor and Reviewer:	PD Dr. Martin Petrick
Co-Reviewer:	PD Dr. Ildikó Bellér-Hann

Diese Veröffentlichung kann kostenfrei im Internet unter *<www.iamo.de/dok/sr_vol73pdf>* heruntergeladen werden.

This publication can be downloaded free from the website *<www.iamo.de/dok/sr_vol73pdf>*.

© 2013
Leibniz-Institut für Agrarentwicklung in Mittel- und Osteuropa (IAMO)
Theodor-Lieser-Straße 2
06120 Halle (Saale)
Tel.: 49 (345) 2928-0
Fax: 49 (345) 2928-199
e-mail: iamo@iamo.de
http://www.iamo.de
ISSN 1436-221X
ISBN 978-3-938584-76-7

ACKNOWLEDGEMENTS

This monograph is the final outcome of my graduate work at the Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO) in Halle (Saale), Germany. I would like to express my deepest gratitude to everyone who contributed in one way or another to the success of this monograph. I particularly thank my supervisor, Martin Petrick, for stimulating discussions and instructive comments at various stages of my PhD studies. I also would like to express my deepest gratitude to Ildikó Bellér-Hann (University of Copenhagen) and to thank Thomas Glauben for helpful comments. In addition I am extremely grateful to Daphne Katranides (Leipzig University) for language editing and outstanding comments.

I also want to thank researchers and administrative staff of Guizhou University and Heinz P. Galler (Martin-Luther University Halle-Wittenberg) for fruitful discussions and supportive comments. My thanks, moreover, go to my colleagues of the China international research group at IAMO: Stephan Brosig, Bernhard Brümmer (University of Göttingen), Jens Frayer, Thomas Herzfeld, Lili Jia, Michael Kopsidis, Kelly Labart (FERDI), Karin Kataria, Daniel Müller, Tursinbek Sultan, Zhanli Sun, Xiaobing Wang (CCAP), Xiaohua Yu (University of Göttingen), Yanjie Zhang and Qiran Zhao for their helpful comments and suggestions throughout the duration of this work; moreover, I thank all other colleagues at IAMO for the supportive working environment. Special thanks go to Balgah Roland Azibo, Kinga Boenning, Gertrud Buchenrieder, Kathrin Happe, Gabriele Mewes and Nadine Wettstein.

This monograph uses data from the China Health and Nutrition Survey (CHNS). I 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 financial support from the 2008 "Pakt für Innovation und Forschung des Bundes und der Länder" is gratefully acknowledged.

Last but not least, I am extremely grateful to my husband, Iván, to my family and to my friends for their support and encouragement.

Bente Castro Campos

Hong Kong, November 2012

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PICTURES

LIST OF MAPS

LIST OF ABBREVIATIONS

AIC	Akaike information criterion
AR	Autonomous Region
BIC	Bayesian information criterion
CCP	Chinese Communist Party
CHNS	China Health and Nutrition Survey
CITS	China International Travel Service
CNY	Chinese Yuan, Renminbi
D-N	deductive-nomological
EU	European Union
EUR	Euro
FN	Field notes
GEV	Generalized extreme value
GTAI	Germany Trade and Invest
HRS	Household Responsibility System
IAMO	Leibniz Institute of Agricultural Development in Central and Eastern Europe
IIA	independence from irrelevant alternatives
IID	independent and identically distributed
IV	independent variable
LR	Likelihood ratio
MNL	Multinomial logit model
MNP	Multinomial probit model
NFRE	Non-farm rural employment
OLOGIT	Ordered logit model
OLS	Ordinary least squares
PhD	Doctoral degree
REAP	Rural Education Action Project

SD	Standard deviation
SLF	Sustainable Livelihood Framework
SIL	Summer Institute of Linguistics
SUEST	Seemingly unrelated estimation
TAR	Tibetan Autonomous Region
US	United States of America
USD	United States dollar ($)
WTO	World Trade Organization
XUAR	Xinjiang Uyghur Autonomous Region

List of important symbols

A	agricultural occupations, majority worker, white worker
a	innate ability
α	job satisfaction element, set of coefficients
ß	estimated coefficient
B	minority worker, black worker
BC	blue-collar occupations
C	set of occupations, cost
c	job search cost, cost
χ^2	Chi^2 statistic
D	index of occupational dissimilarity
d	disutility, compensation
δ	farm specific characteristic, marginal effect
e	accumulated human capital
E	ethnic minority worker, expected value
ε	error term
$\in$	element of
f	females, wives
$f(\cdot)$	density function, production function
F	full income constraint
$F(\cdot)$	cumulative distribution
G	marginal product of labor
g	group of workers
γ	proportion of majority workers
Γ	household characteristics
h	home
H	Han, hypothesis

I	indifference curve
i	individual, additional income
j	occupations, job offer, individual
k	base category
K	amount of companies
L	labor force
λ	productivity
$\Lambda(\cdot)$	logistic cumulative distribution function
m	males, husbands
MP	marginal productivity
μ	unexplained part in wage equation, cut-points in OLOGIT model
N	aggregated flow of labor
n	number of hired workers
NA	non-agricultural occupations
p	prejudiced employer
P	proportion, probability
$Prob$	probability
π	profit, prior belief
ϕ	probability density function of the normal distribution
ψ	probability point in the normal distribution
q	chosen occupation, output
s	threshold value
S	supply curve
soc	secondary occupation
t	time
T	total time
TPP	total production possibility frontier
θ	job preference, proportion, vague signal, productivity indicator
$\hat{\theta}_1$	consistent estimator
$\hat{\theta}_2$	efficient estimator

U	utility
u	unprejudiced employer
V	observable utility
$\hat{\mathbf{V}}$	estimates of asymptotic covariance matrice
W	real wage rate
w	wage, explanatory factor
WC	white-collar occupations
X	explained part of wage equation, agricultural input cost, institutional constraint, characteristic of occupation
x	vector of explanatory variables, net return, input
$\bar{x}$	mean value
Y	monetary income, dependent variable
Z	production function, Z-goods

MONETARY EQUIVALENCE

1 € = 8.997500 Chinese yuan, Renminbi (CNY), average annual exchange rate in 2011

1 € = 1.392189 United States dollar (USD) average annual exchange rate in 2011

1 INTRODUCTION

Ethnic conflicts in Tibet Autonomous Region (TAR) and Xinjiang Uyghur Autonomous Region (XUAR) have been under increasing scrutiny from the West. Today even online shops offer free-Tibet products, and solidarity movements for Tibetans and Uyghurs have sprung up in many major cities in the EU and the US. The West frequently points the finger at China about human rights violations by Chinese security forces against Tibetans and Uyghurs. The West has even considered economic sanctions against Chinese products due to uprisings in Tibetan monasteries (FAZ, 2008). It is striking that ethnic issues in these two autonomous areas have an increasingly international dimension.

One major reason for increasing ethnic tensions in TAR and XUAR is discrimination against Tibetans and Uyghurs in the labor market and the preferences for Han-majority workers in many sectors. GILLEY (2001) finds that in XUAR many jobs are given to Han rather than to Uyghurs. In TAR even jobs in the tourism industry, which is devoted to Tibetan culture, are taken by Han (HILLMAN, 2008). While the roles of Tibetans and Uyghurs in the labor market are widely discussed, there is only limited information about many other ethnic minorities in China, which officially counts 55 ethnic minority groups alongside the Han-majority. These ethnic groups make up only around 8 % of the population, yet in absolute numbers they are roughly 104.49 million people.

It is interesting to note that regarding the poorest area of China, Guizhou province, the labor market literature contradicts received wisdom. Ethnic minorities, which make up 37 % of the population in Guizhou, have higher increases in per capita income than Han between 1988 and 1995 (GUSTAFSSON and LI, 2003). A labor market analysis which considers ethnic minorities separately from each other has yet to be conducted for Guizhou province. Given that in Guizhou the employed persons in rural areas account for roughly 79 % of the total labor force in comparison to 21 % in urban areas, this monograph attempts to investigate the role of the major ethnic minorities, the Bouyei, Miao and Tujia, in comparison to Han in Guizhou's rural labor market.

The objective of the first subchapter is to introduce the differences in labor market outcomes among ethnic groups in China and to acquaint the reader with some of the major determinants of these differences, notably ethnic status, human capital factors and geographic location. A frequently cited reason for lower labor market participation and higher rates of many ethnic minorities in agricultural employment is their comparatively lower educational attainment. Subchapter 1.2, therefore, particularly addresses educational differences among ethnic groups in China.

The Chinese government recognized that ethnic minorities have been increasingly excluded from the job market and has taken several measures to overcome these inequalities. Subchapter 1.3 gives an overview of current preferential policies targeted at ethnic minorities and local governments in autonomous areas. Subchapter 1.4 then introduces the definitions of labor market discrimination and job preferences as well as presents the central research questions of this monograph. Subchapter 1.5 highlights the main scientific controversies. The last subchapter, 1.6, presents the organization of this monograph.

1.1 Ethnic differences in labor market outcomes

Received wisdom says that members of ethnic minorities should be consistently less affluent and less integrated into economic development processes. I intend to shed some light on this issue and postulate that more than one conclusion can be drawn for China overall. To highlight heterogeneous findings I first introduce the situations of the Uyghurs and Tibetans, given that there is evidence that these two groups face labor market discrimination, before I present contradicting results for the Bouyei and Miao in Guizhou, the Koreans in Northeastern China and the Hui minority group spread throughout China among others.

In XUAR there is evidence that agriculture is the main income source of ethnic minorities, while their participation in all other employment sectors is very low (HANNUM and XIE, 1998). James Millward from Georgetown University in Washington, D.C. says that: "Uighurs are simply not hired by Chinese firms. At job fairs, "Uighurs need not apply" signs are standard" (GILLEY, 2001, p. 2). This is consistent with the findings of HOPPER and WEBBER (2009) who find that in Urumqi, the capital of XUAR, many Uyghur migrants only have access to those jobs on which Han look down. The authors find that the labor market is segregated with nearly 90 % of Uyghurs and Han having a boss and co-workers of their same ethnic group. They conclude that for Uyghur migrants the lack of social capital and of network resources are responsible for the disparity in job access between Uyghur and Han migrants in Urumqi. There are many employment sectors in which, because of the strong guanxi relations, which means through connections, among Han, Uyghurs are often not hired; this is particularly true in the construction sector (HOPPER and WEBBER, 2009). Uyghurs have, therefore, established a niche market for demolition, where Han are usually not hired due to their inferior physical strength (HOPPER and WEBBER, 2009). Similar observations are made by BECQUELIN:

> To an observer in Xinjiang, however, it is obvious that fresh waves of Han migrants have been pouring into Urumqi and every other city. Construction work, restaurant employment, hawking and a wide array of petty jobs are taken up by people coming in from Qinghai, Henan and Sichuan − or even further afield. (BECQUELIN, 2000, p. 75)

In an article published in the THE ECONOMIST it was even reported that: "Look", says a young Uighur in Urumqi, "I am a strong man, and well-educated. But Chinese

firms won't give me a job. Yet go down to the railway station and you can see all the Chinese who've just arrived. They'll get jobs. It's a policy, to swamp us" (THE ECONOMIST, 2000, p. 2). GLADNEY (2004), however, also finds that in XUAR science and technology positions are filled by Uzbeks and Tartars instead of by Uyghurs because Uzbeks and Tartars make up a high percentage of the well-educated urban population. It is clearly not simply a majority-minority story in XUAR, but it is much more complex.

Tibetans in TAR face a similar situation to that of Uyghurs in XUAR. There are masses of Han-Chinese job seekers pouring into TAR, often better educated and with better Mandarin skills than their Tibetan counterparts, facilitated by a railway constructed in the summer of 2006 which connects Lhasa and Beijing (KUPFER, 2011a). The economic disparity between Tibetans and Han has been increasing. While Tibetans mainly work in the countryside or as animal breeders, Han mainly work in the administration and the service sectors (KUPFER, 2011a). The author states that the private sector, where salaries are highest, is owned and controlled by Han. Young Tibetans from Lhasa reported that they receive less than a third of Han wages for the same work (KUPFER, 2011a). HILLMAN (2008) stresses that Tibetans are disadvantaged compared to better educated migrants from other provinces, even in the tourism industry, which is mainly devoted to Tibetan culture.

Contradictory results are reported for the Miao, Bouyei and Koreans. In Guizhou and Yunnan provinces in Southern China, GUSTAFSSON and LI (2003) find that per capita income among ethnic minorities increased more than that of Han between 1988 and 1995. The authors explain this phenomenon is due to the fact that ethnic minorities in these provinces particularly benefit from liberalized border trade and a growing tourism industry (BHALLA and QUI, 2006). MAURER-FAZIO et al. (2004, 2005) observe higher labor market participation for the Bouyei and for the Miao than for Han in the 2000 and 1982 censuses, respectively. This is, however, not observable for all ethnic minorities in Guizhou and Yunnan (see HARRELL, 1990, 1995; MCKHANN, 1995). Koreans in Northeast China, who have more years of schooling than the national average, most benefitted from increasing trade between China and South Korea (GLADNEY, 2004, p. 21). Increasing numbers of tourists in Guizhou, Yunnan and the Silk Road in Xinjiang significantly contribute to local economies in these areas (GLADNEY, 2004, p. 21). Also more and more tourists visit Inner Mongolia to see ethnic lifestyles in the wide grasslands (SCHOMANN, 2011).

Another story can be observed for the Hui. They are the largest Muslim group in China, generally speak Mandarin or local dialects and reside throughout the country. While on average more than 80 % of each ethnic minority group in XUAR works in agriculture and husbandry, for the Hui this percentage is 61 % (GLADNEY, 2004). The Hui are engaged in all kinds of activities. GLADNEY states that there is an:

> extensive economic and occupational diversity […] among the Hui, from cadres to clergy, rice farmers to factory workers, schoolteachers to camel drivers, and poets to politicians. In

> the north, the majority of Hui are wheat and dry-rice agriculturists, while in the south, they are primarily engaged in wet-rice cultivation and aquaculture. (GLADNEY, 2004, p. 188)

In urban Hui communities, for example in Oxen Street in Beijing, they often have restaurants and work in niche markets (GLADNEY, 2004, p. 167). Because of "their traditional occupations as small merchants, restaurateurs, butchers, and jewellery craftsmen," the Hui are considered the "Jews of China" (PILLSBURY, 1973 cited in GLADNEY, 2004, p. 167-168) and in many places have a higher distribution in small private businesses and industry than Han (GLADNEY, 2004, p. 286). Also in Fujian the Hui were able to improve their economic situation, particularly through governmental support and their entrepreneurial skills (GLADNEY, 2004). In Shaanxi, Gansu and Ningxia the Hui received lower quality land in comparison to Han and were, hence, forced to develop their entrepreneurial skills (GLADNEY, 2004, p. 292-293). In contrast ZANG (2008) finds that Hui are discriminated against in state employment in Lanzhou (Gansu).

I highlighted heterogeneous labor market situations of some ethnic minorities in China. I showed that the analysis of labor market discrimination and occupational differences requires carefully distinguishing between ethnic groups, occupational segments, regions and time; this analysis also has to take other individual characteristics into account. One fundamental characteristic is educational attainment, which will be addressed in the next subchapter.

1.2 Ethnic differences in educational attainment

Educational attainment is one major employment criterion. Those individuals with more years of education have on average higher chances to work in better paying positions. Although the average education level of 14 ethnic minorities, including the Korean, Manchu, Mongolian and Kazak groups, is higher than the national average (CHINA.ORG.CN, 2005), there are still 41 ethnic minorities with below average levels. Differences in school drop-out rates between ethnic groups find their origin in various factors: school availability, quality and costs, Mandarin language skills, gender and opportunity costs of households.

First, access to education for ethnic minorities is often restrained due to their remote rural location and their higher levels of poverty (GUSTAFFSON and SAI, 2008). Second, higher level and higher quality schools are often situated in cities and are, thus, more difficult for pupils from remote areas to reach. Third, higher educational achievements are closely linked to a fluent command of Mandarin; this is perhaps to be expected, given that Mandarin is the main language of instruction in China. Ethnic minorities, who mainly speak their own languages at home, often have weaker Mandarin language skills than do Han. Fourth, school fees are a barrier to access to education for poorer households among which ethnic minorities in Southern and Western China are often found (cf., GUSTAFFSON and SAI, 2006, 2008). In addition to this lower access to education, some ethnic minorities face higher opportunity costs of education than do Han. For example, as already stressed,

statistically Tibetans and Uyghurs are less likely to work in off-farm jobs (cf., HILLMAN, 2008; GILLEY, 2001). Returns on higher education, thus, may be valued as too low by ethnic minority parents, which keep them from investing further in their children's education. Ethnic minorities, who often work in agriculture in remote areas, benefit from the child's assistance on the farm. Sending the child to school means losing a helping hand in doing household chores; this implies a lower household income in the short run. Many ethnic minorities are often affected by all of these factors which limit access to schooling and employment. The Chinese government has implemented preferential policies for ethnic minorities to overcome these inequalities which are directed to individuals and to autonomous areas.

1.3 Preferential policies

With the foundation of the People's Republic of China in 1949, the government started to reclassify ethnic groups and autonomous areas and progressively set up a preferential policy framework. Fifty-five ethnic minorities alongside the Han-majority have been classified since 1949. Actually China has implemented a very advanced preferential policy framework to tackle labor market discrimination and ethnic inequalities, but one look into TAR or XUAR tells us how poorly these policies have been enforced.

Ethnic minorities enjoy preferential rights (SAUTMAN, 1997) and are protected by law against, among others, labor market discrimination (ROSS et al., 2007). The preferential policies include:

> family planning (exemption from the minimum marriage age and one-child strictures); education (preferential admissions, lowered school fees, boarding schools, remedial programs); employment (extra consideration in hiring and promotion of cadres); business development (special loans and grants, exemptions from certain taxes); and political representation (proportionate or greater numbers of minorities in "people's congresses" and among minority area leaders). (SAUTMAN, 1997, p. 3)

In January 2008 the Employment Promotion Law (中华人民共和国就业促进法) came into force (ROSS et al., 2007). The Employment promotion law covers non-discrimination measures:

> The law expressly provides that workers have equal rights to employment and to establish businesses in accordance with law without respect to ethnicity (nationality in Chinese), race, gender, religious belief or other characteristics (Article 3), and may not be discriminated against in hiring or in the conditions of their employment (Article 26). Similar prohibitions on discrimination were included in the 1994 Labor Law (Articles 12-14), but they were less comprehensive and widely disregarded. (ROSS et al., 2007, p. 2-3)

The Chinese government has also taken several measures to improve autonomous areas. Of the 55 ethnic minorities in China, 44 have their own designated autonomous areas. The population of ethnic minorities practicing regional autonomy accounts for 71 % of the total ethnic minority population, and the area where

such regional autonomy is practiced accounts for 64 % of the physical territory of China (CHINA.ORG.CN, 2005). In these autonomous areas the local ethnic minorities have the right of self-governance. According to the White Papers of the Chinese Government, this implies that local governments in ethnic autonomous areas are granted the rights to:

> Independently managing the ethnic group's internal affairs in its autonomous area; the right to formulate self-government regulations and separate regulations; using and developing the spoken and written languages of the ethnic groups; respecting and guaranteeing the freedom of religious belief of ethnic minorities; retaining or altering the folkways and customs of ethnic groups; independently arranging, managing and developing economic construction; independently developing educational, scientific, technological and cultural undertakings. (CHINA.ORG.CN, 2005, III, p. 1-3)

The same White Papers state that the government supports and assists the ethnic autonomous areas through the following programs:

> Giving prominence to speeding up the development of ethnic minority areas; giving priority to and rationally arranging infrastructure projects in ethnic autonomous areas; strengthening financial support for ethnic minority areas; attaching importance to ecological construction and environmental protection in ethnic autonomous areas; adopting special measures to help ethnic autonomous areas develop education; strengthening assistance to impoverished ethnic minority areas; increasing input into social services in ethnic autonomous areas; assisting ethnic autonomous areas to open wider to the outside world; pairing off more developed areas and ethnic autonomous areas for aid; giving care to special needs of ethnic minorities in production and living. (CHINA.ORG.CN, 2005, IV, p. 1-3)

These measures show that developing ethnic minority areas is a major governmental goal.[1] The literature, however, points out some issues and inequalities. For example ethnic minorities receive independent rights regarding the governmental leadership of the autonomous areas but not regarding the Chinese Communist Party (CCP) leadership of the autonomous area. The CCP leadership still belongs to Han and not to ethnic minorities (ZANG, 1998). MACKERRAS (2004) even stresses that the rich natural resources in autonomous areas are closely linked to economic interests of China and its developmental goals in these regions. For example XUAR has about 30 % and 34 % of China's continental oil and gas reserves, respectively (KUPFER, 2011b). Important Chinese rivers such as the Mekong River, Yellow River and Yangtze River have their starting points in TAR. With respect to inequalities in regional development, GUSTAFSSON and SAI (2008, 2006) find that poverty is mainly observed in Western regions and villages with low average income, while the average economic situation of minority villages in Northeast China is somewhat better compared to the average majority village. Given the preferential policies one would expect that ethnic minority distribution in the labor market would be

[1] There are also programs which focus on the development of all poorer areas and not only on autonomous areas, for example the China Western Development Program (西部大开发-*xibudakaifa*). There are also land policies which have the goal of fostering urban and rural development through efficient land use, land allocation and management strategies (DING, 2003).

positively affected. As shown in the existing literature on ethnic labor market differences, there are, however, various findings and several issues unresolved.

1.4 Resulting research questions

Having shown the diverse labor market situations and educational attainments of ethnic minorities as well as the preferential policies to tackle ethnic inequalities in the previous subchapters, I will now identify the main research questions of this study. Before doing so I want to clarify the major terms of my research and some methodological pitfalls. These are the somehow interrelated concepts of labor market discrimination, prejudice, occupational choice and job preference.

The standard neoclassical approach assumes that labor demand is characterized by the employer's willingness to hire a person, while labor supply is based on individual's utility gained from employment (MENG and MILLER, 1995). On the demand side are employers who demand labor usually with the goal of profit maximization. It is on the demand side of the labor market where job discrimination against ethnic minorities may occur. Labor market discrimination is "a situation in which persons who provide labor market services and who are equally productive in a physical or material sense are treated unequally in a way that is related to an observable characteristic such as race, ethnicity, or gender" (ALTONJI and BLANK, 1999, p. 3168).

By "unequal" the authors mean that "these persons receive different wages or face different demands for their services at a given wage" (ALTONJI and BLANK, 1999, p. 3168). Labor market discrimination is often closely linked to prejudice against a disadvantaged group. Prejudice is defined as "a negative attitude toward an entire category of people, often an ethnic or racial minority" (SCHAEFER, 2007, p. 265).

Looking at the other side of the labor market, the labor supply reflects individual occupational choices based on their utility. Many authors use the term "occupational choice". I believe caution should be exercised when using the term choice-based as the word "choice" includes, according to the ONLINE-CAMBRIDGE DICTIONARY (2011), an "act or the possibility of choosing". Most individuals actually take whatever job is available to them, whether or not they like it. For instance if the labor demand of ethnic minority workers in non-agricultural employment is restricted by discrimination, and if these ethnic minorities have a preference for non-agricultural employment, *ceteris paribus*, then these workers are unable to freely choose their preferred occupations. The term "occupational outcome" rather than "occupational choice" better reflects these circumstances. It indicates that not all individuals have the option to choose their preferred occupations based on their utility; thus, I assume job preferences are developed when individuals are seeking occupations that satisfy their interests and goals and for which they possess the skills, abilities and temperament (GOTTFREDSON, 1981, 1996).

Furthermore there is a huge difference between rural and urban employment possibilities in China. In rural areas subsistence agriculture is still common. Many peasants cannot choose a job from a wide range of occupations. They face, rather, the fundamental decision of continuing agricultural tasks, leaving the sector for available non-agricultural positions in their region or migrating to coastal areas.

Regarding ethnic minorities, I already pointed out in the previous subchapter that their role in the Chinese labor market is quite diverse. Results differ depending on the assigned ethnic minority status, human capital factors and geographic locations among other factors. While there are already some publications about ethnic minorities from the most problematic areas, TAR and XUAR, suggesting that ethnic minorities are discriminated against in the labor market, investigations about ethnic minorities in the poorest area of China, Guizhou province, are relatively scarce. The available studies are, furthermore, contradictory to received wisdom (cf., GUSTAFSSON and LI, 2003; MAURER-FAZIO et al., 2004, 2005). This makes Guizhou, with an ethnic minority population of around 37 %, a case particularly worth studying. If it is true that ethnic minorities in Guizhou are better off economically, then research findings in Guizhou may even direct development in those regions with considerable ethnic job discrimination. Guizhou has a large rural population; according to the WORLD BANK (2008) roughly 85 % of the population of this province lives in rural areas. Given such a demographic distribution, it seems logical to focus on rural areas. The focus of this study, therefore, is on investigating labor market discrimination against ethnic minorities in rural Guizhou.

The attentive reader may already have noticed that a major issue arises in distinguishing labor market discrimination and preferences (cf., SCHMIDT and STRAUSS, 1975). For example imagine a simple regression where the dependent variable is a discrete variable for job, which equals one for a "good job" and zero for a "bad job", and where the independent variable is a discrete variable for ethnic status, which equals one for "ethnic minority" and zero for "ethnic majority", *ceteris paribus*. A significantly negative effect of the ethnic minority status on a "good job" can either indicate that the ethnic minority group is discriminated against (the access to a good job is denied for ethnic minorities) or that the group has a higher preference for the "bad job", *ceteris paribus*. This particular setting indicates that the coefficient for ethnic status can imply completely opposite results. The question arises of how the ethnic effect can be disentangled. It is an open question whether or not ethnic minorities are discriminated against in "good jobs" or simply prefer to work in "bad jobs".

Qualitative inquiries may provide some deeper understanding of this issue. Labor market discrimination is, however, forbidden in China, and field investigations about sensitive topics are restricted. Interviewing employers on their hiring practices and/or disadvantaged groups on their experiences looking for work can lead to inaccurate and misleading results. Employers will not admit that they discriminate against a

disadvantaged group as it is a criminal act; while the disadvantaged group has only a restricted view about their situation as the employers will not openly inform them about their possibly discriminatory behavior.

Taking labor market discrimination as a potential rural development issue in China, this monograph is motivated by two central questions, which are related to content and methodological applications:

1) Are minorities, due to their ethnic affiliation, discriminated against in the rural labor market in Guizhou?
2) To what extent is discrimination empirically measurable?

The two central tasks are now to investigate *labor market discrimination* and *occupational outcomes* of ethnic minorities in relation to their underlying causes. First, this requires theoretical and empirical understanding of available concepts and their respective controversies and limitations. Second, the available empirical evidence from Guizhou will be investigated to formulate applicable research hypotheses, which will then be tested with the required set of methodologies.

1.5 Controversies in theories and methodologies and limitations to current wisdom

This subchapter aims at giving an overview of theoretical and methodological controersies and highlighting the limits of current wisdom. I give only brief introductions of major theoretical concepts and methodologies which are covered in more detail in later sections of this monograph. Finally I provide a brief overview of the labor market in Guizhou in this subchapter.

1.5.1 Theoretical controversies

Labor market discrimination theories can roughly be divided into two groups: taste-based discrimination and statistical discrimination theories. BECKER (1957, 1971) pioneers taste-based discrimination models which assume that a marginal discriminatory employer would pay for his/her special taste a wage premium to majority workers as a result of his/her prejudice against ethnic minority workers. Another group is statistical discrimination models pioneered by PHELPS (1972) and ARROW (1973). The major contribution of this branch of literature is to show that it is not a general distaste of majority employers against ethnic minority workers but employers' *beliefs* that ethnic minority workers are less productive than majority workers which facilitates discrimination.

Ethnic differences in occupational outcomes cannot be reduced to labor market discrimination against ethnic minorities. The occupational exclusion theory of JOHNSON and STAFFORD (1998) postulates that there are at least four reasons for occupational segregation of a disadvantaged group: 1) employer discrimination, 2) institutional discrimination, 3) different abilities and 4) different preferences

(ALTONJI and BLANK, 1999, p. 3176). While the theory covers the major differences which can cause occupational segregation, additional assumptions are required when it comes to empirical testing. The fundamental question of how to empirically disentangle different preferences and discrimination remains unanswered.

As the share of individuals working in agriculture is more than half of the rural population in Guizhou, theories which take into account specific characteristics of farm households have to be considered. BUCHENRIEDER et al. (2001) and MÖLLERS (2006) developed an integrated framework to analyze non-farm rural employment (NFRE). Which job a former agricultural worker receives in the non-farm sector is explained by demand-pull/distress-push concepts. The demand-pull process describes cases in which former agricultural workers receive better paid work in the rural non-farm economy, while the distress-push process describes cases in which former agricultural workers are pushed into poorly paid non-farm work, where wages can even be lower than in agriculture. So far labor market discrimination theories have not been included in the integrated framework, yet labor market discrimination constrains access to NFRE and/or is a crucial distress-push factor for ethnic minorities.

1.5.2 Methodological controversies

The fact that labor market discrimination is illegal in China and that research on sensitive topics regarding ethnic minorities is restricted requires adjusting methodological approaches. Given shortcomings of quantitative methodologies, PETRICK (2004) suggests that researchers should investigate economic behavior more accurately through means such as field work or experiments. In this subchapter I introduce some quantitative and qualitative approaches used for measuring labor market discrimination and differences in occupational outcomes and highlight the pros and cons of each approach.

1.5.2.1 Quantitative methodologies

One important tool for analyzing ethnic differences in occupational outcomes is discrete choice models which measure the causal relationships of a set of explanatory variables on a discrete dependent variable. To analyze different occupational outcomes requires that the discrete dependent variable stands for the considered set of occupations and that the explanatory variables include those factors which cause differences in these occupations. There are, therefore, at least three decisions which have to be made when applying discrete choice models: which particular model setting, which set of occupations and which explanatory variables to use. All three decisions depend on available secondary data, theoretical considerations and statistical testing results.

The decision about which of the many discrete choice models one should apply depends on whether or not the available data satisfy the assumptions made on distributions and variations of the error component of the underlying model's specifications

(e.g., independence from irrelevant alternatives assumption). In addition the specification of the dependent variable, usually an unordered set of available occupations, is based on available databases which often contain only the chosen occupations. Neither the decision-making process nor other occupations considered are observed. A major controversy lies, therefore, in the identification of the set of considered occupations; moreover, researchers are, for computational reasons, seldom able to use the complete set of all available occupations as a dependent variable. Within the empirical application suitable categories often have to be merged depending on their correlations or theoretical considerations.

When choosing explanatory variables, researchers face the challenge of selecting those variables which are of theoretical and statistical importance. When setting up the model, researchers can face problems such as overfitting (too many insignificant variables) or omitted variable biases (the absence of significant variables). Given the extensive theoretical literature on labor (cf., LIST and RASUL, 2011), researchers need to disentangle several interrelated explanatory variables in order to understand labor market discrimination and differences in occupational outcomes. This requires assumptions about linkages between explanatory variables, a major source of controversies among scholars. As already repeatedly highlighted, researchers, moreover, face the difficulty of distinguishing between discrimination and different preferences when a significant coefficient of ethnic status is observed.

Regressions can also face endogeneity problems, meaning that the direction of causal effects can be blurry; for example, more education leads to a better job or to higher wages, versus the wish to access a better job or to earn higher wages causes individuals to study more. It is impossible to pinpoint these tiny differences using the available secondary data. Researchers have to make assumptions about causal relationships, making the correct assumptions about unknown ethnic environments is a particularly difficult undertaking.

1.5.2.2 Qualitative methodologies

Researchers apply qualitative methodologies to understand a limited number of cases and their interrelationships in depth. Qualitative methodologies face at least four shortcomings in measurement: selection bias, inability to capture evolution over time, inability to generalize research findings and researcher bias.

Selection bias and consequently biased results based on truncated samples are major shortcomings of qualitative research; for example, for investigations in China, researchers are requested to work with local partners who generally preselect research areas and interviewees. Crucial issues regarding social inequalities may, therefore, remain untouched as not all areas and individuals are accessible, particularly in TAR and XUAR. Research projects, moreover, are often applied at one single point in time, meaning developments over time are not captured. These two issues (selection bias and inability to capture evolution over time) lead to shortcomings in the generalization of findings. Usually results depend on the

selected area at the particular time of the field work, which means that it is difficult to make conclusions for the whole country.

Another problem of qualitative methodologies is the intervention of the researcher in the inquiries; researchers can, unintentionally or not, influence results based on their own beliefs and ideologies. Participant observation, which is a qualitative tool used to acquire mainly supporting information of field studies, suffers from these shortcomings. Generally it can be said that receiving accurate information is very difficult when investigating forbidden topics such as labor market discrimination. I conclude that combining quantitative and qualitative methodologies is more informative and permits a deeper understanding of labor market discrimination and ethnic differences in occupational outcomes (cf., LIST and RASUL, 2011).

1.5.3 The state of knowledge in Guizhou province

Labor market discrimination and occupational difference by ethnicity are not directly reported in official data; therefore, in this subchapter I seek to show labor market trends based on official statistics for Guizhou. I intend to illustrate that the employment share in primary industry has been by far the highest in Guizhou since 1949, while in the last decade individuals' education has been continuously rising. This may indicate a fiercer competition in employment in secondary or tertiary industries where only few jobs are available for a more and more educated population, which may facilitate labor market discrimination against disadvantaged groups. I, further, introduce the ethnic groups of Guizhou and point out the importance of considering not only the broader categories of Han or Non-Han, but rather seek to point out the importance of analyzing labor market discrimination and differences in occupational outcomes for ethnic minorities separately. Based on the China Health and Nutrition Survey (CHNS) I present data regarding different occupational outcomes of the major ethnic groups of the province to give an additional motivation for my study.

1.5.3.1 General information and labor market development

Guizhou is a mountainous province in southwestern China. The province, which has a land area of 176,100 km², had an official population of 38.5 million people in 2010 (DEUTSCHE BANK, 2012). The western part of Guizhou belongs partly to the Tibetan high plateau and has varying elevations of 1,500-2,800 meters, the central plateau is around 1,000 meters, and the southeast is on average around 600-800 meters (ZHANG, 2003). The province has large water reserves; the upper reaches of Yangtze River and Pearl River as well as Huangguoshu Waterfalls are located in Guizhou. The province contains the main reserves of coal and phosphor within southwestern China and is among the leading providers of bauxite in China (PEOPLE'S DAILY, 2011). With 62 % karsts landforms and 19 % stony desert, Guizhou's land area is, however, very fragile and hinders economic development (CHINA DAILY, 2012). These topographical conditions have a clear impact on

employment options; in remote areas there are often fewer employment possibilities than in better developed areas nearer to urban centers.

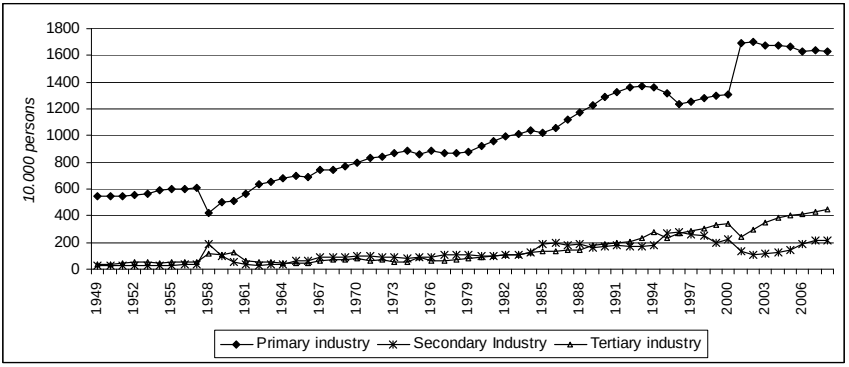
Wei Houkai, deputy director of the Institute for Urban and Environmental Studies at the Chinese Academy of Social Sciences, points out that during the period of the 11th Five-Year Plan (2006-10), the average economic growth of Guizhou is below the national average and that the current income per capita stands at a low level of around US-$2,000 annually (cited in CHINA DAILY, 2012). For comparison the average gross national income of China is US$4,930 in 2011 (WORLD BANK, 2012). The state council announced plans to develop a special economic zone in the Guiyang-Anshun area to boost economic growth of the province (CHINA DAILY, 2012). The increasing infrastructure development and the establishment of more and more industrial enterprises have boosted the development potential of Guizhou (PEOPLE'S DAILY, 2011). Despite this development today primary industry provides by far most of the employment in the province (figure 1-1); in 2008 around 16.3 million people worked in primary industry. Since China joined the WTO in 2001, employment in secondary and tertiary industries has slightly increased, yet with around 2.2 and 4.5 million employees in 2008, respectively, is comparatively low (figure 1-1).

Additional education is one major criterion for leaving employment in primary industry for employment in secondary or tertiary industries. Except during the Great Famine (1958-1961) and during the Cultural Revolution (1966-1976), Guizhou has always shown steady improvement in student enrollment particularly in primary and in regular secondary schooling (figure 1-2). Enrollment in vocational secondary schooling and in universities has greatly increased during the last decade. While in 1997 around 106,700 students graduated from university, this figure increased to around 705,300 students in 2008 (figure 1-2). The existence of relatively few employment possibilities in secondary or tertiary industries and the growth of the educated population create fierce job competition in these industries and may in turn result in labor market discrimination against ethnic minorities in Guizhou.

1.5.3.2 Ethnic groups and occupational outcomes in Guizhou

Guizhou counts 18 officially registered ethnic groups, 17 ethnic minorities and the Han. The ethnic minorities make up more than 37 % of the population (PEOPLE'S DAILY, 2011). More than half of the area in the province is dedicated to autonomous areas. There are three autonomous prefectures and eleven autonomous counties with ethnic minority proportions ranging from around 25 % to 96 % (table A2). There are, hence, many areas where mixed populations are living together but also many villages in which almost all residents belong to the same ethnic group.

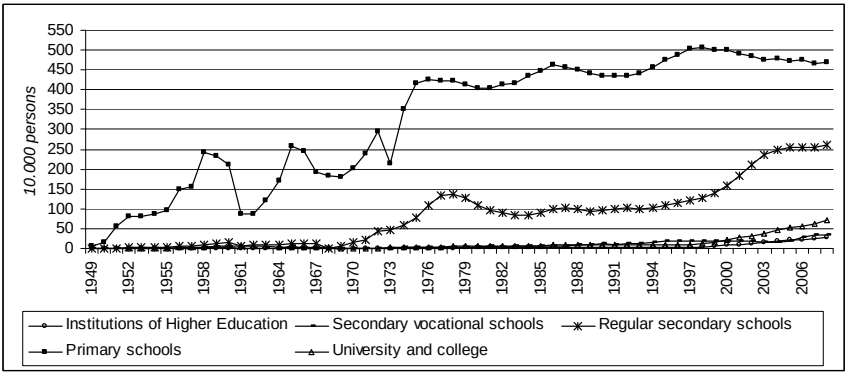
Figure 1-1: Development of employment by industry in Guizhou (1949-2008)



Primary industry is an industry that produces energy or basic materials, such as coal, oil, metals, crops, etc.; secondary industry is an industry that manufactures goods rather than producing raw materials; tertiary industry is an industry that provides services rather than producing goods (ONLINE-CAMBRIDGE DICTIONARY, 2012).

Source: Author based on NATIONAL BUREAU OF STATISTICS OF CHINA (2009).

Figure 1-2: Development of student enrollment in Guizhou (1949-2008)



Source: Author based on NATIONAL BUREAU OF STATISTICS OF CHINA (2009).

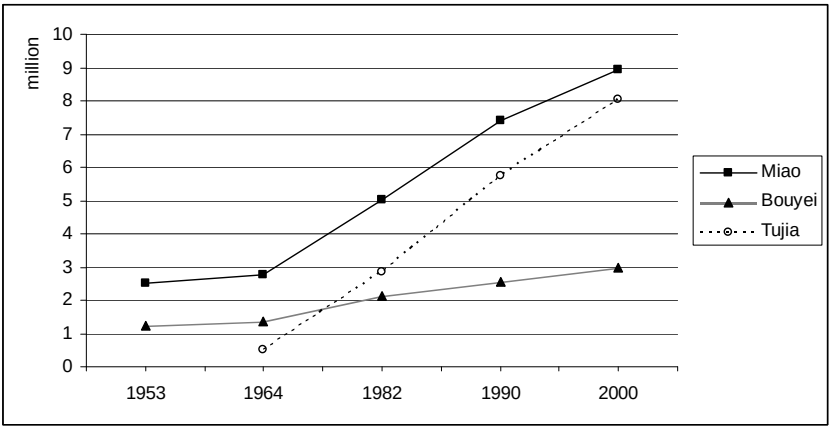
Contrary to received wisdom, ethnic minorities in Guizhou are in a somewhat better economic situation than the average Han due to increasing tourism and border trade (GUSTAFSSON and LI, 2003), but it remains to be answered whether or not this situation is the same for all ethnic minorities and whether or not there have been changes over time. I intend to find out whether there are differences in occupational outcomes between ethnic minorities and to investigate the major ethnic groups of Guizhou, the Bouyei, Miao and Tujia, separately in comparison to Han. The major reason for concentrating on these three ethnic minorities is

the secondary data about each of these groups available in the CHNS. I can, therefore, combine econometric analysis based on these data and conduct supporting field work to increase the accuracy of measuring economic behavior.

I will briefly introduce these three ethnic minorities. They are all mainly distributed in southern China. Census data from 1953 to 2000 show that the total population share of each group steadily increased (figure 1-3). This is related to demographic changes and to reclassifications. There were 8.9 million Miao distributed in the provinces Guizhou, Hunan, Yunnan, Guangxi, Chongqing, Hubei and Sichuan in 2000. The number of Tujia, who were first mentioned in the 1964 census, has sharply increased over the years. There were eight million Tujia distributed in Hunan, Hubei, Chongqing and Guizhou in 2000. The stark increases for both Tujia and Miao populations are strongly related to reclassifications rather than to demographic changes (e.g., SCHEIN, 2000, p. 69, ZHOU, 2003, p. 14; BROWN, 2001, p. 56). The Bouyei, with a naturally increasing share to almost three million people in 2000, have their major concentration in Guizhou.

The economics literature on these three groups is very scarce, and ethnographic studies are limited. The working papers of MAURER-FAZIO et al. (2004, 2005) are the only publications I am aware of which directly analyze the Bouyei, Miao and Tujia separately in labor market research. The authors compare the labor market participation rates of these three groups along with other ethnic minorities and Han in the available population censuses.

Figure 1-3: Total populations of the Bouyei, Miao and Tujia between 1953 and 2000



Source: Author based on table A1. Ethnic Statistical Yearbook 2007, population figures are from the 2000 population census; population figures of the years 1990, 1982, 1964, 1953 are taken from ZHOU (2003, p. 12-13).

They find that in 1982 the Miao had a 7 % higher labor force participation probability than Han, while in 2000 there was no statistically significant difference (MAURER-FAZIO et al., 2004, p. 12). The authors observed that in 2000 the Bouyei had a 14 % higher probability of being in the labor force than Han (MAURER-FAZIO et al., 2004, p. 17), while their results for the Tujia were not statistically significant. This is in line with the more general inquiry of GUSTAFSSON and LI (2003), which finds a larger increase in per capita income among ethnic minorities than among Han between 1988 and 1995 in Guizhou and Yunnan. The results suggest that Han are actually underrepresented in the labor market in Guizhou.

Few ethnographic studies of the three minorities are available. Apart from linguistic inquiries of Bouyei languages conducted by researchers of SIL International, I am not aware of any available publications regarding the Bouyei. To the best of my knowledge, the only recent available publications regarding the Miao are SCHEIN (2000) and DIAMOND (1995). In her ethnography SCHEIN (2000) studies several aspects of Miao living mainly in the Xijiang community in Guizhou. I cite several of her important findings when setting up the conceptual framework of this study in the next chapter. DIAMOND (1995) analyzes the ethnic classification of Miao. She concludes that there is diversity among Miao in all four aspects of STALIN's definition: language, territory, economic life and culture.[2] BROWN (2001) who analyzes the ethnic classification of Tujia also finds that the official classification of Tujia in Enshi prefecture in Hubei differs from their actual identity and culture. This also indicates that a deeper analysis of ethnic minorities demands caution as their applied definitions are not always accurate.

MACKERRAS gives very brief general descriptions of ethnic minorities in China. Regarding the Bouyei, Miao and Tujia he writes, respectively:

> The Bouyei way of life is similar to the Miao and their language is closely related to those of the Zhuang and Dai. They practise polytheism and ancestor worship.
>
> The Miao are one of the most ancient of China's nationalities, tracing their origins back more than 4,000 years. In China some people whom the state classifies as Miao regard themselves as Hmong. Groups of this designation are found in many of the countries of mainland Southeast Asia. Before the modernisation of farming methods the Miao grew millet and buckwheat using the slash-and-burn method.
>
> The Tujia farm rice and corn collect fruit and fell trees for lumber. In most ways they are very similar to the Han people. (MACKERRAS, 2003, p. 183-191)

The majority of the people in Guizhou work in primary industry (figure 1-1). Among the total labor force of around 22.9 million people in 2008, the employed persons in

---

[2] The communist government classified ethnic minorities or similarly ethnic nationalities based on a theory developed by STALIN (1913, 1953). Theoretically those groups with a common language, territory, economic life, psychological make-up and culture received the same ethnic status.

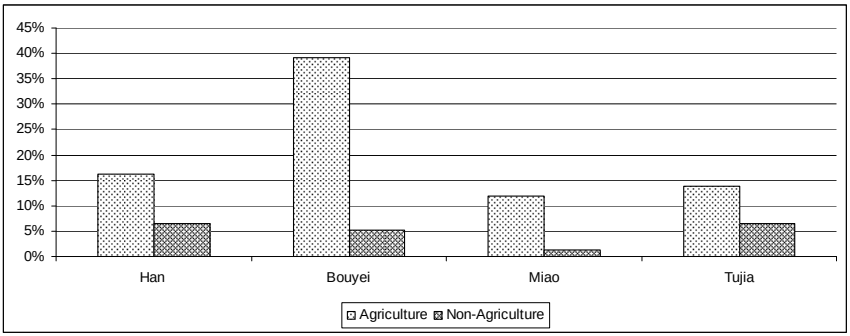
rural areas account for 79 % and in urban areas for 21 % (NATIONAL BUREAU OF STATISTICS OF CHINA, 2009, table 25-2). When looking at differences between agricultural and non-agricultural employment in rural areas, the huge share of agricultural employment is striking. Not only the Bouyei, Miao and Tujia but also Han are mainly working in agriculture (figure 1-4). The greatest share of agricultural employment is observable for the Bouyei. The shares of non-agricultural employment for the three groups are comparatively low, with the lowest share for the Miao.

To sum up it remains to be answered whether or not the different occupational outcomes of the three groups in comparison to Han are statistically significant considering multivariable settings. As I already pointed out, the investigation of labor market discrimination and occupational differences by ethnicity is a complicated undertaking which requires multidisciplinary theories and mixed methodologies. It can be said that the empirical knowledge of labor market discrimination and ethnic differences in occupational outcomes in Guizhou is still insufficient and unclear. These issues still await a precise theoretical and empirical investigation.

1.6 Organization of the monograph

The major goal of this study is to link the different theoretical concepts with a rigorous empirical examination of the Guizhou case. To give the reader a better orientation, the overall structure of this monograph is presented in table 1.1. Following the two leading research questions as stated in subchapter 1.4, in the second chapter I first discuss major theories and empirical methodologies to analyze labor market discrimination and ethnic differences in occupational outcomes and wages (subchapters 2.1 to 2.4). Based on the most relevant theoretical outcomes and empirical results I derive the major determinants of ethnic differences in occupational outcomes and wages of which labor market discrimination is only one of many.

Figure 1-4: Frequency of agricultural versus non-agricultural employment by ethnicity in 2004



Source: Author based on CHNS sample 2004.

The conceptual framework for the empirical analysis and the leading hypotheses are then developed in subchapter 2.5. The reader particularly interested in theoretical and empirical discussions, their linkage and a broader theoretical reflection on China can focus on the second chapter of this monograph. For a quick overview, I suggest having a look at figures 2.4 and 2.5, which show my study's integrated theoretical framework and conceptual framework, respectively.

The third chapter describes the qualitative approach and presents field work results from Guizhou. The reader particularly interested in qualitative methodologies can focus on this case study in chapter three. I also include several photos here. For a more general discussion of qualitative methodologies the interested reader may also have a look at subchapter 2.4.2, where I discuss qualitative approaches to measure labor market discrimination. The qualitative portion of my work plays a complementary, "corrective" role to the quantitative portion.

The fourth chapter focuses on the quantitative part of the Guizhou case. In this chapter I first describe the database and provide descriptive statistics. I then explain the econometric modeling approach and provide estimation results. The reader with particular interest in the quantitative part of the Guizhou case can focus on this chapter and may also follow a more general discussion on quantitative methodologies in subchapter 2.4.1.

The last chapter combines the theoretical and empirical conclusions of the study, gives policy implications and suggests areas for future investigation. Throughout the monograph I provide short summaries of the major contents at the beginning of each chapter and interim conclusions when required.

Table 1.1: Structure of the monograph and presentation in the text

Are minorities, due to their ethnic affiliation, discriminated against in the rural labor market?	*To what extent is discrimination empirically measurable?*
Theoretical discussion (2.1 to 2.3)	Empirical discussion (2.4)
Major determinants of ethnic differences in occupational outcomes and wages: theoretical perspectives (2.3.1) and empirical perspectives (2.4.3)	
Development of testable hypotheses and empirical research methodology (2.5)	
Presentation of qualitative approach and results (3) General discussion of qualitative methodologies (2.4.2)	
Presentation of quantitative approach and results (4) General discussion of quantitative methodologies (2.4.1)	
Conclusions (5)	

Source: Author.

2 LABOR MARKET DISCRIMINATION AND ETHNIC DIFFERENCES IN OCCUPATIONAL OUTCOMES IN THEORETICAL AND EMPIRICAL PERSPECTIVES

Taking labor market discrimination as a potential rural development issue in China, this monograph is motivated by two central questions. The first question, are minorities, due to their ethnic affiliation, discriminated against in the rural labor market in Guizhou, requires at least a theoretical analysis of *ethnic minorities*, *labor market discrimination* and *occupational outcomes* with particular consideration of the *rural* labor market. The interdisciplinary character of the question, furthermore, requires integrating theoretical knowledge and hypotheses into the research design. To solve the second question, to what extent is discrimination empirically measurable, requires determining whether or not empirical methodologies actually answer the content-based question.

In subchapter 2.1 I first briefly review the major paradigms regarding theory construction before explaining the research design of this study. In subchapter 2.2 I discuss theoretical perspectives. In sections 2.2.1 to 2.2.4 I introduce theories of group differences, labor market discrimination theories, occupational "choice" theories, farm household theories and non-farm rural employment theories. In section 2.2.5 I provide theoretical conclusions. In subchapter 2.3 I link relevant theoretical concepts into one single theoretical framework and discuss the literature on labor market discrimination of ethnic minorities in China. Subchapter 2.4 focuses on the extent to which discrimination is empirically measurable and provides a discussion of suitable empirical approaches. In section 2.4.1 and 2.4.2 I review quantitative and qualitative methodologies, respectively. Section 2.4.3 gives interim conclusions. In subchapter 2.5 I link the theoretical and empirical considerations. In section 2.5.1 I derive the leading hypotheses of my study. In section 2.5.2 I explain how I combine quantitative and qualitative methodologies to test the hypotheses. In subchapter 2.6 I draw conclusions.

2.1 The role and position of theory

Researchers who enquire about ethnic minorities usually follow a qualitative approach. They build their theoretical framework and derive hypotheses based on direct field observations *throughout the research process* or do not consider theoretical underpinnings at all; they stick, rather, to descriptions of the research field. Researchers who investigate labor markets usually follow a quantitative approach which is based on the deductive nomological (D-N) paradigm. Quantitative

researchers, therefore, must use existing theories to define their theoretical framework and hypotheses. The German term "Werturteilsstreit" refers to the debate over the degree to which researchers' own values, including personal and political opinions or ideological goals, influence their scientific work and whether or not these *normative beliefs* are valid to explain theories (SCHNELL et al., 1998, p. 83). The basic logical problem of the qualitative paradigm is that researchers implicitly have theoretical knowledge before they go to the field and are, therefore, not starting from a completely "clean state" (KELLE, 1997, p. 23). Advocates of the D-N paradigm postulate that interpretative and theory-building methodologies cannot generally provide valid and reliable results, given that these methodologies have a strong normative nature (KELLE, 1997, p. 11). The D-N paradigm, which strictly evaluates (verifies or falsifies) *existing* theories and hypotheses, is the mainstream paradigm used in labor market inquiries.

As I was not raised in rural Guizhou and lack necessary background knowledge, I cannot accurately analyze my content-based question solely with the D-N paradigm because the theoretical framework and hypotheses have to be defined *before the empirical phase*. The literature on ethnic minorities and their role in Guizhou's rural labor market is, moreover, very scarce. It is, thus, inappropriate to formulate definite hypotheses based solely on the scarce literature without supporting knowledge of the research field. Given that the strict quantitative and qualitative divide is deceptive (THEESFELD, 2005), and given that the nature of my research problem requires an integrated approach, I combine both quantitative and qualitative strategies to analyze labor market discrimination and ethnic differences in occupational outcomes.

2.2 Theoretical perspectives

To understand labor market discrimination and ethnic differences in occupational outcomes requires theories which explain why labor market discrimination and occupational differences exist. Different occupational outcomes among ethnic minorities do not necessarily imply that ethnic minorities are discriminated against in the labor market. It is a very complex issue of many interlinked factors which are very difficult to capture empirically. To understand causes of labor market inequalities, one must not only understand the labor market itself but also understand the evolution of heterogeneous worker characteristics. In section 2.2.1 I first introduce theories which explain group differences, then provide an overview of labor market discrimination theories in section 2.2.2 and occupational "choice" theories in section 2.2.3. Given my research focus on rural areas, I also consider farm household models in section 2.2.4 and non-farm rural employment theories in section 2.2.5. Finally in section 2.2.6 I draw theoretical conclusions.

2.2.1 Theories of group differences

In competitive labor markets differences in job preferences and human capital affect labor market participation rates and the distribution of occupations and wages. The competitive theory of group differences underlies the hypothesis that "group differences in wages, occupations, and employment patterns are the consequence of preference and skill differences rather than discrimination" (ALTONJI and BLANK, 1999, p. 3164).

Job preferences of young people evolve depending on their interests, goals, skills, abilities and temperament when they are growing up (GOTTFREDSON, 1981). Evolution of different job preferences is, moreover, closely related to differences in child-rearing practices, educational systems, comparative advantages and human capital (ALTONJI and BLANK, 1999). Parents strongly influence the job preferences of their children; for example, ALTONJI and BLANK (1999) state that parents who are convinced that their daughter will face discrimination in the job market attempt to form her preferences for more traditional functions to prevent her from future discrimination in the labor market. This may also hold true for racial or ethnic minorities who are discriminated against in the labor market.

Skill differences are linked to differences in comparative advantages, human capital accumulations and preferences. Theories in economics of the family suggest that comparative advantages evolve out of biological differences (e.g., child bearing and physical strength) between women and men (ALTONJI and BLANK, 1999). While thirty years ago women had a comparative advantage in home production and men in the labor market, particularly for those tasks which require physical strength, in the US the increasing importance of interpersonal and cognitive skills resulted in a more and more similar occupational structure among genders (ALTONJI and BLANK, 1999). Closely linked to different comparative advantages are differences in human capital accumulations before entry into the labor market. Evidence from the US suggests that skill differences between ethnic and racial groups are strongly influenced by family backgrounds, neighborhoods and school quality. Many researchers find that in the US African Americans and Hispanics accumulate lower human capital before entering the labor market than do Whites because ethnic minorities are often poorer, grow up in impoverished neighborhoods and receive lower quality schooling (ALTONJI and BLANK, 1999). These factors have negative implications for labor force participation rates as well as for occupational and wage distributions for the disadvantaged groups. LIST and RASUL (2011, p. 140), based on GLEWWE and KREMER (2006), postulate that educational production is a function of child characteristics (including "innate ability"), household characteristics, school and teacher characteristics (quality) and costs related to schooling, where school and teacher characteristics (quality) and costs related to schooling are both linked to education policies and local community characteristics. Opportunity costs of schooling and expected returns on schooling are, moreover, closely linked to the aforementioned factors (ALTONJI and BLANK, 1999). While it is theoretically

possible to disentangle factors which cause pre-labor market differences among ethnic groups, it is empirically very complicated to separate group differences as a result of pre-labor market differences from labor market discrimination given that each concept has many underlying often interlinked causes.

2.2.2 Labor market discrimination theories

The demand side of the labor market is characterized by employers who demand labor with the goal of profit maximization. The labor supply reflects individual occupational outcomes based on workers utility gained from employment. It is on the demand side, in other words on the side of the employer, where job discrimination against ethnic minorities may occur. The microeconomics literature of labor market discrimination tries to find an explanation for this phenomenon and distinguishes broadly between taste-based discrimination theories, theories of statistical discrimination and theories of occupational exclusion (ALTONJI and BLANK, 1999). The literature of each of these particular areas is vast, and there is no room here to do it justice. In the following I, thus, discuss only the most important theoretical contributions of each of these three theoretical approaches.

2.2.2.1 Taste-based discrimination, Becker's model of employer discrimination

Taste-based discrimination models were pioneered by BECKER (1957, 1971). BECKER analyzes discriminatory behaviors of employers, employees and consumers in their choices for employees, employers and sellers, respectively. He assumes that a marginal discriminatory actor is willing to pay for his/her special "taste" based on his/her level of prejudice against the disadvantaged group. BECKER characterizes prejudice as "a distaste, or aversion to cross-racial contact." Prejudice is defined as "a negative attitude toward an entire category of people, often an ethnic or racial minority" (SCHAEFER, 2007, p. 265).

When explaining employer discrimination, BECKER assumes a competitive market with constant returns to scale in production. He assumes that black and white workers are perfect substitutes in production and that employers are all whites with differing levels of prejudice against black workers. Following CHARLES and GURYAN (2008), in BECKER's model an employer's utility (V_i) is based on his/her profit (π_i) and the number of black employees (L_b) in the enterprise. Every black employee is assumed to bring disutility of $d_i \geq 0$. Employer utility is, thus,

$$V_i = \pi_i - d_i L_b, \tag{2-1}$$

where $\pi_i = f(L_a + L_b) - w_a L_a - w_b L_b$ is employer's profit, $f(\cdot)$ is the production function with constant returns to scale and w_a and w_b stand for white and black wages, respectively (CHARLES and GURYAN, 2008, p. 777). Employers allocate white and black labor (L_a and L_b) with the goal of maximizing their utility, equation (2-1).

The labor choices under utility maximization (L_a^* and L_b^*) satisfy the following conditions,

$f'(L_b^* + L_a^*) - w_a \leq 0$, with equality if $L_a^* > 0$,

$f'(L_b^* + L_a^*) - w_b - d_i \leq 0$, with equality if $L_b^* > 0$ (2-2)

(CHARLES and GURYAN, 2008, p. 778). Based on these assumptions, condition (2-2) indicates that an employer hires labor up to the point when the marginal product equals marginal cost. The marginal cost for whites is the wage for whites, w_a. The marginal cost for blacks is the wage for blacks, and the prejudice of the employer, $w_b + d_i$. At this point BECKER shows that prejudice makes employers act as if wages for blacks are higher than they really are (CHARLES and GURYAN, 2008, p. 778). Given that white and black workers are assumed to be perfect substitutes, employers hire white workers as long as $w_a < w_b + d_i$ and black workers as long as $w_b < w_a - d_i$. In this way the market divides between black and white workers; the least prejudiced employers hire black workers, the most prejudiced employers hire white workers, and the marginal employer in equilibrium is indifferent. The prejudice of the employer in equilibrium, d_i^*, the "marginal discriminator" is the same as the marginal wage gap,

$$w_a^* = w_b^* + d_i^* \tag{2-3}$$

(CHARLES and GURYAN, 2008, p. 778).

Four key results of BECKER's basic employer prejudice discrimination model are:

> (a) that the marginal employer matters more than the average prejudice for relative wage differences; (b) that the number (or fraction) of blacks in the workforce is negatively related to racial wage gaps, with prejudice held constant; (c) that prejudice in the right tail of the employer prejudice distribution should not matter for racial differences whereas higher prejudice in the left tail of the prejudice distribution should affect racial wage gaps; and (d) that the mechanism that generates these patterns is the tendency of the market to segregate blacks from the most prejudiced whites. (CHARLES and GURYAN, 2008, p. 780-781)

ARROW (1971, 1973) criticizes the BECKER model for the fact that in the long run prejudiced employers are unable to survive in competitive markets because they have to pay more to employ white workers. Prejudiced employers have to pay $w_a = w_b + d_i$ while unprejudiced employers only have to pay w_b. Prejudiced employers, thus, make lower profits and will be driven out of the market in the long run. The model does not adequately explain the relationship of persistent wage differences and prejudice.

2.2.2.2 Taste based discrimination, Black's equilibrium search model

Equilibrium labor market search models are another branch of competitive taste-based discrimination models. BLACK (1995), who examines employer-employee interactions in an equilibrium search model of employer discrimination, assumes

that one part $(1-\gamma)$ of the total work force are minority workers (type-*B*) who are discriminated against and that the other part (γ) are majority workers (type-*A*) who are not discriminated against. BLACK assumes that all workers have the same productivity and preferences and that job search costs (*c*) are the same for both types *A* and *B*. There are two kinds of employers: those employers who only follow the goal of profit maximization without prejudice (*u*), and those employers who are prejudiced against type-*B* workers and only hire type-*A* workers (*p*). The share of *p*-employers makes up θ of all companies who only hire type-*A* workers for wage w_{pa}. The share of *u*-employers, who make up $1-\theta$, seek to maximize profits and pay wage w_{ua} to type-*A* workers and wage w_{ub} to type-*B* workers.

During a particular period a given job provides a worker utility, which equals his/her total wage plus a "match-specific job satisfaction component" (α) (ALTONJI and BLANK, 1999, p. 3172). While the employer is only aware of the distribution of α, the worker knows α before deciding on a job offer. A prejudiced employer *p* offers wage w_{pa} with probability θ to a type-*A* worker and an unprejudiced employer *u* offers w_{ua} with probability $1-\theta$. The type-*A* worker has a reservation utility to take a job (u^a) given the arrival probability of the two job offers,

$$u^a = f^a(\overset{-}{c}, \overset{?}{\theta}, \overset{+}{w_{pa}}, \overset{+}{w_{ua}}, \beta_\alpha), \tag{2-4}$$

where f^a is the probability density function and β_α "is the parameter vector of the distribution of α" (ALTONJI and BLANK, 1999, p. 3173). If search costs increase, then reservation utility will decrease, *ceteris paribus*, yet an increase in wages will increase reservation utility. A type-*A* worker takes a job offer when $w_{ja} + \alpha > u^a$, $(j = u, p)$. In contrast a type-*B* worker only receives a job offer from unprejudiced employers *u* with probability $1-\theta$. The reservation utility of a type-*B* worker is,

$$u^b = f^b(\overset{-}{c}, \overset{-}{\theta}, \overset{+}{w_{ub}}, \beta_\alpha). \tag{2-5}$$

(ALTONJI and BLANK, 1999, p. 3173).

As in the type-*A* case, higher search costs lead to a decreasing reservation utility, *ceteris paribus*. The higher the probability of prejudiced employers θ in the market, the lower the reservation utility of type-*B* workers because type-*B* workers will not receive job offers from prejudiced employers. If a type-*B* worker happens to get a job offer from an unprejudiced employer *u* and if the utility of this offer is larger than the reservation value, $w_{ub} + \alpha > u^b$, the type-*B* worker will take the job offer. A logical consequence of the BLACK model is that if $w_{pa} \geq w_{ua} \geq w_{ub}$ then $u^b < u^a$ because type-*B* workers only receive job offers from $1-\theta$ of all employers (ALTONJI and BLANK, 1999, p. 3173).

In the BLACK model employers also decide on wage levels. Given their awareness of higher job search costs for type-*B* workers, employers have some "monopsonistic power" over type-*B* workers, meaning, they are the only ones who control the market (BLACK, 1995). While prejudiced employers only make wage offers to type-*A* workers, unprejudiced employers offer lower wages to type-*B* than to type-*A* workers. This is because unprejudiced employers know about the comparatively smaller fraction of unprejudiced employers in the market and are, thus, aware of the lower reservation utility level of type-*B* workers. This means that even if there are only unprejudiced employers in the market, type-*B* workers will face wage discrimination.

This is in contrast to BECKER'S employer discrimination model, where the market segregates between prejudiced employers who only employ type-*A* workers and unprejudiced employers who mainly employ type-*B* workers (ALTONJI and BLANK, 1999, p. 3174). As in the BECKER model of employer discrimination, in the Black model prejudiced employers also pay higher wages to type-*A* workers and, thus, make lower profits than unprejudiced employers. Prejudiced employers will be driven out of the market in the long run, *ceteris paribus*. Wage discrimination against type-*B* workers, however, persists among unprejudiced employers in the long run.

An increase in the share of type-*B* workers, *ceteris paribus*, increases the wage gap and segregates the labor market in the BECKER model. What happens in the BLACK model is difficult to capture. An increase in the share of type-*B* workers $(1-\gamma)$ results in a lower share of type-*A* workers γ in the market. A lower share of type-*A* workers indicates that matching probabilities of prejudiced employers and type-*A* workers decline. Prejudiced employers may face a shortage of type-*A* workers. One alternative is that prejudiced employers keep hiring only type-*A* workers and, therefore, reduce production capacities and lose profits, *ceteris paribus*. Another alternative is that prejudiced employers hire type-*B* workers despite their prejudice. The reservation utility of type-B workers will increase, as might their wages. As pointed out by ALTONJI and BLANK (1999, p. 3174), a higher share of type-*B* workers in the market, therefore, increases type-*B* workers wages.

2.2.2.3 Statistical discrimination

The pioneering scholars in the field of statistical discrimination are ARROW (1971, 1973) and PHELPS (1972). Statistical discrimination models analyze employers' hiring and wage decisions under the assumption that employers face imperfect information about workers performance. Given limited information, employers use those characteristics of workers which are easily observed, for example, race or gender, to make predictions about job performance for the entire group. If an individual of one group happens to perform comparatively worse than an individual of another group, and if employers' use these facts to make the same predictions about job performance for all individuals of the disadvantaged group, then

employers statistically discriminate against the disadvantaged group. What causes discrimination is not, as suggested by BECKER, general distaste for type-*B* workers; rather it is the employer belief (ARROW, 1971, p. 25) that type-*B* workers are less productive than type-*A* workers.

Statistical discrimination studies are broadly classified into two groups: those analyzing "prior beliefs" about productivity of different groups and the resulting employment and wage decisions of employers and those analyzing effects of group differences in the precision of information, meaning, exact and accurate in form, time and detail about individual productivity (ALTONJI and BLANK, 1999, p. 3181). LANG (1986), moreover, developed a language discrimination theory indicating that only individuals who speak the same language can actually work together.

Of the statistical discrimination models I first introduce the model developed by COATE and LOURY (1993), which falls in the range of statistical discrimination models with group differences in prior beliefs. Second, I present the statistical discrimination model with group differences in information quality developed by LUNDBERG and STARTZ (1983). Third, I introduce LANG's (1986) theory of language discrimination.

Prior beliefs and signals

COATE and LOURY (1993) make an important theoretical contribution to the prior-belief studies. The authors closely follow the work of ARROW (1973). They derive a job-assignment model in which employers are only aware of group identities, but not of workers productivity before the job assignment. The authors basically conclude that workers with the same initial abilities acquire different skills based on employers' prior beliefs about workers productivity.

The model assumes two groups of workers ($g = A, B$); *A* stands for majority workers, and *B* for minority workers. Every employer offers two jobs, task *0* and task *1*. Every worker is able to do task *0*, as no particular skills are required. Task *1* requires specific skills. Only skilled workers are, thus, able to do the job. Workers who do task *1* receive a wage premium (*w*). The company receives a net-return $x_q > 0$, if a qualified worker is appointed to task *1*. The company makes a loss $-x_u < 0$, if the worker is unqualified for the job.

In the job matching process an employer first observes if a worker belongs to group *A* or *B*. The employer has a prior-belief (π_g) about each group's performance. In the hiring process the employer also receives a vague signal ($\theta \in [0,1]$) about a worker's qualification. The signal can be based on the results of an interview, a test or other job-related inquiries. The signal is higher if the worker is qualified and lower if not. Based on the signal θ and on the prior belief π_g, an employer derives the probability of a worker's qualification for task *1*. The employer

sets threshold values

$$s_g = s^*(\pi_g) \tag{2-6}$$

for each group (ALTONJI and BLANK, 1999, p. 3182). The threshold value decreases if the belief π about the group increases.

Workers initially have the same skills. To qualify for task *1*, workers need to make an investment in training for cost *c*. A worker makes this investment only if his/her expected benefits are larger than the training costs. The expected benefits of the investment are an increased probability of being assigned to task *1* and of receiving the wage premium *w*. A worker evaluates his/her probability of being assigned to task *1* based on the threshold values, with probability of $1 - F_q(s)$ if qualified and $1 - F_u(s)$ if unqualified. The model assumes that $\beta(s) \equiv w[F_u(s) - F_q(s)]$ is "the expected benefit of investment for any worker facing the standard *s*" (COATE and LOURY, 1993, p. 1225). A worker will invest if $c \leq \beta(s)$. The share of workers qualified for task *1* is $G(\beta(s))$.

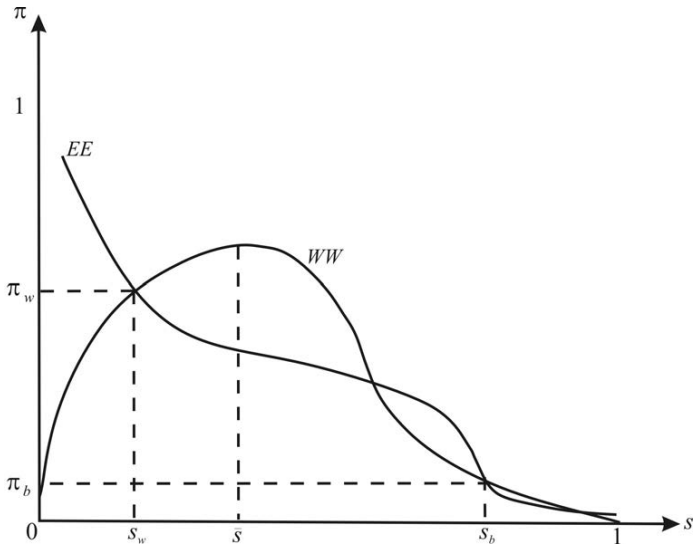
An equilibrium is observed if prior beliefs (π_b, π_w) satisfy the condition

$$\pi_g = G(\beta(s^*(\pi_g))) \qquad g = b, w. \tag{2-7}$$

(COATE and LOURY, 1993, p. 1225). There is a discriminatory equilibrium $\pi_b < \pi_w$ if (2-7) has multiple solutions. A discriminatory outcome may occur if an employer is uncertain that workers of group *B* will qualify for task *1*, which leads to a lower probability $\pi_b < \pi_w$. The expected benefits from investment in job training decline for type-*B* workers, and fewer type-*B* workers will invest in training, which confirms the initial negative belief of employers about type-*B* workers. This proves the major conclusion of COATE and LOURY (1993) that employers' initially negative beliefs about a group's performance are self-confirming.

COATE and LOURY (1993) also illustrate their analysis graphically (see figure 2-1). The vertical axis measures the belief π. The horizontal axis measures the threshold values (standards) *s*. The downward sloping curve *EE* shows the function $\{(s,\pi) \mid s^* = (\pi)\}$, which represents the pairs of standards-beliefs under optimal employer behavior; *WW* shows the function $\{(s,\pi) \mid \pi = G(\beta(s))\}$, which represents the pairs of standards and the share of workers investing in training (COATE and LOURY, 1993, p. 1226). The intersections of *EE* and *WW* illustrate all possible equilibriums. Multiple interactions imply that there are discriminatory equilibriums (figure 2-1). Assuming the extreme case, that employers believe a group is unable to perform task *1*, then, $(s,\pi) = (1,0)$. This implies that $G(0) = 0$ and shows that the belief that a group is unable to perform task *1* is self-confirming; workers from the disadvantaged group are not willing to invest in additional training because employers assign them to task *0* anyway (figure 2-1).

Figure 2-1: An equilibrium with negative stereotypes against B's



Source: COATE and LOURY (1993, p. 1225).

The main conclusion of COATE and LOURY (1993) is that only when employers have the same beliefs about the productivity of both the favored and disadvantaged groups can equal labor market outcomes among the groups be realized. The authors conclude that this is also the goal of affirmative action policies. They argue, however, that affirmative action policies can also raise labor market inequalities as a result of increasing stereotypes against the disadvantaged group. They base their last argument on the fact that employers are, due to affirmative action policies, forced to hire a certain amount of workers from the disadvantaged group.

The authors assume that employers, then, set lower hiring standards for workers from the disadvantaged group. The disadvantaged group is, thus, not required to invest in additional skills because of the job privilege. The stereotypes about lower skill levels and productivity against the disadvantaged group will remain in the labor market and result in a "patronizing equilibrium" (COATE and LOURY, 1993, p. 1230).

Information quality about productivity

The second group of statistical discrimination models considers the effects of group differences in the "precision" of information, or information quality, about workers' productivity. The model was developed by AIGNER and CAIN (1977) and extended by LUNDBERG and STARTZ (1983) and by LUNDBERG (1991). There are three major consequences when differences in information quality are observed. First, employers expect lower productivity of a disadvantaged group if the information

uncertainty about a disadvantaged group is comparatively higher than for the advantaged group. Second, if employers have difficulty identifying the productivity of a disadvantaged group, then wages of the disadvantaged group cannot be directly linked to productivity. This implies that the disadvantaged group loses incentives to invest in further skills as their real productivity will not be recognized by employers anyway. Even if all workers have the same innate abilities, the disadvantaged group is on average less productive than the favored group in the equilibrium of these models. Third, imprecise information about workers' productivity impedes their mobility decisions. OETTINGER (1996) uses a dynamic approach with multiple periods to measure job mobility decisions. The author argues that wage gaps between type-*B* and type-*W* workers get larger over time because type-*B* workers are more uncertain about taking new job offers as employers constantly undervalue type-*B* workers' productivity, which also impedes mobility decisions of type-*B* workers.

Based on ALTONJI and BLANK (1999, p. 3189-3190), I introduce the statistical discrimination model with information uncertainty, which originated with LUNDBERG and STARTZ (1983). LUNDBERG and STARTZ (1983) assume that every worker has a marginal productivity (MP_i), which depends on his/her innate abilities (a_i) and on his/her accumulated human capital (e_i) so that

$$MP_i = a_i + e_i \,. \tag{2-8}$$

Every worker knows about his/her innate abilities and invests in human capital based on the marginal cost for investing in additional skills and on the corresponding marginal increase in wages.

The marginal cost depends on the accumulated human capital, so that

$$C'(e_i) = ce_i \,, \tag{2-9}$$

c is assumed to be equal for the entire workforce. Employers observe whether a worker is type-*A* or type-*B* and assign a productivity indicator (θ_i) for each worker. The productivity indicator depends on the worker's marginal product and on an error component, so that

$$\theta_i = MP_i + \varepsilon_i \,. \tag{2-10}$$

A worker receives a wage $w_i = E(w_i \mid \theta_i)$, which depends on his/her productivity indicator. With the assumption of jointly normal and independently distributed errors, the wage is

$$w_i = \overline{MP} + \beta(\theta_i - \overline{\theta}) \,. \tag{2-11}$$

β stands for the effect of an additional investment in human capital on wages, *ceteris paribus*. Assuming that c and the average value of a_i are the same for groups *A* and *B*, but that θ is less accurate for group *B* than for group *A* ($\beta_B < \beta_A$),

there are group differences in the average w_i. This means that statistical discrimination as a result of information inequality about productivity leads to lower average wages for the group with lower productivity indicators. There is statistical discrimination if employers use different wage equations for each group. As a result of statistical discrimination against type-*B* workers, the returns on human capital for type-*B* workers are also lower than for group *A*. In what seems a rational response, type-*B* workers decrease investments in human capital. The lower productivity indicator of type-*B* workers is, thus, self-confirming. This result is similar to the one obtained by COATE and LOURY (1993).

LUNDBERG and STARTZ (1983) argue that affirmative action policies reduce the differences in human capital investment and wages between the two groups. They suggest that if the government implemented a law which forbids employers to use different wage equations based on the productivity indicators of the two groups, human capital and wage differences would disappear. This is in sharp contrast to COATE and LOURY (1993), who postulate that affirmative action policies actually reduce employment standards for type-*B* workers; the result is that type-*B* workers have no incentive to invest in additional human capital, and, therefore, the initial negative belief about their abilities is self-confirming.

Language

In view of the many language communities in the US, LANG developed a language theory of discrimination, he defines language as "all aspects of verbal and non-verbal communication by which individuals transmit information" (LANG, 1986, p. 364). In LANG's model there are two possible situations between two language groups; the two language groups can communicate, or they cannot communicate. The marginal cost of learning a new language is constant for all actors, which are employers and workers in LANG's model. Actors have no utility gain from learning a new language. As learning a new language is expensive, the market tends to limit interactions of actors speaking different languages. In an extreme scenario there will be complete occupational segregation between language communities. There will only be interaction between the two language communities if the utility of learning a new language exceeds the transaction costs of learning the language.

I will now introduce two cases of LANG'S language theory of discrimination. In the first case he assumes that workers speak either white (*W*) or black (*B*). Given that the capital-labor ratios of the two language groups vary, the two language groups trade with each other. The author assumes that whites have more capital relative to labor. First the author analyzes the situation when a white employer hires a fixed labor force (LANG, 1986, p. 370). A mixed workforce is costly as part of the workforce is required to learn a new language. The white employer has three hiring alternatives: 1) only whites, 2) blacks who speak *W* or 3) monolingual blacks who speak only *B* and learning *B* her-/himself. Each of these three alternatives

comes with different costs (c) for the white employers, which are given in equations (2-12) to (2-14). If the white employer hires only whites, then his/her costs are

$$c = nw_w \,, \tag{2-12}$$

where n stands for the number of hired workers and w_w stands for the wage of whites. If the white employer wants to employ bilingual blacks, then s/he has to compensate them for their additional language skills. The employer, thus, has to pay

$$c = n(w_b + d) \,, \tag{2-13}$$

where w_b is the wage for blacks and d is a compensation for learning a new language. If the white employer, however, decides to employ only monolingual blacks, then the employer has to invest in learning B language, so that the costs will be

$$c = nw_b + d \,. \tag{2-14}$$

Equations (2-13) and (2-14) indicate that the costs given in (2-13) are larger than in (2-14) when $n > 1$. The white employer will, hence, learn B language and hire monolingual blacks instead of hiring bilingual blacks. The wage gap between blacks and whites is derived from equations (2-12) and (2-14). LANG sets both equations equal and identifies the black-white wage gap as

$$\frac{d}{n} = w_w - w_b \,. \tag{2-15}$$

In the model the white employer who hires only monolingual blacks receives a compensation for his/her effort to learn B language and makes higher profits. In a real world scenario the white employer, however, has to pay his/her compensation from his/her own profit, so that actual results are more complex than in LANG's model; moreover, in my view in an economic situation with many redundant workers, employers may not compensate blacks for additional language skills. White employers usually require the same language skills for all workers, independent of racial background.

In the second case LANG (1986) assumes that a white employer requires two types of employees (supervisors and workers) who need to communicate. Every supervisor controls n workers. Altogether the employer has four hiring alternatives. S/he can employ 1) a white supervisor and white workers, 2) a black supervisor and white workers, 3) a white supervisor and black workers or 4) a black supervisor and black workers. LANG (1986) again assumes statistical discrimination by race based on different wage equations. The wage for white supervisors (w_{ws}) is

$$w_{ws} = w_w + i \,, \tag{2-16}$$

where *i* stands for additional income, a compensation differential, which supervisors receive for additional skills required for the position. The employer has thus total labor costs

$$c = (n+1)w_w + i\,. \tag{2-17}$$

In the case of hiring a black bilingual supervisor and black workers, the employer has to pay the corresponding wages and compensation differentials for the task of supervising and for learning a new language. The wage for black supervisors (w_{bs}) is, thus,

$$w_{bs} = w_b + i + d\,. \tag{2-18}$$

The employer has to pay total labor costs

$$c = w_{bs} + nw_b\,. \tag{2-19}$$

To derive the black-white wage gaps, the equations have to be equal. The author sets equations (2-17) and (2-19) equal and inserts the corresponding wages for supervisors and workers into both equations. This leads to the black-white wage gaps for supervisors (2-20) and workers (2-21), respectively,

$$w_{bs} = w_{ws} + nd/(n+1) \tag{2-20}$$

and

$$w_b = w_w - d/(n+1)\,. \tag{2-21}$$

Equation (2-20) suggests that black supervisors receive higher wages when they supervise black workers than they would receive for supervising white workers. The author, however, argues that it would not pay off for white supervisors to learn language *B* as the compensation differential for learning a new language is insufficient. The market tends to segregate in a way that black workers are matched with black supervisors and white workers with white supervisors driven by the fact that transaction costs are minimized when maximal segregation occurs. As in the BECKER model of "taste-based discrimination", LANG's language model also leads to market segregation. The major conclusion of the model is that only a reduction of language differences will decrease black and white wage-gaps and occupational segregation. This makes sense from an economic point of view, but considering human heritage the theory will become more complex.

2.2.2.4 Occupational segregation

Up to now most discussions of occupational segregation have focused on the role of women in the labor market and, to a lesser extent, on racial or ethnic differences.[3]

[3] Along with "occupational segregation" the terms "occupational exclusion" and "occupational dissimilarity" are used for analyzing occupational distributions (JOHNSON and STAFFORD, 1998, p. 72).

There are basically four reasons for occupational differences: 1) employer discrimination, 2) institutional discrimination, 3) abilities and 4) preferences (ALTONJI and BLANK, 1999, p. 3176).

The most important scholars in this research area are BERGMANN (1974) and JOHNSON and STAFFORD (1998). BERGMANN (1974) analyzes the effects of occupational segregation when employers discriminate by gender and race. Her analysis was extended by JOHNSON and STAFFORD (1998), who concentrate on gender segregation in the US. JOHNSON and STAFFORD (1998) analyze the effects of gender differences on employer discrimination, human capital and institutional constraints. Based on ALTONJI and BLANK (1999, p. 3177-3179), I introduce the occupational exclusion model, which originated with JOHNSON and STAFFORD (1998).

The authors suppose a hypothetical economy with two available occupations (*j=1,2*) for producing a single good. The workforce consists of men (*m*) and women (*f*), *g=m,f*, where *g* stands for gender. All women are λ_j as productive as men in job *j*. Women have a comparative advantage in the second occupation, so that $\lambda_2 > \lambda_1$. Women's performance in the second occupation is relatively better than men's performance in the second occupation. The aggregated flow of labor in each occupation is

$$N_j = L_{mj} + \lambda_j L_{fj}, \qquad j = 1,2 . \tag{2-22}$$

The marginal product of labor, the additional worker employed, is denoted by *G*. For both occupations *G* depends on the amount of labor (*N*). For occupations *1* and *2* the marginal products of labor are, thus, $G_1(N_1, N_2)$ and $G_2(N_1, N_2)$, respectively. The model assumes employers have the same hiring preferences regarding gender. In this way the model is not confronted with the critical question of whether or not prejudiced employers can last in the long run (ALTONJI and BLANK, 1999, p. 3177).

Along with gender differences in comparative advantages, JOHNSON and STAFFORD (1998) assume that employers discriminate against women and have a disutility (d_1 or d_2) for hiring women in occupations 1 or 2, respectively. Employers hire men until their marginal products equal their wages in both occupations,

$$W_{m1} = G_1, \quad W_{m2} = G_2 . \tag{2-23}$$

Women are hired until their wages equal their marginal product altered by disutility components and productivity factors, so that

$$W_{f1} = (1 - d_1)\lambda_1 G_1, \quad W_{f2} = (1 - d_2)\lambda_2 G_2 . \tag{2-24}$$

To control for wage effects on labor supply, JOHNSON and STAFFORD (1998) assume that the labor supply for both men and women is fixed and that the labor market clears. The total labor supply is, thus,

$$L_g = L_{g1} + L_{g2} . \tag{2-25}$$

The relative labor supply for occupations *1* and *2* depends on the relative wages and preferences for men and women. The *desired* relative labor supply of group *g* is,

$$\frac{L_{g1}^{s}}{L_{g2}^{s}} = \theta_g \psi_g \left(\frac{W_{g1}}{W_{g2}} \right), \tag{2-26}$$

where θ_g is a taste parameter which differs between men and women and $\psi'(\cdot) > 0$ (ALTONJI and BLANK, 1999, p. 3178). The authors provide no additional information about the distribution of ψ. To measure the actual relative labor supply, however, institutional constraints on women's employment in job 1 are implemented in the model; examples of the sorts of pressures or constraints on women's employment would include social pressure or even prohibition of women doing what is typically men's work. These constraints are captured with X_g. The authors assume that the *actual* labor supply is the product of the *desired* relative labor supply (equation 2-26) and X_g,

$$\frac{L_{g1}}{L_{g2}} = X_g \frac{L_{g1}^{s}}{L_{g2}^{s}} = X_g \theta_g \psi_g \left(\frac{W_{g1}}{W_{g2}} \right), \qquad g = m, f\,. \tag{2-27}$$

ALTONJI and BLANK (1999, p. 3178) assume the extreme case of prohibition of women in occupation *1*. This implies that $X_f = 0$ and that $L_{f1} / L_{f2} = 0$, as no woman is allowed to work in occupation *1*.

Based on previous assumptions and equations, JOHNSON and STAFFORD (1998) derive the wage ratios for men and women based on their marginal products, respectively,

$$\frac{W_{m1}}{W_{m2}} = \frac{G_1(N_1, N_2)}{G_2(N_1, N_2)}, \qquad \frac{W_{f1}}{W_{f2}} = \frac{(1-d_1)\lambda_1 G_1(N_1, N_2)}{(1-d_2)\lambda_2 G_2(N_1, N_2)}. \tag{2-28}$$

As JOHNSON and STAFFORD (1998) assume that women have a comparative advantage in occupation *2* ($\lambda_2 > \lambda_1$) and/or face larger employer discrimination in occupation *1* ($d_1 > d_2$) the wage ratio is larger for men than women. The shares of each group in occupation *1* are given as

$$P_{g1} = \frac{L_{g1}}{L_g} = \frac{X_g \theta_g \psi_g \left(\frac{W_{g1}}{W_{g2}} \right)}{1 + X_g \theta_g \psi_g \left(\frac{W_{g1}}{W_{g2}} \right)}. \tag{2-29}$$

The authors calculate the index of occupational dissimilarity on the basis of equation (2-29), which is denoted as $D = P_{m1} - P_{f1}$ and which gives the gender difference in the occupational distribution. Several important conclusions can be drawn from *D*. The following conclusions are based on the assumption that all

other effects are equal. First, *D* decreases if women's comparative advantage in occupation *1* (λ_1) increases, so that the ratio λ_1 / λ_2 increases. Second, *D* decreases if women face lower disutility *d* (employer prejudice) in occupation *1*, which implies that the ratio $(1-d_1)/(1-d_2)$ decreases. Third, if only one of these cases, women get more competitive in occupation *1* or face lower employer prejudice in occupation *1*, occurs, then women's wage ratio W_{f1}/W_{f2} increases relative to W_{m1}/W_{m2} (cf., equation 2-28). The relative supply of women to occupation *1*, hence, increases. Fourth, *D* decreases if women increase their taste for occupation *1* as the ratio θ_f / θ_m increases. Fifth, *D* decreases if X_f / X_m increases, which implies that social pressure and/or institutional constraints decrease.

JOHNSON and STAFFORD's (1998) model facilitates the analysis of occupational exclusion of disadvantaged groups based on employer discrimination, tastes, institutional constraints and social norms. This model is, however, limited by the underlying assumptions. A major weakness is that the causes for occupational segregation are not clearly postulated. Researchers must dig deeper to better understand the underlying factors behind the four reasons for occupational exclusion.

2.2.3 Occupational choice theory

The theoretical framework of occupational "choices" is based on the assumption that individuals act as *homo oeconomicus* in their job decisions. This means that individuals know about all available occupations in their occupational choice sets, evaluate each occupation on the basis of its characteristics, associate a level of satisfaction with each occupation, compare the occupations based on their perceived satisfaction levels and choose the most attractive occupation given environmental constraints.

Recently five new models have contributed to the theoretical rigor of occupational choice literature. BROWN et al. (2008) develop a model to untangle supply and demand in occupational choice. ASTEBRO et al. (2008) develop an occupational choice model which focuses on self-employment. JACOBS (2007) develops a static occupational choice model for developing countries. DROST (2002) incorporates both dynamics in occupational choices and the risk of unemployment. KIMURA and YASUI (2006) develop an overlapping generations model, combining occupational and fertility choices.

The standard approach of choice theory assumes that an individual (i) acts as *homo oeconomicus* in his/her job decision. This means that i knows the occupations (j) available in his/her choice set ($C(q) \in C$); i evaluates each occupation $j \in C(i)$ on the basis of its characteristics (X_j); i associates a level of satisfaction to each occupation; i compares the occupations based on the perceived level of satisfaction and chooses the most attractive occupation given environmental constraints. Econometricians are, however, unable to observe this whole

decision-making process and, hence, treat satisfaction levels as random utility. Given the imperfect information, utility is divided into an explained and an unexplained part. To analyze occupational outcomes, it is common practice to use a random utility function, such as

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \tag{2-30}$$

where V_{ij} is the observable part and expressed by the explanatory variables and ε_{ij} is the random part and unspecified in the observed part of the random utility function. It is crucial to notice that the various multinomial discrete choice models depend on assumptions made on distributions and variations (among *j* and/or *i*) of the error component ε_{ij}.

The rationality assumption in occupational choice theory is regarded as very unrealistic. The existence of bounded rationality, which includes limitations in information, time and cognitive abilities in decision making, is seen as more realistic by some researchers.[4] While these so-called occupational "choice" models assume underlying random utility functions to capture job decisions, vocational psychologists assume that individuals constantly change their potential occupations without an underlying utility function. GOTTFREDSON's (1981) theory of circumscription and compromise provides an insight into the processes of how vocational choices are developed from birth to adolescence. The author assumes that with increasing age young people adapt their social space of potential occupations depending on their interests, goals, skills, abilities and temperament.

Researchers are, nevertheless, often limited to the usage of occupational choice theory which follows the logic of utility maximization and can be directly linked to mathematical modeling of decision making. Random utility functions are, thus, widely accepted to measure occupational choices without consideration for actual decision-making processes.

2.2.4 Farm household theories

As far more than half of the rural population in Guizhou works in agriculture, theories which account for this situation are required. Farm households have the particularity that they make not only consumption, but also production, decisions, which involve labor allocations on or off the farm. The general assumption in farm household models is that a household seeks to maximize utility from final consumption subject to constraints in the household's production function, time allocation and monetary income (ELLIS, 1993, p. 123-145).

---

[4] For example, SIMON (1956), KLEIN (2001), GIGERENZER and SELTEN (2001) and REINA (2005) discuss the concept of bounded rationality.

Farm household models are usually based on strong assumptions regarding the functional form of utility and constraints. In reality it is far more difficult to determine the behavior of households and their job decisions. The basic model does not, for example, account for separate interests of individual household members and possible changes of individual preferences over time. It also does not account for households with different ethnic affiliations, yet employer discrimination against ethnic minorities is a potential barrier for off-farm work and, thus, influences time allocations of ethnic minority households. I present a standard farm household model introduced by ELLIS (1993, p. 123-145) in section 2.2.4.1 and depict the model graphically in section 2.2.4.2.

2.2.4.1 Model specification

BECKER (1965) extends the basic farm household theory of CHAYANOV (1986) with "New Home economics" theories. I briefly describe one branch of these theoretical extensions following ELLIS (1993, p. 123-145), who bases theoretical specifications on MICHAEL and BECKER (1973). In farm household models the observational unit is not a single individual, but the overall agricultural household. The goal of the household is to maximize utility (*U*) from final consumption subject to constraints in the household production function (*Z*), time allocation (*T*) and monetary income (*Y*).

The utility function of the household is given as

$$U = f(Z_1, Z_2, \ldots, Z_n), \tag{2-31}$$

where *Z* stands for *Z-goods* which are produced for household consumption and not for market consumption.

The constraint in a household's production function of *Z-goods* is based on the time spent (T_i) and necessary inputs, goods and services (x_i), purchased on the market. The home production function is

$$Z = f(x_i, T_i). \tag{2-32}$$

The total time constraint takes the form

$$T = T_w + \sum T_i, \tag{2-33}$$

where T_w is the time used for wage employment outside of the farm and T_i is the time used for producing *Z*-goods on the farm.

The time used for wage work (T_w) and the corresponding wage rate (*w*) determine the monetary income constraint (*Y)* of the household. In equilibrium *Y* equals the value of inputs used for producing *Z*. The value of inputs (x_i) is determined by multiplying each item with corresponding prices (p_i). The income constraint is thus

$$Y = wT_w = \sum p_i x_i \quad (2\text{-}34)$$

The income and time constraints form the "full income constraint" (*F*) by evaluating the household's time (*T*) at the market wage rate,

$$F = wT = w\sum T_i + \sum p_i x_i \quad (2\text{-}35)$$

The ratio of marginal utility received from producing *Z*-goods, meaning the marginal rate of substitution between any pair of *Z*-goods, equals the ratio of their marginal production costs.

Depending on the theoretical model setting, consumption and production decisions are either made simultaneously or sequentially by the household. When a competitive labor market functions perfectly, then separability of households' production and consumption decisions can be applied. In this model setting, "farm households first make the optimal farm production decisions, and then decide on the optimal level of consumption and leisure" (WANG, 2007, p. 22). When the labor market, however, functions imperfectly, which is usually the case, then production and consumption decisions are simultaneously determined (WANG, 2007, p. 22).

Some researchers analyze time allocation of rural households between home time, farm work and off-farm work by implementing risk behavior (FINKELSHTAIN and CHALFANT, 1991; FAFCHAMPS, 1992), credit constraints (DE JANVRY et al., 1991) and transaction costs for accessing product markets (KEY et al., 2000) into the model. Other researchers analyze the supply of off-farm work by the household head (HUFFMAN, 1980; SUMNER, 1982), interactive household employment decisions between husbands and wives (HUFFMAN and LANGE, 1989; TOKEL and HUFFMAN, 1991; LASS and GEMPESAW II, 1992; SKOUFIAS, 1994; SADOULET et al., 1998; AHITUV and KIMHI, 2002; BENJAMIN and KIMHI, 2006), and joint decisions of hiring external workers and off-farm work of household heads (BENJAMIN et. al., 1996; FINDEIS and LASS, 1994). LOW (1986) distinguishes among household members and their comparative advantages for wage work in his study area (countries bordering South Africa) where subsistence farming is basically done by women, children, the aged and infirm, while men pursue wage employment.

2.2.4.2 Graphical depiction

ELLIS (1993, p. 128-129) explains the home production model based on the assumptions that 1) a household produces only one single good *Z*, 2) *U* is based on maximizing *Z* and leisure time, and 3) a single price (*p*) is used for assessing inputs (see figure 2-2). In figure 2-2 the time constraint is given on the horizontal axis. It shows the time endowment ($T_)$ of the household divided into time used for home work (T_Z), for wage work (T_w) and for leisure (T_H). The opportunity cost of time is based on the real market wage (w/p). The line *0-F* shows that total real income increases with time available for wage work and leisure. At point *F*

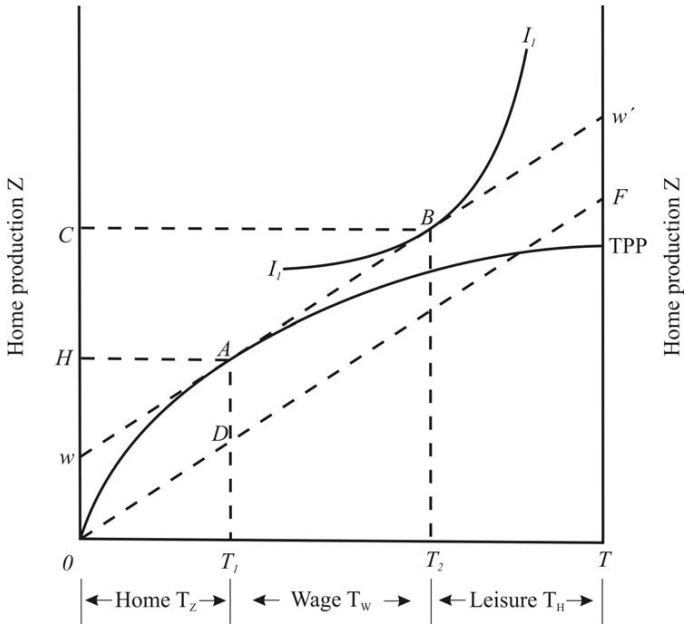
the total available time is valued at the real wage rate, which gives the full opportunity costs of time for the household (wT / p).

Figure 2-2 further shows the total production possibility frontier (*TPP*) of the household for producing *Z* in relation to *T* used, an indifference curve *I*, which gives utility levels related to possible combinations of T_H and production of *Z* and a wage line (ww') tangent to *I* and *TPP*. The wage line provides the opportunity costs of *T* at market prices.

At point *A* the household production of *Z* is in equilibrium. This means that the marginal physical production (*MPP*) of home work, which is the slope of the *TPP*, equals the real wage rate ($MPP = w / p$). This is the part of the total cost the household has to pay for producing *Z* at point *H*. At point *B* the figure shows the household consumption of *Z* in equilibrium. It gives the marginal rate of substitution between leisure and consumption of *Z*, (MU_L / MU_Z).

The money income constraint (*Y*) is shown graphically as the distance *CH* on the vertical axis. This indicates that money used for inputs cannot exceed the market wage (*w*) multiplied by the time used for wage work (T_w). The household's "full income" is given by moving distance *0F* upward to ww'; distance *AD* is the profit.

Figure 2-2: The home production model



Source: ELLIS (1993, p. 129).

2.2.5 Non-farm rural employment theories

BUCHENRIEDER and MÖLLERS' (2006) integrated theoretical framework captures the major kinds of theories for analyzing NFRE. Labor market discrimination as a potential rural development issue was, however, not considered in their theoretical framework, yet labor market discrimination constrains access to NFRE and/or is a crucial distress-push factor for accessing NFRE for racial or ethnic minorities, women and the elderly.

In the next section I closely follow BUCHENRIEDER and MÖLLERS (2006) to introduce their integrated theoretical framework, which links theoretical contributions of the sustainable livelihood framework (SLF), demand-pull and distress-push concept, a welfare model and a behavioral theory. In section 2.2.5.1 I present the SLF and demand-pull and distress-push concepts as well as their linkages. In section 2.2.5.2 I explain demand-pull and distress-push movements in a welfare model developed by BUCHENRIEDER and MÖLLERS. In section 2.2.5.3 I present AJZEN's (1985) theory of planned behavior as part of the integrated theoretical framework.

2.2.5.1 The sustainable livelihood framework and demand-pull/distress- push concepts

The sustainable livelihood framework (SLF) explains livelihood strategies in a framework considering access to capital, institutions and structures. It considers how capital assets, such as natural, physical, human, social and financial assets, linked to the socio-economic structure of society and to formal and informal institutions influence which kind of non-farm jobs individuals obtain. To better understand labor movements away from agriculture towards various occupations in the rural non-farm sector, demand-pull and distress-push concepts are used for explaining varying outcomes (cf., EFSTRATOGLOU, 1990; BARRETT et al., 2001). NFRE is usually characterized as "highly lucrative at the top end with mainly formal wage employment and modern capitalized enterprises, but very menial at the bottom end, where traditional artisan skills and poorly paid manual labor predominate" (START, 2001, p. 496).

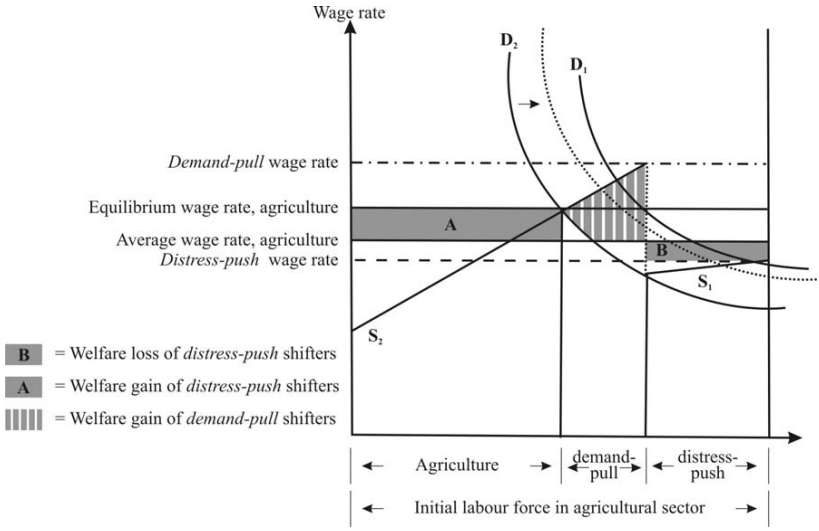
The demand-pull process describes the case in which a former agricultural worker receives better jobs at the top end in the rural non-farm economy, while the distress-push process describes the case in which a former agricultural worker is pushed into poorly paid non-farm jobs. Which class of jobs a worker can access depends strongly on his/her capital assets interlinked with the socio-economic structure of society and formal and informal institutions as given in the SLF. It is obvious that particularly demand-pull processes can bring advantages in terms of income diversification to the household. The distress-push mechanisms, however, also can increase total household income, reduce vulnerability and improve households' risk management (START, 2001).

MÖLLERS (2006, p. 496) brings together several factors which influence demand-pull and distress-push situations in line with the SLF. Labor market discrimination against ethnic minorities is, however, not considered an important determinant of rural employment decisions. Labor market discrimination is, however, indeed, a crucial distress-push factor for some ethnic groups, for example, Tibetans and Uyghurs in China, and should be included in the analysis of NFRE. Labor market discrimination even forces some ethnic minorities to remain in agriculture because employment in low level non-farm jobs is denied in some cases (cf., subchapter 1.1).

2.2.5.2 Demand-pull and distress-push dynamics in a welfare model

BUCHENRIEDER and MÖLLERS (2006) explain labor allocation processes induced by demand-pull and distress-push dynamics in a welfare model (see figure 2-3). The main result of their model is that either demand-pull or distress-push factors both bring welfare benefits to the household and society, which is in line with START (2001). BUCHENRIEDER and MÖLLERS's model consists of two labor supply curves. S_1 and S_2 are the labor supply curves for distress-push shifters and for demand-pull shifters, respectively. By shifters the authors mean workers who change or shift from agricultural to non-agricultural work. Demand-pull shifts occur when the feasible wage rate in the non-farm sector is higher than the average wage rate in agriculture. Distress-push shifts occur when the wage rate in the non-farm sector is no higher or even lower than the average wage rate in agriculture.

Figure 2-3: A basic model of welfare gains with demand-pull and distress-push labor shifts



Source: BUCHENRIEDER and MÖLLERS (2006, p. 6).

The authors assume that household members who receive an income from non-farm employment contribute the entire income to the household in the short run. Household members, who are represented with S_1, are unable to shift to S_2 due to high shifting costs (e.g., as a result of lacking capital assets, inadequate state structures and institutions). Although the distress-push wage rate can be lower than the average wage rate in agriculture, household members in a distress-push situation will only shift their labor to low-paid, non-farm jobs until S_1 intersects *D*, the labor demand curve in agriculture. Those household members with zero or no opportunity costs of agricultural labor can work at the lower distress-push wage rate in order to increase aggregated household welfare.

The difference between the shaded areas *A* and *B* indicates the welfare gain induced by the distress-push shifters, where *A* gives the wage gain for those who stay in agriculture and *B* the wage loss for those who leave the agricultural sector. BUCHENRIEDER and MÖLLERS's model show that it is a rational decision of the household to diversify income sources as total household income will increase even if some household members receive only below-average wages.

S_2 shows shifts from agriculture to better-paid, demand-pull employment. The demand-pull wage rate is higher than the equilibrium wage rate in agriculture. Workers, thus, change from agriculture to the demand-pull sector until the wage difference exceeds the shifting costs. The striped triangle in figure 2-3 shows the welfare gain from the labor shift. The labor shift from agriculture to the demand-pull sector increases the average wage rate for workers staying in agriculture and, hence, reduces workers incentive to shift. The dotted line in figure 2-3 illustrates this case. At the point where S_2 intersects with the demand-pull wage rate, there are no shifting costs.

2.2.5.3 Theory of planned behavior

AJZEN's (1985) theory of planned behavior is another element of BUCHENRIEDER and MÖLLERS's theoretical framework. AJZEN's theory incorporates the decision-making processes which underlie the different outcomes of NFRE; it uses attitudes, norms and behavioral constraints to explain rational decisions of utility-maximizing households.

AJZEN assumes that three "beliefs", behavioral beliefs, normative beliefs and control beliefs, direct human actions. Behavioral beliefs result in a subjective attitude towards a decision. For example, if a household has the belief that NFRE brings secure and high wages, this leads to a positive attitude towards NFRE. In AJZEN's theory, normative beliefs express the expectations of others regarding the decision and the willingness of someone to fulfill them. Normative beliefs capture how the social environment (e.g., family, friends, peers, etc.) influences individual behavior. For example, on a successful farm parents may expect that a child will continue farming, which may influence his/her decision to stay in or leave the agricultural sector. Control beliefs influence someone's evaluation whether or not a

decision is genuinely suitable for himself/herself in relation to his/her capital assets as given in the SLF (e.g., natural, physical, human, social and financial assets). For example, if an individual has insufficient education for the desired non-agricultural job, s/he will not be able to work in the non-agricultural job despite his/her positive attitude and despite positive family expectations towards the non-agricultural job.

These three beliefs along with their underlying concepts continue to form the behavioral intention of individuals. The more positive the attitude, and the higher the subjective norm and perceived control towards NFRE, the more probable it is that a household member will shift away from agriculture towards NFRE (BUCHENRIEDER and MÖLLERS, 2006).

2.2.6 Interim conclusions

The available theories for explaining ethnic differences in occupations and wages can broadly be divided into two types: those which focus on pre-labor market group differences in preferences and skills and those which focus on employer discrimination in the labor market. The market actors are employers and employees who both seek to maximize their utility based on a set of assumptions which differ depending on the theory.

Taste-based discrimination models pioneered by BECKER (1957, 1971) assume that a marginal discriminatory employer has distaste against black workers and will, therefore, pay a wage premium to white workers. Under these models the market segregates between prejudiced employers who only employ white workers and unprejudiced employers who mainly employ black workers; these models further project that prejudiced employers make lower profits and will be driven out of the market in the long run and that wages will equalize between races. The theory is, therefore, unable to explain persistent wage differences.

BLACK's (1995) equilibrium search model of employer discrimination, which falls in the category of taste-based discrimination, predicts that unprejudiced employers are aware that black workers face higher job search costs as the proportion of unprejudiced employers among the total employers is comparatively small. Under this model unprejudiced employers, therefore, offer lower wages to black workers than to white workers. This further indicates that even if there are only unprejudiced employers in the market, black workers will face wage discrimination. In contrast, as in the BECKER model, prejudiced employers hire only white workers for higher wages, make, therefore, lower profits than unprejudiced employers and will be driven out of the market in the long run.

Statistical discrimination models pioneered by ARROW (1971, 1973) and PHELPS (1972) analyze hiring and wage decisions of employers with imperfect information about workers' performance. They assume that, based on limited information about worker performance, employers statistically discriminate by using race

or ethnic status to make predictions about job performance of all workers of the same group. COATE and LOURY (1993) analyze differences in employers' prior beliefs about workers' productivity. The authors find that employers' initially negative beliefs about job performance of blacks can be self-confirming when black workers are not willing to invest in additional training because they know that they are assigned to lower skilled positions anyway. LUNDBERG and STARTZ (1983) come to similar conclusions assuming differences in information quality about individual productivity. LANG (1986) developed a language theory of discrimination and finds that different language communities will only interact if the utility of learning a new language exceeds the transaction costs of learning that language; only if language barriers are reduced, will wage-gaps between and occupational segregation of ethnic groups decline. LANG's theory is straightforward from an economic point of view, but it is in contrast to preserving human heritage as local languages tend to disappear.

While taste-based discrimination and statistical discrimination theories focus on the labor demand side (e.g., distaste, prior beliefs or information uncertainty) with strong underlying assumptions, the occupational exclusion model of JOHNSON and STAFFORD (1998) manages to combine the important factors of the theories of group differences, taste-based and statistical discrimination. The authors assume that there are four major causes of occupational segregation: employer discrimination, institutional discrimination, workers' abilities and workers' preferences. These concepts contribute to the understanding of ethnic differences in competitive labor markets; however, they give no theoretical guidance about how different job preferences evolve, and they cannot be directly applied to rural labor markets with large shares of subsistent farmers. To capture rural labor market peculiarities, farm household models and non-farm rural employment theories can be used.

In farm household models the goal of the household is to maximize utility from final consumption, subject to constraints in the household's production function, time allocation and monetary income (ELLIS, 1993, p. 123-145). The basic farm household models follow a neoclassical logic. If wage work off the farm provides higher income than farm work, a farm household decides to hire in external workers on the farm and reduces its own time for farm work. The rural non-farm sector can be analyzed with non-farm rural employment theories. The demand-pull and distress-push concepts are one branch of the theoretical framework of NFRE (BUCHENRIEDER and MÖLLERS, 2006). The demand-pull process describes cases in which former agricultural workers receive better paid non-farm jobs, while the distress-push process describes cases in which former agricultural workers receive poorly paid non-farm jobs. It can be shown in a welfare model that both demand-pull and distress-push factors bring welfare benefits to the household and society (BUCHENRIEDER and MÖLLERS, 2006).

Finally theories regarding evolution of job preferences shed some light on the decision-making processes behind occupational outcomes and wages. AJZEN's (1985)

theory of planned behavior assumes that three beliefs, behavioral beliefs, normative beliefs and control beliefs, guide human actions. In this model individuals think about decisions based on their subjective attitudes, the perceptions of the social environment regarding the decision and check whether or not their given capital assets allow them to make the decision. The theory of circumscription and compromise by GOTTFREDSON (1981) postulates that job preferences evolve when young people adopt their social space of potential occupations, depending on their interests, goals, skills, abilities and temperament while they are getting older. While all theoretical approaches are important, authors rarely provide clear instructions on how to empirically measure their theories.

2.3 Theoretical linkages and reflections on China

In this subchapter I first link the theoretical concepts, then make theoretical reflections on labor market discrimination in China. As already pointed out, different occupational outcomes and wages among ethnic groups do not necessarily imply that ethnic minorities face discrimination in the labor market; it can also mean that ethnic minorities have lower abilities or simply prefer other jobs. The occupational exclusion theory of JOHNSON and STAFFORD (1998) combines the leading sources for different occupational outcomes and serves as a precedent for linking the relevant theoretical concepts discussed previously into one single framework. The framework is meant to highlight the theoretical complexity when ethnic differences in occupations and wages are observed. The framework serves as a basis for more thorough, theoretically guided empirical research on labor market discrimination.

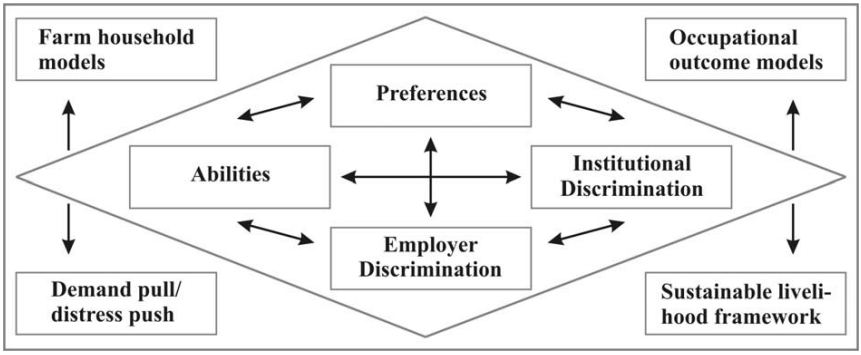
2.3.1 Theoretical linkages

Taking JOHNSON and STAFFORD (1998) as the benchmark model, it is possible to interlink the various theoretical approaches to get a comprehensive theoretical framework for analyzing ethnic differences in occupational outcomes and wages. It is also possible to apply the framework for analyzing racial discrimination, discrimination against women or the elderly, etc.

JOHNSON and STAFFORD (1998) assume four major causes for occupational segregation of a disadvantaged group: differences in employer discrimination, institutional discrimination, abilities and preferences. I use these four factors to arrange the theoretical concepts in a diamond[5] of theories (figure 2-4). These four major theoretical approaches for analyzing ethnic differences in occupations and wages are linked among each other with several causal directions. They can, furthermore, be implemented in other models such as farm household models, demand-pull/ distress-push concepts, occupational outcome models and the sustainable livelihood framework.

[5] I was inspired by PORTER (1990) who also uses a diamond to show theoretical interactions.

Figure 2-4: The diamond of theories for measuring ethnic differences in occupational outcomes and wages in an integrated theoretical framework



Source: Author.

Table 2.1 depicts the classification of some of the previously discussed theoretical frameworks based on the four major principles of occupational exclusion by JOHNSON and STAFFORD (1998). First, I classify theories which explain employer discrimination (third column of table 2.1); taste-based and statistical discrimination theories fall into this category (cf., subchapter 2.2.2). Employers can influence occupational and wage distributions of ethnic minorities through at least five discriminatory practices: distaste (prejudice), negative belief (stereotype), information uncertainty about productivity, negative signals about abilities and language discrimination. All these factors are constraints for ethnic minorities and can be directly implemented in other theoretical approaches as shown in figure 2-4.

For example employer discrimination impedes access to non-farm rural employment for ethnic minorities. Depending on the level of employer discrimination, ethnic minorities are, thus, in a distress-push situation and either work in badly paid non-agricultural jobs or stay in agriculture. Employer discrimination, therefore, also influences the time allocation of ethnic minorities in farm household models.

Second, I classify theories which fall within the range of institutional discrimination (fourth column of table 2.1). In China the laws (institutions) are in favor of ethnic minorities (cf., subchapter 1.3). Whether affirmative action policies do in fact benefit ethnic minorities is still under discussion. LUNDBERG and STARTZ (1983) postulate that the government merely needs to implement a law which forbids employers to use different wage equations for ethnic groups, and human capital and wage differences will disappear. COATE and LOURY (1993), however, suggest that affirmative action policies will actually reduce employment standards for ethnic minorities given the "ethnic quota". Ethnic minorities would, thus, have no incentive to invest in additional human capital given the job privileges, and the stereotype about lower job performance of ethnic minorities will remain.

Table 2-1: Classification of theoretical concepts in the diamond of theories

Preferences	Abilities	Employer Discrimination	Institutional Discrimination
Theories of group differences	Theories of group differences	Taste-based discrimination theories	Affirmative action theories
Occupational choice theory	Human capital theories	Statistical discrimination theories	
Theory of circumscription and compromise			
Theory of planned behavior			

Source: Author.

Third, I classify theories which analyze differences in abilities (second column of table 2.1). As there are various approaches (LIST and RASUL, 2011; ALTONJI and BLANK, 1999), I use the terms "theories of group differences" and "human capital theories" to classify theories focusing on differences in abilities. Ethnic differences in abilities are closely linked to differences in comparative advantages, human capital accumulations and preferences (ALTONJI and BLANK, 1999). This indicates that there are linkages between preferences and abilities in the theoretical diamond (see figure 2-4). Educational attainment is one important human capital factor, it depends on child characteristics (including "innate ability"), household characteristics, school and teacher characteristics (quality) and costs related to schooling, where school and teacher characteristics (quality) and prices related to schooling are both linked to education policies and local community characteristics (LIST and RASUL, 2011, p. 140, based on GLEWWE and KREMER, 2006). Opportunity costs of schooling and expected returns on schooling are also closely linked to these factors (ALTONJI and BLANK, 1999).

Fourth, preferences evolve out of individual decision-making processes. As JOHNSON and STAFFORD (1998) give no guidance about how and why job decisions are made, applicable theories in this area are occupational choice theories, the theory of planned behavior (AJZEN, 1985), the theory of circumscription and compromise (GOTTFREDSON, 1981) and theories of group differences with focus on preferences (first column in table 2.1). Occupational choice theory assumes that individuals act as *homo oeconomicus* in their job decisions. This means that individuals know about all available occupations, evaluate each occupation on the basis of its characteristics, associate to each occupation a level of satisfaction, compare the occupations by their perceived satisfaction levels and finally choose the most attractive occupation given environmental constraints. AJZEN's (1985) theory of planned behavior requires analyzing the decision-making processes of individuals in relation to their subjective attitudes, the social environment and control factors. The theory of circumscription and compromise (GOTTFREDSON, 1981) requires

analyzing how young people adjust their occupational choices while growing up, depending on their interests, goals, skills, abilities and temperament. Evolution of different job preferences, moreover, depends on different child-rearing practices, educational systems, comparative advantages and human capital (ALTONJI and BLANK, 1999).

To sum up individually the theories contribute to a better understanding of ethnic differences in occupations and wages. Each theory, however, only identifies some determinants for explaining ethnic differences in occupations and wages. An integrated theoretical framework as shown in the diamond of theories in figure 2-4 is, therefore, essential for theoretically driven empirical research and may also give policy makers a comprehensive theoretical overview of the complex issue of labor market discrimination. Researchers should make well- founded assumptions to guide their investigations as there are often several causal directions among the four major theoretical concepts.

2.3.2 Reflections on China

In this subchapter I discuss ethnic differences in occupational outcomes and wages in China in light of the theoretical knowledge acquired in the last sections. This requires analyzing the available literature on China in relation to the theoretical framework (figure 2-4), focusing on the four theoretical branches: preferences, abilities, employer discrimination and institutional discrimination.

Regarding preferences there is actually no available information about how job decisions evolve. Most researchers assume that individuals, independent of their ethnic affiliations, prefer higher wages from non-agricultural employment, rather than lower profits from agriculture. This is a reasonable assumption given significant wage gaps among industries and between rural and urban areas in China. The CHINA LABOR BULLETIN notes that in 2009 the average monthly wage in primary industry was only 1,196 CNY, while it was 4,846 CNY in computer services and 5,033 CNY in financial services. At the beginning of 2011, the annual per capita disposable income of rural households was only around 5,153 CNY, while the average disposable urban household income was around 17,175 CNY. There is, however, still analysis to be done on distress-push and demand-pull mechanisms within the non-farm rural sector, particularly given that an unskilled laborer may even earn less in non-agricultural employment than in agriculture (BUCHENRIEDER and MÖLLERS, 2006).

Given the assumption that, on average, individuals independent of their ethnic affiliations prefer to work in non-agricultural employment, three possible constraints which may hinder ethnic minorities from working in the non-farm sector remain: institutional discrimination, employer discrimination and abilities. As previously stated, there is no institutional discrimination (discrimination by law) in China because there is an elaborated preferential policy framework, which assists ethnic minorities to acquire higher education and which strictly prohibits job discrimi-

nation against ethnic minorities (cf., subchapter 1.3). There is, moreover, some empirical evidence that some ethnic minorities have benefited from preferential policies in education; in the 2000 census, for example, the years of schooling of 14 ethnic minorities, including Korean, Manchu, Mongolian and Kazak groups, are above the national average (CHINA.ORG.CN, 2005). In contrast policies regarding employer discrimination seem to be poorly enforced in TAR or XUAR. There is evidence that Uyghurs and Tibetans face difficulties finding better-paying jobs in non-farm sectors (GILLEY, 2001; HILLMAN, 2008). This situation could be a result of employer discrimination and differences in abilities, the two remaining concepts of the integrated theoretical framework.

2.3.2.1 Evidence of ethnic differences in employer discrimination

Taste-based and statistical discrimination theories can be used for analyzing employer discrimination. To analyze taste-based and statistical discrimination theories in respect to China, I concentrate on the major discriminatory factors: distaste (prejudice) of marginal actors and negative beliefs (stereotypes) against the entire ethnic group, which are concepts of taste-based and statistical discrimination theories, respectively.

If there is statistical discrimination, which implies stereotypes against an entire ethnic group, then there must have been taste-based discrimination before, as prejudices of some actors could turn into generally accepted stereotypes. Some researchers find that there is statistical discrimination against Tibetans and Uyghurs. HILLMAN (2008) states that Tibetans are disadvantaged compared to better educated migrants from other provinces, even in the tourism industry, which is mainly devoted to Tibetan culture. In XUAR Uyghurs are not hired by Chinese companies (GILLEY, 2001). An article published in the THE ECONOMIST (2000) reports that Chinese companies in Urumqi give jobs to Han Chinese rather than to Uyghurs. This is consistent with HOPPER and WEBBER's findings (2009, p. 187), which show that in Urumqi Uyghur migrants work in those jobs on which Han look down. The literature, therefore, suggests that the Tibetans in TAR and the Uyghurs in XUAR face statistical discrimination. In contrast it seems that other ethnic minorities in XUAR are not discriminated against. GLADNEY (2004), for example, finds that in XUAR science and technology positions are filled by Uzbeks and Tartars instead of by Uyghurs because Uzbeks and Tartars make up a high percentage of the well-educated urban population.

In contrast to the situations in XUAR and in TAR, is the situation in Guizhou and Yunnan provinces in southern China. GUSTAFSSON and LI (2003) find that between 1988 and 1995 ethnic minorities had higher increases in per capita income than Han. The Bouyei and Miao have even a higher labor market participation probability than Han in the 2000 and in the 1982 census, respectively (MAURER-FAZIO et al., 2004, 2005). Although at odds with the received wisdom, the literature suggests that Han may face employer discrimination in Guizhou. There is, however, no evidence of statistical discrimination, a general stereotype against all Han workers, but there

is evidence of taste-based discrimination, a distaste of some employers against Han workers. Given that ethnic minorities mainly have job advantages in the growing tourism industry (BHALLA and QUI, 2006), it seems to be the case that Han are somehow excluded in this growing sector which is devoted to ethnic customs and traditions. For example, the Sani from the counties Lunan and Luliang in Yunnan province sell embroidery (e.g., "beautiful bags" in the center of Kunming and in the region of the stone forest (HARRELL, 1995, p. 65). Many other of Yunnan's ethnic minorities are, however, mainly working in agriculture and pastoralism complemented by "specialized work in trades such as mule-skinning, carpentry, basket-making, and coppersmithing" (MCKHANN, 1995, p. 51). Other ethnic groups, such as the Han, Bai, Lisu, Pumi, Tibetan, Hui, and Yi, which surround the Naxi and Mosuo areas in northwestern Yunnan have almost the same working patterns (MCKHANN, 1995, p. 51). A usual Nuosu family in Yunnan also primarily engage in subsistence agriculture with limited income from other sources (HARRELL, 1990, p. 529).

Koreans in Northeast China benefit to a greater extent than Han from increasing trade between China and South Korea (GLADNEY, 2004, p. 21). An important aspect is probably the language advantage of the Korean minority when doing business with South Korea. This is in line with LANG's language theory of discrimination (1986). Applying LANG's theory, Han will only learn Korean if the utility of learning Korean exceeds the transaction costs of learning Korean, *ceteris paribus*.

Another story can be observed for the Hui, the largest Muslim group in China residing throughout the country. In many places, for example, in Fujian, Shaanxi, Gansu and Ningxia, the Hui have a comparatively higher distribution in small private businesses and industries than do Han (GLADNEY, 2004). In the capital Lanzhou of Gansu province, the Hui, however, face discrimination in state employment (ZANG, 2008). There is, therefore, some taste-based discrimination against Hui workers, but no statistical discrimination, against Hui workers.

To sum up, the available literature suggests that in XUAR and in TAR the Uyghurs and Tibetans, respectively, face statistical discrimination. The case of the Hui, meanwhile, is quite diverse: in most of China they apply their entrepreneurial skills without any barriers, yet in Lanzhou there is distaste against Hui workers in state employment. In contrast the literature suggests that in Guizhou and Yunnan Han actually face taste-based discrimination. This can basically be explained by the fact that ethnic minorities have higher employment probabilities in the increasing tourism sector, which is devoted to local ethnic minority culture. As there are 55 classified ethnic minorities alongside the Han people in China, there is definitely a need for more empirical examinations of ethnic differences in occupational outcomes and wages.

2.3.2.2 Evidence of ethnic differences in abilities

There is no room here to reflect on the literature in light of the many theories explaining differences in abilities. In China the relationship between ethnicity, education and access to better jobs and income opportunities does not follow a clear pattern. In addition to being influenced by ethnic status, educational attainment is also closely linked to school availability, quality, expenses, Mandarin language skills, gender and opportunity costs of households.

School availability, quality and expenses

School availability and quality in China depend primarily on an individual's place of residence independent of ethnic status. Only primary schools (grades 1 to 6) are located in villages, while schools for obtaining grades 7 to 9 are available in small towns and schools for obtaining grades 10 to 12, in cities (KAI MING, 2003). The distance to secondary education schools can be very great for rural children, who are often ethnic minorities. The establishment of boarding schools in border, pastoral and mountainous regions improves this situation, but the study environment and living conditions in these schools are often poor (SAUTMAN, 1997).

Beyond the heterogeneous geographic availability of schools, financial accessibility of schooling also adds to the inequality of access to education. Tuition fees are additional to costs of school supplies and travel expenditures, so that education has slowly turned into a larger part of the household's budget (GUSTAFSSON and LI, 2003), reducing access probability to secondary or higher education for children from poorer families, many of whom are ethnic minorities from rural areas (WORLD BANK, 2009). Uyghur peasants in XUAR "do not see the point in educating their children to the end of primary school, since further education is not considered open to most peasants. But if they do not send their children to school, they are fined" (BELLÉR-HANN, 1997, p. 107).

Some girls of the Hui minority of Lijiashang in Ningxia Hui AR report that in their families, who for their Hui status are allowed to have three children, only boys attend school after elementary schooling because of high tuition fees (KOLONKO, 2005). This is also the case for many Yi families of Liang Mountains in Sichuan province, where the Yi account for 98 % of the population and Han for only 2 % (LUO, 2008). According to the author, Yi children stop schooling when parents are unable to work the fields themselves.

With the introduction of the K-9 rural education program meant universalize education until the age of nine, all ethnic groups are to be given equal access to this compulsory education (Education Law of September 2006). This law, indeed, stipulates the elimination of expenses for tuition and books in primary schools. The Rural Education Action Project (REAP), however, stresses that it is cumbersome to change the established educational system (REAP, 2008).

Mandarin language skills

Higher educational achievements are closely linked to fluent command of Mandarin. Mandarin is the main language in China and has been used in the Chinese educational system since the beginning of the 20th century (GLADNEY, 2004, p.7-8). It is also true that fluency in Mandarin is required in most workplaces. Many ethnic minorities, however, speak their local languages at home.[6] Minority children begin to learn Mandarin when they first enter school at the age of six or seven.[7] Even today many individuals particularly from rural areas have very poor Mandarin language skills; this hinders their ability to find jobs. LANG's language theory of discrimination (1986) postulates that this situation will only change if the utility of learning a new language exceeds the transaction costs of learning the language. This could imply that only if additional Mandarin language skills increase the probability of finding a good job in future will the effort of learning Mandarin be acceptable to ethnic minorities.

Weak Mandarin skills are linked to poor quality of education in remote rural areas, where often even the teachers make mistakes when speaking Mandarin. This impedes ethnic minorities' access to demand-pull NFRE as for Yi men from Liang Mountains in Sichuan, who are unable to find work in Beijing because of their weak Mandarin language skills. Olivia Kraef says that:

> Many of them cannot speak Chinese very well; because the teachers in their schools are Yi and do not correctly know the Chinese language. Only some few can gain a foothold within the cultural sector, for example, as Yi-dancers or as mediators of Yi culture. Many, however, find no work, hang around, or are struggling doing occasional jobs. (own translation from LUO, 2008, p. 3)

The example of a Tujia boy from Jishou also clarifies the significance of Mandarin: "Oh, long ago we became Chinese… [W]ith Tujia language you cannot reach very far. It is delivered only verbal and has no script. I learned it only from my granny and grandpa, yet not very thoroughly. I only know some simple expressions" (own translation from LUO, 2008, p. 1).

---

[6] Broadly the languages spoken in China can be classified into four language families: Sino-Tibetan (e.g., Mandarin, Tibetan, Kam-Tai, Miao-Yao), Turkic-Altaic (e.g., Kazakh, Uyghur, Mongolian, Manchu-Tungus, Korean), Austro-Asiatic (e.g., Hmong, Vietnamese) and Indo-European (e.g., Tajik, Russian) (GLADNEY, 2004, p. 7). Han not only speak Mandarin but have different "linguistic groupings", which include Mandarin, Wu, Yue, Xiang, Hakka, Gan, Southern Min and Northern Min, with additional subgroups among them (GLADNEY, 2004, p. 7). The number of Han who can speak an ethnic minority language might not be larger than the number of Germans speaking Turkish (GERBER, 2011).

[7] The ethnic minorities in remote rural areas often live in compact communities without much Han influence. If the number of Han-Chinese children in a primary school surpasses the number of ethnic minority children, the language of instruction will change to Mandarin instead of the ethnic minority language in order to guarantee equal opportunities (GERBER, 2011). From middle school to university, the major language of instruction is usually Mandarin.

This example also holds true for many other residents of Jishou and other ethnic townships in China. In Jishou 70 % of the population are Tujia, but on the streets, in restaurants and at the train station, one hears only Mandarin (LUO, 2008). In contrast the Hui, who are spread throughout China, do not have their own language. They speak Mandarin or Chinese dialects depending on their location (GLADNEY, 2004, p. 271). The Dongxian, Baoan and Salar, who speak a mixture of Chinese, Turkish and Mongolian, are also instructed in Mandarin, while the Uyghur, Kazakh, Kyrgyz and Tajik are mainly instructed in their mother tongues in primary and secondary schools (GLADNEY, 2004, p. 272).

In addition to problems caused by Mandarin's status as the main language of instruction in China, another issue for some ethnic minorities is the syllabus in state schooling. Muslims are often unhappy about the strong focus on Mandarin and mathematics rather than content related to Islam, such as the Quran, Arabic and Persian (GLADNEY, 2004, p. 278). Men and women, moreover, traditionally do not pray and study together, so that orthodox Muslims refuse to have their children educated in Chinese state schools (GLADNEY, 2004, p. 286). Regarding the comparatively high drop-out rates among Muslim girls, the author emphasizes that "until Chinese educational policy recognizes "cultural levels" that are based on other knowledge traditions and languages, many more conservative Muslims might continue to resist sending their children – especially their daughters – to state schools" (GLADNEY, 2004, p. 281).

Opportunity costs of education

High opportunity costs of education influence schooling and employment outcomes. Children who are attending school cannot help with household chores. This is a negative incentive for parents to send children to school and is particularly observable in poorer families and rural agricultural households in China; ethnic minorities make up a huge part of both poorer families and agricultural households (GUSTAFSSON and SAI, 2008). In Gansu the dropout rate increased with the implementation of the household responsibility system[8], which triggered the necessity for more income-generating labor on the farm (GLADNEY, 2004, p. 278).

Children's involvement in household's chores is also influenced by the availability of schools. If children have to spend the week or, depending on the boarding school, the whole term at school, they cannot help with household chores after school. Households in remote areas, thus, face higher opportunity costs of sending

[8] "Household Responsibility System (HRS) was implemented nationally in China at the end of 1978 to replace the previous communal system. Under HRS, the land in the village is distributed equally in quantity and quality to the households according to family size with land management rights vested in households but land ownership rights remaining in the village. Under HRS, households sign the contracts with the local village; these contracts link various taxes and quotas to the plots of contracted land but allow the households to retain the residual income after fulfilling the quotas and taxes" (WANG, 2007, p. 2).

their children to school than do households closer to townships or cities. This holds true for children, for example, for Kazakh and Kyrgyz herders' children who are left in school during herding season and can only join their parents during holidays (GLADNEY, 2004, p. 274).

Expectations that further education will not provide higher income, moreover, negatively influence how long a child goes to school. Overall returns on education appear lower in rural areas, though DE BRAUW and ROZELLE (2007) and ZHANG et al. (2002) argue that the aggregate rates of returns on education have increased over time. DE BRAUW and ROZELLE (2007) also point out that private returns on education may be very low for some individuals based on the fact that some employers still use non-market factors (e.g., *guanxi*, which means through connections) for assigning jobs, rather than giving the position to the most skilled and qualified worker.

Finally some authors have observed different educational preferences between ethnic groups. For example BHALLA and QUI (2006) suggest that the Miao value education more highly than do other ethnic groups. In contrast GLADNEY (2004, p. 274) finds that Hui parents prefer that their children help with family businesses rather than acquire additional education; this, however, may be related to the syllabus set by the Chinese Ministry of Education and the devaluation of content related to Islam in public schools.

These heterogeneous findings show that there is no clear pattern regarding educational differences of ethnic groups in China. To obtain higher education, Mandarin is fundamental. The availability and quality of schools depends on geographic location. The opportunity costs of education are, furthermore, particularly high for agricultural households in remote regions, where most of the autonomous areas assigned to ethnic minorities are located. Ethnic minorities often face these hurdles which hinder their access to better schooling and employment. With the implementation of preferential policies, the Chinese government seeks to overcome these inequalities.

2.4 Empirical research methodologies

While all theoretical approaches to understanding the phenomenon of ethnic differences in occupational outcomes and wages are useful, investigating them empirically is cumbersome since employer discrimination is forbidden and only limited data are available, particularly in China, where investigations about sensitive topics, for example ethnic minorities, are restricted. The second question, to what extent discrimination is empirically measurable, is, thus, the center of discussion in this subchapter.

I introduce major differences between the quantitative and the qualitative approach to acquaint the reader with their particularities (table 2.2). The two approaches have at least four main differences: "the level of measurement, the number of observation

Table 2-2: Differences of mainstream qualitative and quantitative methodologies

Differences	Qualitative approach	Quantitative approach
Level of measurement	nominal	ordinal and higher
Size of the *N*	small	large
Statistical tests	no	yes
Depth of analysis	thick, detailed knowledge of specific cases	thin, limited knowledge of each case

Source: Author, based on SEAWRIGHT and COLLIER (2004, p. 301-302).

(size of the *N*), the application of statistical tests, and depth of analyses" (SEAWRIGHT and COLLIER, 2004, p. 301-302).

First, the level of measurement differs in that qualitative data have a nominal level and quantitative data have ordinal or higher levels of measurement. The nominal level of measurement classifies data based on numbers, words and letters, while the ordinal level of measurement gives some ordered relationships of the data (STATISTICS SOLUTIONS, 2012). Second, regarding the number of observations, the qualitative approach uses small-*N* research, and the quantitative approach, large-*N* research. SEAWRIGHT and COLLIER (2004, p. 301) suggest that the dividing line lies somewhere between 10 and 20 observations. Third, the quantitative approach uses statistical tests to analyze data. The qualitative approach does not follow strict statistical testing; instead, it uses a narrative approach for analyzing data, which is as important as statistical testing is in the quantitative approach. Fourth, quantitative researchers analyze a large set of observations but receive only limited knowledge of each case; this is referred to as thin analysis. Researchers applying qualitative analysis to get detailed knowledge of specific cases; this is referred to as thick analysis (SEAWRIGHT and COLLIER, 2004, p. 301-302). In subchapters 2.4.1 and 2.4.2 I describe the most important quantitative and qualitative approaches which can be used for analyzing ethnic differences in occupational outcomes and wages. In subchapter 2.4.3 I draw some conclusions regarding their major shortcomings.

2.4.1 Quantitative approaches

Quantitative methodologies have been extensively used for analyzing labor market discrimination and occupational outcomes. The quantitative approaches most widely used in economics and the social sciences to analyze occupational differences are discrete choice models, wage equations and segregation indices, which I discuss in sections 2.4.1.1, 2.4.1.2 and 2.4.1.3, respectively. The computational accuracy of each of these methodologies has improved with time, yet the fundamental question of how to empirically disentangle different preferences and discrimination remains unanswered. To control for these issues I make assumptions about causal relationships; these are laid out in subchapter 2.5.

2.4.1.1 Discrete choice models

Occupational outcome models, one kind of discrete choice model, measure the causal relationships of a set of explanatory variables on a discrete dependent set of occupations. BOSKIN (1974) and SCHMIDT and STRAUSS (1975) are the pioneers of this approach. The major issue in occupational outcome analysis is that it is difficult to disentangle effects from discrimination and preferences when a significant ethnic coefficient is observed. It is necessary to make a-priori assumptions about job preferences to receive straightforward modeling results.

Researchers have to make at least three decisions when applying occupational outcome models: 1) which set of occupations to use as dependent variable, 2) which explanatory variables to use and 3) which particular model setting to apply. All three decisions depend on available data, theoretical considerations and results of statistical testing.

Dependent variable, set of occupations

The dependent variable (left-hand side variable) is a discrete, usually unordered variable which includes the set of available occupations. In the practical application of these models, the categorization of this variable mainly depends on available secondary datasets and usually considers the actual occupations of people, without information on how occupations were chosen or on other occupations considered in this decision-making process.

For example a person living in rural Guizhou may consider a much smaller list of job possibilities than an individual living in the centre of Beijing. BLAU et al. (1956) long ago pointed out the difficulty of answering the question of whether or not individuals really rank occupations based on their preferences and expectations. Available statistical data (revealed preference data) do not reveal whether a farmer chose to be a farmer because it was his/her exclusive choice, whether s/he chose to be a farmer out of a set of available agricultural positions, or whether s/he also considered a completely different set of occupations. Researchers can use choice experiments for analyzing decision-making processes and, therefore, obtain stated preference data, which are more meaningful for analyzing decision-making processes than revealed preference data.

Given the wide range of available occupations, several ways of categorizing different job types exist. Table 2.3 gives some examples of frequently used approaches. In the empirical application the number of observations crucially influences how many job categories can be used. From a computational point of view, the most commonly used approach is to have at least 30 respondents in each considered job category.[9] This means that groups with less than 30 respondents are usually collapsed to form a broader job category which combines similar occupations.

[9] Colleagues referred to this number in one of our departmental meetings at IAMO.

Within the empirical examination suitable categories can, thus, be merged, depending on underlying correlations of categories (cf., chapter four).

After having identified the reduced set of occupations for the econometric analysis, it is necessary to give numerical values to the alternatives. The standard way is to use 0 for the most frequently chosen alternative; the remaining alternatives are then numbered, starting with 1 up to the number of alternatives considered. If researchers, for example, consider a choice set with three alternatives, they will have the numerical values 0, 1 and 2. There is, however, evidence that the estimated results are different when the scope of the numerical values changes; when the three alternatives are numbered, for example, 0, 3 and 6 instead of 0, 1 and 2, different solutions may occur.

Independent variables

Regarding explanatory variables (V_{ij} in the random utility function 2-30), researchers face the challenge of selecting only those variables which are significant without having problems of overfitting (too many insignificant variables), or of omitted variable biases (the absence of significant variables) in the regression.

Table 2-3: Categorization of occupations

Social status JONES and MCMILLAN (2001), LE and MILLER (2001)
Holland's six occupational types LARSON et al. (2002), PORTER and UMBACH (2006), ROSENBLOOM et al. (2008)
Skilled, semi-skilled, unskilled DARDEN (2005)
Good jobs and bad jobs JUNANKAR and MAHUTEAU (2005), MAHUTEAU and JUNANKAR (2008)
White-collar and blue-collar occupations BJERK (2007), HAM, et al.(2009)
Menial, blue-collar, craft, white-collar and professional SCHMIDT and STRAUSS (1975)
Agriculture; professional and managerial; clerical, sales, service; manufacturing and transportation HANNUM and XIE (1998)
Labor market participation states in agriculture: labor services off-farm "selling", on-farm labor "hiring", simultaneously "selling" and "hiring", do not participate on either side "autarky" BROSIG et al. (2007)
Non-state sector, state-sector, redistributive agencies ZANG (2008)

Source: Author, extension of HAM, et al. (2009).

Based on the theoretical discussion in the previous subchapter, I find that explanatory variables must control for differences in preferences, abilities, employer discrimination and institutional discrimination. In the literature researchers commonly use 1) racial, ethnic or national status, 2) gender, 3) education and 4) age as the main independent variables for analyzing different occupational outcomes. Additional variables are included depending on the research focus and on statistical significance. Interaction terms among the variables or square values of the variables are, moreover, included to capture nonlinear effects.

For example SCHMIDT and STRAUSS (1975) use race, gender, educational attainment (school years) and labor market experience (age minus years of schooling minus five). HANNUM and XIE (1998) use nationality, gender, age, education (illiterate, junior high school and senior high school) and residence status. ZANG (2008) uses age, male, marital status, native, CCP membership, education (illiterate and semiliterate, primary school, junior high, senior high), era of labor force entry (1949-59, 1960-79, 1980-89), father CCP, father state worker, father professional and Hui status. These factors are, however, not always the sole cause of differences in occupational outcomes; there are many interlinked causes seldom observed in secondary datasets (figure 2-5).

Model setting

The theoretical approach of choice models as given in subchapter 2.2.3 suggests that individuals choose the job which brings them the greatest utility. Researchers, however, are uncertain about which job alternative brings the greatest utility to individuals; thus, researchers compute probabilities. The decision about which discrete choice model one can apply depends on whether or not the available data satisfy the assumptions made on distributions and on variations of the error component of the underlying model specifications (cf., TRAIN, 2009; GREENE, 2008; HENSHER et al., 2005).

The multinomial logit model (MNL) is the "workhorse" for choice analysis (HENSHER et al., 2005). While binary models compare two alternatives (e.g., agricultural or non-agricultural occupations), the MNL model can be used to compare more than two unordered alternatives (e.g., agricultural, blue-collar, white-collar or professional occupations). One major assumption of the MNL is the independence from irrelevant alternatives (IIA) assumption as a result of independent and identically distributed (IID) error terms. It implies that job alternatives are uncorrelated among each other; for example, under this assumption the decision to work in agriculture does not depend on whether only one or several alternative occupations are available; this is very unrealistic in a real world scenario. MNL models are estimated by maximizing likelihood functions. The validity of the IIA assumption can be tested with Hausman tests (HAUSMAN and MCFADDEN, 1984). The IID/IIA assumption is only justified if all correlations can be explained in the modeled part of the utility function (cf., equation 2-30).

If the IIA assumption does not hold, models such as nested logit, cross-nested logit, mixed logit or multinomial probit models can be used, as these models relax the underlying IID/IIA assumption. This means that the covariance structure of the underlying error term matrix allows for different forms of correlations among alternatives. In addition to demanding case-specific variables, these models also require alternative-specific variables in the empirical analysis. Alternative-specific variables are specific to the outcome categories (e.g., the wage rate for each occupation), while case-specific variables are specific to the individual, for example, human capital factors (e.g., ethnic status, education, gender and age).

In nested or cross-nested models some alternatives become closer substitutes for each other (e.g., white-collar and professional positions and/or agriculture and blue-collar positions) and are grouped into nests. The IIA assumption holds within nests, but not across nests. The error component has an extreme value distribution. Nested models are often interchangeably named generalized extreme value (GEV) models.

Mixed logit models and multinomial probit models (MNP) account for random representation of taste heterogeneity. The error terms are assumed to have multivariate normal distributions that are heteroskedastic and correlated. The models allow for a general covariance structure of the error term, so that the choice for one job alternative depends on all remaining alternatives. The probabilities from the multivariate normal distribution are evaluated using simulation techniques because of the absence of a closed-form solution; this comes with high estimation costs in comparison to the MNL or nested models, which have closed-form solutions.

2.4.1.2 Wage equations

When wages are used as a dependent (left-hand side) variable, this variable is usually continuous rather than discrete. Wage differences between groups are then usually, as in the discrete models (equation 2-30), analyzed by dividing the right-hand side of the equation in an explained and an unexplained part. The wage difference is calculated first by estimating for each group a wage equation, then by taking the difference of these two equations (ALTONJI and BLANK, 1999, p. 3153).

The technique of decomposing inter-group differences in wage equations is attributed to BLINDER (1973) and OAXACA (1973). Following ALTONJI and BLANK (1999, p. 3153-3154), the wage equations for individual *i* in group *1* and individual *j* in group *2* at time *t* are

$$W_{1it} = ß_{1t} X_{1it} + \mu_{1it} \tag{2-36}$$

$$W_{2jt} = ß_{2t} X_{2jt} + \mu_{2jt} \tag{2-37}$$

where $ß$ is the estimated coefficient, X the explained part and μ the unexplained part in both equations. The equations underlie the assumptions that the expected

values of the explained and unexplained parts are independent from each other, $E(\mu_{1it} \mid X_{1it}) = 0$ and $E(\mu_{2jt} \mid X_{2jt}) = 0$, respectively.

The differences in the mean wages can then be estimated in two major ways, as shown by ALTONJI and BLANK (1999, p. 3156):

$$W_{1t} - W_{2t} = (X_{1t} - X_{2t})\beta_{1t} + (\beta_{1t} - \beta_{2t})X_{2t} \tag{2-38}$$

$$W_{1t} - W_{2t} = (X_{1t} - X_{2t})\beta_{2t} + (\beta_{1t} - \beta_{2t})X_{1t} \tag{2-39}$$

The authors, however, emphasize that the two approaches lead to varying results, so that researchers frequently report both and/or use the average of the two results.

Coming back to the right-hand side of the equations, the first part shows the difference in the explained component ($X_{1t} - X_{2t}$), and the second part, the difference in the unexplained component ($\beta_{1t} - \beta_{2t}$) between the two groups. The explained part shows the average differences in the variables used, such as education or age, while the unexplained part shows the differences in the estimated coefficients and, thus, measures differing returns between the two groups given the same characteristics. The differences in wages, which result from the second component of the equations, are considered to be the "share due to discrimination" (ALTONJI and BLANK, 1999, p. 3156).

The authors, however, point out that it is misleading to say that the unexplained part of the regression explains discrimination, as any variable omitted from the explained part will also affect the coefficients. In addition to the effects of discrimination, the unexplained part also includes "unobserved group differences in productivity and tastes" (ALTONJI and BLANK, 1999, p. 3156). At the same time pre-labor market discrimination can also cause differences in group characteristics, i.e., the explained part of the regression. If, for example, one group suffers from pre-labor market discrimination, these circumstances are reflected in the given characteristics, which in turn affect the returns from labor.

Many scholars have discussed the shortcomings of the Blinder-Oaxaca decomposition technique or have extended it.[10] For example some important contributions are made by OAXACA and RANSOM (1999), who examine the effects of omitted variables in decomposing wage equations. FAIRLIE (1999, 2006) extends the Blinder-Oaxaca decomposition with a decomposition that can be used to analyze non-linearity, such as estimates from logit or probit models. JONES (1983) and CAIN (1986) further discuss the challenge of interpreting the unexplained part of decomposed wage equations. Drawing reliable conclusions about labor market

[10] BLINDER (1973) had 2,324 and OAXACA (1973) 3,886 registered citations in the google-scholar search engine on July 11, 2011, which implies the extensive use of their approach in the literature.

discrimination with wage equations is very difficult because of the complexity of disentangling discrimination effects.

2.4.1.3 Numerical indices of employment segregation

To analyze secondary datasets based on differences in job distributions, some researchers calculate segregation indices. Depending on the research focus, segregation indices are used to numerically measure employment segregation between genders and among ethnic groups and races. The basic idea of the indices is to measure whether or not members of the groups under consideration are equally distributed among all available occupations depending on their total shares in the work force. For example JOHNSON and STAFFORD (1998, p. 77), who study the tendency for men and women to perform differently in the labor market, use $D = P_m - P_f$ for measuring the gender proportion in an occupation. *D* stands for the index of occupational dissimilarity, P_m and P_f for the proportion of men and women, respectively. *D* equals 0 when the employment shares of *m* and *f* are identical. *D* increases when the share of *m* increases or when the share of *f* decreases, *ceteris paribus*.

An equal distribution of occupations across groups is defined as integration (HUTCHENS, 2004); an unequal distribution of occupations across groups is defined as segregation, exclusion or dissimilarity (JOHNSON and STAFFORD, 1998, p. 72). CHAKRAVARTY and SILBER (2007), HUTCHENS (1991, 2001, 2004), FLÜCKIGER and SILBER (1999), DUNCAN and DUNCAN (1955), among others, elaborated segregation indices. Researchers basically use segregation indices for analyzing available secondary data on occupational distributions, yet the actual sources of the observed segregation or the driving forces behind job decisions are not considered with the indices. As in occupational outcome models, only the final job decisions are considered in segregation indices.

In addition to the measurement challenges, there is a social debate about whether or not a total integration, which implies an equal job distribution among ethnic groups, is required by society. Is it really necessary to have equal job distributions between ethnic groups if the occupational choices of individuals exactly represent their job preferences?

2.4.2 Qualitative approaches

I highlighted many challenges in the application of quantitative methodologies for measuring ethnic differences in occupational outcomes and wages in the last subchapter. When confronted with the limitations of quantitative methodologies, I drew attention to qualitative approaches as a suitable way for gathering additional information about underlying sources for ethnic differences in occupational outcomes and wages. Qualitative methodologies, however, can also only be partly applied; they can even be inappropriate for identifying the sources of different occupational outcomes.

In general qualitative methodologies face at least four potential shortcomings: selection bias, inability to capture evolution over time, inability to generalize research findings and researcher bias. Selection bias and consequently biased results based on truncated samples are major shortcomings of qualitative research (e.g., self selection of individuals into the experiment, the so called Hawthorne-effect (LIST and RASUL, 2011, p. 129)), which also holds true in interviews.

For investigations in China, researchers should be aware that research areas and interviewees are pre-selected by the local partners. Crucial issues regarding social inequalities, therefore, may remain undiscovered as not all areas and individuals are accessible, particularly in TAR and XUAR. With qualitative methodologies it is also difficult to capture evolution over time. Qualitative research projects are often at a single point in time and are not repeated over a longer time period, which is related to project length and funding limitations.

These two issues, selection bias and the inability to capture evolution over time, lead to shortcomings in the generalization of findings. Results usually depend on the selected area at the particular time of the field work. This means that conclusions for the whole country, often required by mainstream economists, can in many cases not be drawn and cross-country comparisons are, therefore, difficult to make. Another problem of qualitative methodologies is the intervention of the researcher. Researchers have their own beliefs and ideologies which they intentionally or unintentionally can use in interviews, field experiments or by observing the field to conduct their preferred results.

I give a brief overview of some qualitative methodologies in the economic and social sciences, which are frequently used to tackle group differences in diverse circumstances: audit studies, interviews and participant observation. I particularly draw attention to the difficulty of correctly measuring employer discrimination as in the quantitative part.

2.4.2.1 Audit studies

The main shortcomings of quantitative methodologies, that not all factors which can potentially cause discrimination can be disentangled and controlled for, can partly be overcome by using audit studies for measuring hiring decisions (cf., ALTONJI and BLANK, 1999, p. 3192-3194). Audit studies fall in the range of quasi-experimental methodologies (QUILLIAN, 2006, p. 303) and serve to shed light on employers' hiring decisions.

There are two kinds of audit studies. In the first approach researchers send out identical job applications by changing only the applicant's name and, depending on the research focus, the ethnic status, race or gender of the applicant. Researchers then measure whether or not there are differences in the acceptance or denial probabilities between the groups considered. In the second approach researchers not only send out applications as in the first approach, but also directly send out individuals with different ethnic status, race or gender to interviews with potential

employers. The second approach, therefore, requires the researcher to know the exact requirements for the open job position, to search for suitable probands and to train them accordingly. ALTONJI and BLANK (1999, p. 3192) suggest that the treatment of the applicant by the employer before (e.g., waiting time), during and after the interview, moreover, can be compared between the groups of interest.

With audit studies researchers identify discrimination if the probabilities of getting an interview in the first approach or of receiving a job offer in the second approach differ significantly between the groups considered. As audit studies control all factors and only vary in ethnic status and name, researchers are able to measure employer discrimination against ethnic minorities more accurately than with other approaches. Audit studies, however, are also limited. HECKMAN and SIEGELMAN (1993) criticize two weaknesses. First, the authors argue that there is not a double blind approach in audit studies. The probands know about the measurement goal of the study and may adapt their behavior, so that they achieve the desired outcome of the study. Second, ethnic, race or gender effects are overrepresented in audit studies. All potential factors aside from ethnic status, race or gender are held constant with the methodology, which is a very artificial approach to measure real labor market outcomes.

To be absolutely accurate, a job rejection can also occur because of "name discrimination" without any relation to the ethnic status, race or gender of the applicant. This effect cannot be disentangled with audit studies either. In the second approach, moreover, researchers can only control their probands, but fail to control real applicants in the analysis as their application profile is unknown to them.

2.4.2.2 Interviews

Sociologists usually apply long, structured interviews in their studies of social inequalities. This might be an appropriate way to get information about decision-making processes for identifying the roots for different occupational outcomes, yet interviewees may not tell the truth when asked about subjects such as employer discrimination which are taboo or even illegal.

In China job discrimination by ethnicity, race, gender or religious affiliation is forbidden by law (cf., ROSS et al., 2007). Researchers who interview members of the advantaged group on their hiring practices and/or members of the disadvantaged groups on their experiences looking for work can obtain inaccurate and misleading results. The advantaged group may not admit that they discriminate against the disadvantaged group as it is a criminal act, while the disadvantaged group may not be aware that they are discriminated against. QUILLIAN (2006), who analyzes discrimination in the US, even states that "the strong normative prohibition against discrimination" leads the advantaged group to downplay their actual discriminatory behavior, even when discrimination is not punishable by law (QUILLIAN, 2006, p. 303). The disadvantaged group, however, has only a restricted view about their

situation in the labor market as the advantaged group will not openly inform them of their discriminatory behavior. The disadvantaged group, thus, has only a perception about their actual situation, which can either be understated or overstated depending on previous experiences (QUILLIAN, 2006, p. 303).

To secure information accuracy, interview results and statistical figures can be combined. Unfortunately statistical figures on labor market discrimination are generally not available. As the researcher has an active part in the inquiries, researchers might, moreover, unintentionally or not, influence results.

2.4.2.3 Participant observation

Researchers of many disciplines apply participant observation to acquire supporting information from field studies and/or to get a first impression of the research field in order to develop suitable hypotheses. Participant observation requires that researchers get progressively involved with the field environment and people and that researchers make observations based on a broader consideration of the field to more concrete attention to the research questions (FLICK, 1995).

The observation process is divided into descriptive, focused and selective phases (SCHMIDT-LAUBER, 2007; SPRADLEY, 1980). In the descriptive phase researchers get an orientation to the field and obtain unspecific descriptions. The goal of the descriptive phase is to capture the complexity of the field and to develop more concrete hypotheses. Researchers further seek to find out whether or not the chosen research field serves to answer their stated research questions. In the focused phase researchers pay particular attention to specific problems, processes and persons involved in order to answer the initial research question. The selective phase, finally, serves to find additional records and examples of the identified patterns and forms of behavior of the local actors. The following list of concerns can be used to systematically and comprehensively capture a particular research field at a specific time:

> space: the physical place or places; actor: the people involved; activity: a set of related acts people do; object: the physical things that are present; act: single actions that people do; event: a set of related activities that people carry out; time: the sequencing that takes place over time; goal: the things people are trying to accomplish; feeling: the emotions felt and expressed. (WOLFINGER, 2002, p. 91 based on SPRADLEY, 1980, p. 78)

Results obtained with participant observation are normative in nature. As with interviews there is always the possibility that researchers, unintentionally or not, influence research outcomes based on their own beliefs and ideologies.

2.4.3 Interim conclusions

The empirical investigation of ethnic differences in occupational outcomes and wages is a challenging task as employer and institutional discrimination is forbidden and data are limited. In China investigations about taboo subjects, such as discrimination against ethnic minorities are, moreover, restricted, especially for

foreign scholars. The second question, to what extent is discrimination empirically measurable, is, thus, difficult to answer. I find that quantitative and qualitative methodologies both have shortcomings, particularly when it comes to disentangling the effects of preferences, abilities and employer discrimination. The right set of relevant assumptions is, therefore, the key for obtaining accurate results.

The application of occupational outcome models or wage equations to analyze ethnic differences in occupational outcomes and wages is often affected by omitted variable biases (the absence of significant variables), over-fitting (too many insignificant variables) or endogeneity problems (unclear causal relationships). The application of segregation indices is, moreover, problematic as it is questionable whether an equal job distribution of all ethnic groups should really be achieved, given that an unequal job distribution can also reflect different individual job preferences rather than employer discrimination. Qualitative methodologies face at least four potential shortcomings, which are selection bias, inability to capture evolution over time, inability to generalize research findings and researcher bias, which can influence results.

The empirical analysis of labor market discrimination, thus, requires well-defined assumptions to clarify relationships among the variables. Without assumptions investigation results can contradict each other, for example, when a significant negative coefficient of the ethnic status is observed in occupational outcome models (cf., SCHMIDT and STRAUSS, 1975). Making the right assumptions on variables in occupational outcome models requires an in-depth analysis of job preferences and related human capital factors of the ethnic groups and individuals considered with a thorough literature review and qualitative investigations. In line with PETRICK (2004), I, therefore, suggest that combining methodologies increases the plausibility of results.

2.5 Linking theoretical and empirical approaches

To analyze the content-based research question, I combine quantitative and qualitative methodologies to get more reliable results. This requires a research strategy which includes *theoretical foreknowledge* and, thus, controls and explains relationships. The research strategy must also accommodate the use of empirical results to adjust *testable hypotheses throughout the research process*. This approach is known as "circular theorizing" (KORF, 2004, p. 31, cited in THEESFELD, 2005, p. 97). This means that the theoretical framework steers the empirical application, then the empirical results adjust the theoretical framework.

In my investigation I use the diamond of theoretical principles (see figure 2-4) as theoretical foreknowledge to explain ethnic differences in occupations. I adjust the theoretical foreknowledge to my case study based on relevant information collected during field work and on additional literature from later stages of the research process. I explain how I develop suitable assumptions and hypotheses

based on this research strategy in subchapter 2.5.1. In subchapter 2.5.2 I then explain how I empirically investigate the derived hypotheses using mixed methodologies before I present the qualitative and quantitative approach and then the results of my study in chapters three and four, respectively.

2.5.1 Assumptions and hypotheses

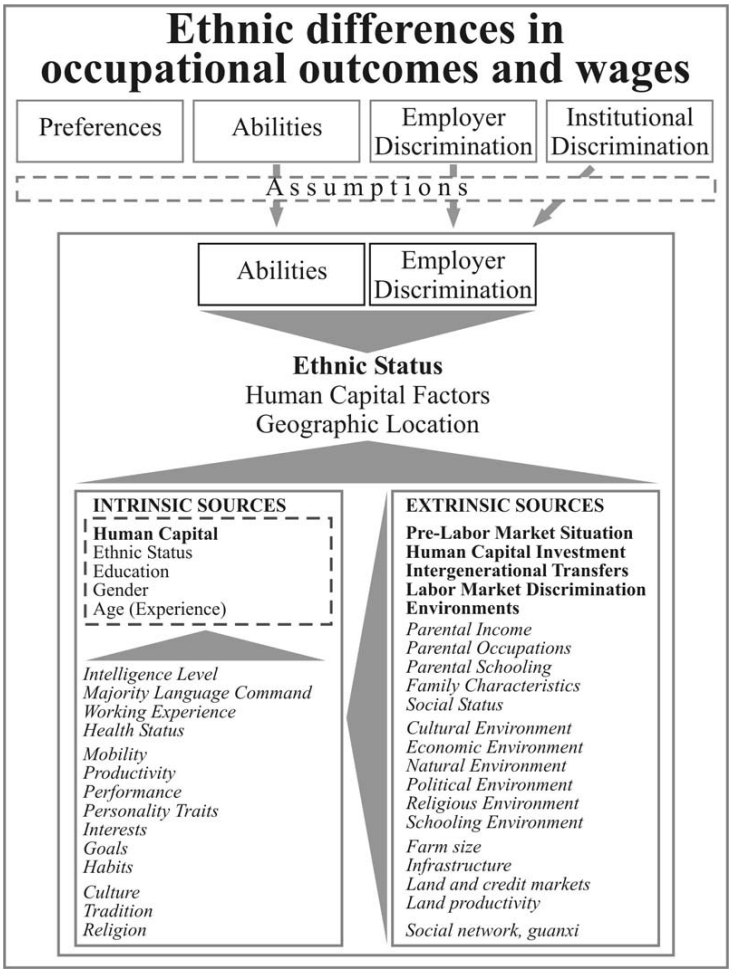
In the theoretical discussion I concluded that JOHNSON and STAFFORD's occupational exclusion model (1998) actually serves to organize available theoretical concepts for measuring ethnic differences in occupational outcomes and wages (cf., figure 2-4). I highlighted that the four major causes for ethnic differences in occupational outcomes and wages are differences in preferences, in abilities and in both employer and institutional discrimination. In the discussion of empirical methodologies, I presented major quantitative and qualitative research tools which can be applied for analyzing whether or not ethnic minorities in Guizhou are discriminated against in the rural labor market. I seek to connect the theoretical and empirical concepts by developing the necessary set of assumptions and deriving testable hypotheses.

2.5.1.1 Assumptions included in the theoretical concept

Figure 2-5 serves as a guide linking the four theoretical concepts and related variables. The top of figure 2-5 presents the four theoretical principles. To disentangle these four effects, I make two major assumptions about job preferences and about employer and institutional discrimination. These assumptions are required because, otherwise, results have no accurate interpretation (cf., subchapter 2.4). Regarding job preferences I have repeatedly stressed that a significantly negative ethnic coefficient on a discrete set of occupations can either indicate ethnic differences in discrimination or in preferences, which are completely opposite outcomes and insufficient either for drawing reliable conclusions or for guiding policy (cf., SCHMIDT and STRAUSS, 1975).

For the analysis of the Guizhou case, I first assume that in a free choice scenario without constraints, all individuals regardless of ethnic affiliation prefer non-agricultural (*NA*) over agricultural (*A*) positions. I make this assumption based on the fact that average workers in Guizhou consider *A* inferior work, which reinforces higher individual preferences for *NA* positions. The first major objection to work in *A* is related to the low status of peasants in Guizhou. A second major objection to *A* is that long exposure to strong sun during fieldwork darkens the skin; in China skin darkened by the sun is generally seen as an inferior characteristic of farmers, while being white and tall is the beauty norm. SCHEIN (2000, p. 241), for example, points out that on the marriage market in Miao communities, good characteristics of a potential partner are a "fleshy body" and a "clean and fair skin", while inferior characteristics are a skinny body and dark skin, related to hard physical labor in the sun and to poverty.

Figure 2-5: Conceptual framework



Source: Author based on own field work and JOHNSON and STAFFORD (1998); ALTONJI and BLANK (1999); HAM et al. (2009); BUCHENRIEDER and MÖLLERS (2006); GOTTFREDSON (1981); ARROW (1973).

SCHEIN (2000) additionally provides a personal example about the importance of a fair skin for a Miao woman in Guizhou:

> This standard of beauty was made poignantly clear to me when my spouse came to visit from the United States during the summer months. For weeks, I was troubled that a close friend – a nineteen-year-old woman who had visited me regularly earlier in the year – never came to visit. After what seemed an incomprehensible several weeks of absence, she came abashedly to our door, apologizing that she had stayed away so long. She confided

> to me that she had been ashamed (*bu hao yisi*) to be seen by my spouse because her hard outdoor work (*ganhou*) had blackened and emaciated her. (SCHEIN, 2000, p. 241)

The preference for white skin is also obvious when you want to buy a simple facial cream in an average store in China; it is actually quite impossible to find a facial cream without whitening formula. SCHEIN, based on POTTER (1983), concludes that even in remotest areas in China, urban workers are seen as superior over peasantry and mental tasks over manual labor. SCHEIN finds that many women prefer a potential partner with a regular salary over a simple peasant. During my stay at Guizhou University in winter term 2010, a researcher familiar with the topic even informed me that in Dong communities in southeastern Guizhou, no one would actually freely choose to be a farmer if other occupations were available. Considering all the aforementioned facts, I, thus, assume that on average *NA* positions are preferred over *A* positions. It could, however, also be argued that some manual jobs in the *NA* sector require similar, hard work in harsh conditions and, therefore, may also be closely linked to lower social status. Based on conversations with locals and researchers familiar with the topic as well as on the literature, no direct evidence is, however, available to support this statement.

Second, I make assumptions regarding the theoretical concepts of employer and institutional discrimination. I assume that positive institutional discrimination towards ethnic minority workers (status *E*) as is found in Guizhou facilitates the access of status *E* workers to the *NA* sector. Preferential policies directly influence the lives of ethnic minorities in several private (e.g., family planning) and corporate sectors (e.g., school and university admission, reduction of taxes) (cf., subchapter 1.3; BROWN, 2001, p. 57; SAUTMAN, 1997, p. 3). A Han employee of Guizhou University, for example, informed me that for ethnic minorities with PhD degrees, it is fairly easy to get better employment in the university, while it is much harder for Han with PhD degrees to get good positions. Under the assumptions that the positive institutional framework is well enforced and equally enforced for all ethnic minorities considered, I exclude institutional discrimination from the empirical investigation of my study case. This means that all measureable discrimination against ethnic minorities results from employer discrimination and not from the institutional framework.

I made two assumptions to narrow down the theoretical concepts of my research. First, all individuals regardless of their ethnic affiliations prefer to work in the *NA* sector. Second, there is only employer discrimination and no institutional discrimination; therefore, I reduce the theoretical concepts for explaining ethnic differences in occupational outcomes to differences in abilities and employer discrimination. These two effects are, thus, the core links between the theoretical framework and the empirical application (cf., figure 2-5).

2.5.1.2 Derivation of testable hypotheses

The next step is to empirically measure ethnic differences in abilities and in employer discrimination. Based on the theoretical discussion I postulate that ethnic minorities face two constraints for accessing employment in the *NA* sector: 1) individual abilities are not adequate to perform *NA* work, and 2) employers in the *NA* sector discriminate against status *E* workers. I also find that individuals' geographic locations influence their job outcomes. For example MOHAPATRA (2004) finds that agriculture is the main occupation in remote regions and in the lowest-income regions, that there are more available occupations in more developed regions, that agriculture is substituted by migration for working in non-agricultural employment in the slightly more developed and lower-middle income areas, that micro enterprises are established in the higher middle-income and more developed regions and that large-scale manufacturing dominates the richest and most developed villages near city centers. If job migration is, however, not possible, *NA* work can only be chosen if *NA* work is available in surrounding areas. Individuals are, thus, constrained in their access to *NA* employment if 3) *NA* work is not available in the area.

Considering these three constraints, I focus the empirical application on the following independent factors: 1) individual abilities, 2) ethnic status and 3) geographic location. The lower part of figure 2-5 depicts several variables, which can directly or indirectly be linked to these three factors. I divide sources for occupational differences into intrinsic and extrinsic sources, which are interrelated with each other (lower part of figure 2-5). While intrinsic sources are based on important and basic characteristics of a person, extrinsic sources come from outside and are not directly linked to the person's characteristics. Each of these single factors, however, has underlying theoretical approaches and disciplines, yet because of data constraints and complexity, I cannot cover all of the factors in this monograph.

In line with previous studies (cf., SCHMIDT and STRAUSS, 1975; HANNUM and XIE, 1998; ZANG, 2008), I focus the empirical analysis of different occupational outcomes on human capital factors, including ethnic status, education, gender and age; I also consider geographic location as depicted in the center of figure 2-5. I now derive the major hypotheses related to human capital factors and geographic location.

Human capital

Common factors to cover human capital are 1) ethnic status, 2) gender, 3) education and 4) age. I derive testable hypotheses for these four factors:

Ethnic status

If ethnic minority workers (*E*) receive negative discrimination in the *NA* sector, then depending on the degree of employer discrimination, their employment

share in the *NA* sector is very low or absent, and consequently status *E* workers have a higher share in sector *A*, *ceteris paribus*. Sources for negative discrimination against status *E* workers by Han (*H*) employers can result from prejudices against status *E* workers (cf., BECKER, 1957, 1971). The prejudices are often not directly against the ethnic affiliation, but rather against an agglomeration of inadequate factors, which status *E* workers have accumulated.

Depending on the occupation within the *NA* sector, Han employers can have negative beliefs about the *wenhua* (文化) levels of ethnic minorities. SCHEIN (2000, p. 174) points out that *wenhua* is related to general education and particularly Mandarin language skills, while in Chinese language classes and in dictionaries[11], *wenhua* is usually referred to as culture, civilization or education, schooling, literacy. BRAUN (2005, p. 112) translates "wenhua shuiping di" as "low level of education".

This initial negative belief about inadequate *wenhua* levels of status *E* workers has become a stereotype in Guizhou, which causes statistical discrimination against ethnic minorities (cf., subchapter 2.2.2.3). Given this situation, even ethnic minorities themselves perceive their own *wenhua* as inferior to that of the Han people. A researcher familiar with the topic informed me that the perception of inferiority of ethnic minorities is often expressed as 我们没有文化, (women meiyou wenhua = we have no "wenhua")[12]. This implies that in order to find employment in many branches of the *NA* sector and to fit into the mainstream language and cultural system, ethnic minorities in Guizhou tend to deny their ethnic identity.

SCHEIN (2000, p. 192) stresses that after decollectivization, modernization became a huge concern of Miao people, to the extent that they would even give up ancient customs if those are not profitable. If status *E* workers are, however, unable to fulfill the *wenhua* level demanded by Han and, therefore, lose optimism about better job prospects and wages, status *E* workers may also lose their incentive to work harder and/or to invest in human capital. This puts status *E* workers in a vicious cycle and further reduces their chances of accessing the *NA* sector. In this case the initial negative belief about a lower *wenhua* level of status *E* workers is self-confirming (cf., COATE and LOURY, 1993).

This leads to the first testable hypothesis:

> H1: Being an ethnic minority negatively influences access to non-agricultural employment.

---

[11] HTTP://BAIKE.BAIDU.COM/VIEW/3537.HTM, HTTP://WWW.NCIKU.COM/SEARCH/ZH/DETAIL/文化/1315983

[12] It is striking to notice that an older Hui in Hezhou gave exactly the same answer that he "had no culture" after he was asked about his cultural level (GLADNEY, 2004, p. 262).

During field observations I found that in Guizhou levels of sinicization[13] differ by ethnic minority groups; therefore, I focus not only on the more general distinction between status *H* and status *E* workers in occupational outcomes, but also analyze differences in occupational outcomes of the Bouyei, Miao and Tujia groups separately.

I also have to consider that results may differ within the *NA*-sector. For example there are preferential policies which assign ethnic minorities to state employment in their designated autonomous areas (cf., subchapter 1.3). Ethnic minorities, moreover, have a comparative advantage in the tourism industry, which is mainly devoted to traditional ethnic lifestyles (cf., GUSTAFSSON and LI, 2003; BHALLA and QUI, 2006).

The importance of *wenhua* makes clear that education is crucial for accessing many jobs in the *NA* sector, indicating that the combination of ethnic minority status and educational attainment is relevant in the analysis. I first make general hypotheses about education and then consider interactions of ethnic status and education.

Education

With the assumption that the education necessary for accessing *NA* is higher than for accessing *A*, *ceteris paribus*, I deduce that regardless of their job preferences only those individuals with sufficient educational attainment can work in *NA*.

This leads to the second testable hypothesis:

> H2: More years of education positively influence access to non-agricultural employment.

Given the impediments to education laid out in 2.3.2.2, I assume that ethnic minorities in rural Guizhou have more limited opportunities to attain higher education than do Han. This can at least have five major interrelated reasons. First, on average ethnic minorities have lower Mandarin language skills. A researcher familiar with the topic informed me that many ethnic minorities have the first contact with Mandarin in primary school as ethnic minorities often live in compact communities without much Han influence. Ethnic minorities, therefore, start learning Mandarin as a second language in school, which reduces their overall performance in other courses as well. 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 Mandarin instead of their mother tongue and adapting to the unfamiliar Han culture can lead to culture shock for ethnic minorities, which may in turn increase their school drop-out rate (FN, 2010, p. 35). This also negatively affects access to senior high school or university, to which

[13] To sinizice means to modify by Chinese influence (HTTP://WWW.MERRIAM-WEBSTER.COM/DICTIONARY/SINICIZE).

access is restricted by an entrance examination. If ethnic minorities expect to fail in the entrance examination, this could lead to a higher drop-out rate for ethnic minorities beginning as early as primary school, regardless of the fact that ethnic minorities benefit from preferential policies. The lower schooling levels of ethnic minorities in return restrict their access to many branches of the *NA* sector and result in higher employment probabilities in sector *A* for ethnic minorities.

Second, there are fewer schools in remoter ethnic minority villages, and these schools offer lower quality education than those schools found in more developed areas. School availability and quality in China depend primarily on the individual's place of residence. Only primary schools (grades 1 to 6) are located in villages, while schools for obtaining grades 7 to 9 are only available in nearby small towns and schools for obtaining grades 10 to 12 are found in the closest city (KAI MING, 2003). The distance to secondary schools can be very great for rural children, often ethnic minorities. The establishment of boarding schools improves this situation, but studying and living conditions in boarding schools are often poor (SAUTMAN, 1997).

Third, high opportunity costs of education, which involve the costs when children continue schooling rather than help with household chores, represent a negative incentive for parents to send children to school. This is particularly the case for poorer families and households involved in agriculture, who would benefit from the help of children at home, in the field or in a paid off-farm position (BASU and TZANNATOS, 2003). Rural people and members of ethnic minorities make up a huge part of poorer and/or agricultural households (GUSTAFSSON and SAI, 2008). The involvement of children in the household's chores is also closely linked to the availability of schools. If children have to spend the week or, depending on the boarding school, the whole term at a boarding school, they cannot contribute to the household's chores after school. Households in remote areas, thus, face higher opportunity costs of sending their children to school than do households closer to townships.

Fourth, a low expectation of future income from higher education also negatively influences the duration of schooling. Overall returns on education appear lower in rural areas, even if DE BRAUW and ROZELLE (2007) and ZHANG et al. (2002) argue that the aggregate rates of returns on education have increased over time in these areas. If status *E* workers are, however, unable to fulfill the demanded *wenhua* level of the majority and lose optimism about better job prospects and wages, status *E* workers can lose their motivation to work harder and/or to invest in human capital, which puts them in a vicious cycle and further reduces their chances of finding a job in the *NA* sector. Parents of ethnic minority children, thus, consider an additional year of education for a child to be more a financial burden than a present investment in higher future earnings.

Fifth, parents' schooling, occupations and income influence investment in human capital of children, which is crucial for developing individual abilities and for

employment chances (ALTONJI and BLANK, 1999, p. 3167). Particularly for poorer households, among which ethnic minorities in Southern and Western China are often found, school fees are a barrier for going to school (GUSTAFFSON and SAI, 2006, 2008).

Given the above-mentioned impediments, I assume that ethnic minorities in rural Guizhou have more limited opportunities to attain higher education than do Han.

This leads to the third testable hypothesis:

> H3: The lower educational achievement of ethnic minorities negatively influences their access to non-agricultural employment.

Educational differences between the Bouyei, Miao and Tujia groups must also be considered. The Bouyei may have higher educational levels than the Miao based on their higher degree of sinicization in some areas of Guizhou.

Gender

Although in China women are said to hold up "half the sky" (半边天 – ban-biantian) (GLADNEY, 2004, p. 71), in the modern industrialized nation women frequently face constraints in employment in "better" jobs and higher wages. The reasons in China are often established traditions and their corresponding gender roles, as well as the few available positions in the rural *NA* sector (MENG and MILLER, 1995).

MENG and MILLER (1995) investigate occupational segregation and its impact on wage discrimination against females in China's rural industrial sector. The authors find that wage discrimination is greater than occupational segregation between men and women in China's rural industrial sector. They explain this phenomenon as a result of Chinese traditions and fewer available better paid positions in the rural industrial sector. Similar observations have been made in Miao communities in Guizhou; "femininity stood unquestioned as the inferior rank in a vertical social ordering" (SCHEIN, 2000, p. 174).

ALTONJI and BLANK (1999, p. 3166), moreover, point out that men and women have different comparative advantages in competitive labor markets; for biological reasons women have a comparative advantage in home production. Home time is defined as

> time which is not spent in directly productive and labor market activities. It includes family maintenance (cooking, fetching wood and water, tending the house); family reproduction (pregnancies, rearing the children, attending the elders); socialization (relationships within the family and with neighbours and the community, festivals, religious practices); and leisure (relaxation, pleasure, and sleep). (SADOULET and DE JANVRY, 1995, pp. 143-144)

For Miao groups in Guizhou, traditionally "young men might go to towns and cities to seek labor opportunities, while women would invariably stay home to

till the fields, possibly joining their husbands for a sojourn during slack season" (SCHEIN, 2000, p. 174). A similar line of argument has been formulated by BELLÉR-HANN about villages in southern XUAR, where women mainly take care of the house and children, while men generate income (BELLÉR-HANN, 1998a, p. 703). BELLÉR-HANN (1998, p. 9; 1998a, p. 708) even finds in her interviews that agriculture "is not an income-generating activity" for the villagers. Although women are said to take care of the house and children, women are actually engaged in many kinds of work. The difference between men's and women's work is that women's tasks are considered by both men and women only supportive to the work done by men (BELLÉR-HANN, 1998a, p. 709).

I, thus, assume that because of the restricted number of occupations in the *NA* sector, as well as traditional and biological reasons, women devote more hours to home time in rural areas in China,[14] which reduces their available working time. As implied by SCHEIN (2000, p. 174), some household chores also include agricultural production at home, which increases women's distribution in sector *A*. Male workers in contrast have higher chances to work in the *NA* sector for traditional and biological reasons.

This leads to the fourth testable hypothesis:

> H4: Being female (male) negatively (positively) influences access to non-agricultural employment.

Age

Age is also an important determinant in occupational outcome analysis. With information about an individual's age, it is possible to capture experience, labor force entry data and tradition. Given the fact that the set of occupations has been increasing since the economic opening of China in 1978, younger workers benefit more from new job opportunities than do older workers. In remote rural areas older workers, who have been working for their whole lives in agriculture or related activities, are often unwilling to invest in job training or to migrate to more developed areas to find a job in contrast to younger workers. Compared to younger workers, older workers also face more severe health problems, and those in rural areas are often illiterate. In her study about "work and gender among Uyghur villagers in southern Xinjiang", BELLÉR-HANN finds that:

> During the years of collectivization open apprenticing was not possible, although the trade could sometimes be learnt from a father who worked secretly at home. Such conditions limited the options of many young men who were brought up in the 1960s and 1970s. They typically started physical work at the age of ten or eleven to earn workpoints and through them grain for their families and many have remained illiterate. These men,

[14] Among China's unmarried women, however, migration to the Pearl-River delta is very common. Unmarried women often work for some years on the assembly line before they return home with their savings (GTAI, 2011, p. 39-40).

> now in their thirties and forties, represent a lost generation. …Today their options are extremely limited, many make cash through working as hired labourers, competing with each other on the casual labour market (*medikar baziri*). (BELLÉR-HANN, 1998, p. 8-9)

In the newly implemented Employment Promotion Law (cf., subchapter 1.3), age is not considered a potential source for discrimination; therefore, discrimination against the elderly is not prosecuted under the law (ROSS et al., 2007). Based on these facts I assume older individuals have on average higher probabilities to work in *A* than younger individuals in rural Guizhou.

This leads to the fifth testable hypothesis:

H5: Older age negatively influences access to non-agricultural employment.

To conclude I show the importance of human capital factors in a job advertisement of Huaxi Hotel located in Huaxi district of Guiyang (picture 2-1). The advertisement was placed at the entrance of the hotel. It indicates that the hotel is looking for personnel as 服务员 (fuwuyuan – waitress), 传菜员 (zhuancaiyuan – server) and 保安 (baoan – security personnel). The hotel positions with direct contact to guests (e.g., 服务员 – fuwuyuan and 传菜员 – zhuancaiyuan) require a certain beauty standard (height), age, culture and education, as well as Mandarin language skills. There is, moreover, a distinction between women and men. While women are obliged to apply for available jobs as waitress, men are obliged to apply as security personnel or as server. The ethnic status is not a direct concern in the advertisement, but ethnic minorities happen to accumulate lower Mandarin skills and *wenhua* requirements, which may reduce their chances of employment in Huaxi hotel. Only those ethnic minorities who speak Mandarin, have the required education and are completely integrated into the majority culture have a chance to get a job at Huaxi hotel.

Geographic location

Guizhou is a mountainous province in southwestern China. The western part of Guizhou belongs partly to the Tibetan high plateau and has varying elevations of 1,500-2,800 meters, the central plateau has an altitude of around 1,000 meters, and the Southeast of around 600-800 meters (ZHANG, 2003). As a result of the settlement of Han in the lowlands, ethnic minorities were pulled back to live in the mountainous areas of Guizhou (ZHANG, 2003, p. 282). Given the fact that ethnic minorities mainly live in remote rural areas, they face restrictions on available jobs if they are not willing to migrate to more developed regions. GUSTAFSSON and SAI (2006, 2008) find that geographic location is one major source for differences in poverty and income.

Picture 2-1: Job advertisement

We need the following personnel:

Waitress: 5 women. Age 18-28.
Height above 1.55 m, culture above junior middle school level, fluent in Mandarin.
Server: 2 men. Age 18-25,
Height above 1.60 m, culture above junior middle school level, can speak Mandarin.
Security: 1 men. Age 22-35.
demobilized soldiers with working experience have priority.

Huaxi Hotel

Source: Author.

This is in line with the findings of SCHEIN regarding Miao communities in Xijiang:

> Outlying villages, despite their widely ranging features, were uniformly described by Xijiang inhabitants as "backward" (luohou), "poor" (qiong), "small places" (xiao difang). Their villages were "dirty", often for lack of water, "dark" for lack of electricity, more reliant on homemade products (whether food, tools or clothing), less sophisticated in their technologies, and less savvy about the ways of the world… Peasants from "small places" signified the lowest level of a social hierarchy in which Xijiang residents were positioned at or near the top. (SCHEIN, 2000, p. 244)

These findings indicate that, in addition to the majority-minority segregation, there is also a kind of village ranking among residents of the same ethnic group. SCHEIN finds that in Miao communities "villages were evaluated according to size, proximity to the road and to long-distance bus routes, presence of a periodic market, availability of goods, electricity, and television etc." (SCHEIN, 2000, p. 240).

To capture major geographic particularities, I basically consider county dummies and village size in the econometric application. With the consideration of different counties, it is possible to capture geographic characteristics for each county. With the consideration of village size, measured by the registered households within the village, it is possible to capture local infrastructure characteristics. I assume that larger villages have positive effects on accessing employment in the *NA* sector. Larger villages are generally better connected with paved roads, bus routes and sometimes train stations than are smaller villages; access to *NA* employment is, thus, easier in larger than in smaller villages. It is, for example, possible for residents of better connected villages to commute to developed areas on a daily basis. Additionally local markets can develop more easily in larger than in smaller villages.

This leads to the sixth and seventh testable hypotheses:

H6: There are on average differences in occupational outcomes depending on the counties considered.

H7: Larger villages positively influence access to non-agricultural employment.

I developed hypotheses regarding the effects of ethnic status, education, gender, age and geographic location on occupational outcomes. The hypotheses were tested with qualitative and quantitative methodologies, and this is laid out in chapters three and four, respectively.

2.5.2 Methodological triangulation

Triangulation is an empirical research approach which derives empirical evidence with more than one methodology or data source to increase the amount of information in theory and hypothesis testing (SEAWRIGHT and COLLIER, 2004, p. 310).[15] I apply methodological triangulation by combining occupational outcome models in the quantitative portion and participant observations in the qualitative portion to analyze ethnic differences in occupational outcomes in rural Guizhou based on major human capital factors and geographic location.

I use occupational outcome models rather than wage equations or segregation indices as with occupational outcome models I can better reflect differences in job categories, most importantly between agricultural and non-agricultural sectors. I use participant observation rather than audit studies or structured interviews in the qualitative portion as field investigations about sensitive topics are restricted in China. I am, moreover, not required to do a full survey in the area as there is secondary data information conducted by the China Health and Nutrition Survey (CHNS) freely available. The field observations (chapter three) mainly have a supporting function for better understanding the results of the occupational outcome models (chapter four). I, furthermore, use field observations to refine hypotheses H1 to H7.

Table 2.4 relates the derived hypotheses H1 to H7 to the empirical application of occupational outcome models (second column) and to participant observation (third column). I use the key variables of the hypotheses as independent variables (IV) in occupational outcome models. The core factors which influence occupational outcomes are ethnic status, years of education, gender, age and geographic location. The CHNS data sample for Guizhou provides information about the Bouyei, Miao and Tujia in addition to the Han; therefore, I concentrate the empirical application on the three ethnic minorities and can combine a rigorous econometric application with supporting field observations for the three groups. The dependent variable in the occupational outcome models is a discrete set of occupations with major focus on binary comparisons of agricultural and non-agricultural sectors. I test

[15] This approach was first applied by CAMPBELL and FISKE (1959, p. 38-39) who name the philosopher FEIGL (1958) as pioneer of triangulation.

Table 2-4: Methodological triangulation and hypothesis testing

Hypotheses	Occupational outcome models	Participant observation
H1: Being an ethnic minority negatively influences access to non-agricultural employment.	IV – dummies for Bouyei, Miao, Tujia (Han base group)	D, F, S
H2: More years of education positively influence access to non-agricultural employment.	IV – years of education	D, F, S
H3: The lower educational achievement of ethnic minorities negatively influences their access to non-agricultural employment.	IV – interaction terms between Bouyei, Miao, Tujia and education	D, F, S
H4: Being female (male) negatively (positively) influences access to non-agricultural employment.	IV – dummy for male	D, F
H5: Older age negatively influences access to non-agricultural employment.	IV – age in years	D, F
H6: There are on average differences in occupational outcomes depending on the counties considered.	IV – dummies for counties	D
H7: Larger villages positively influence access to non-agricultural employment.	IV – number of registered households in villages	D

IV stands for independent variable. D, F, S stand for descriptive, focused and selective phases, respectively.

Source: Author.

H1 to H7 by measuring the effects of major explanatory variables on the discrete set of occupations (see table 2.4). I use dummy variables for Bouyei, Miao, Tujia and Han-base group for testing H1. I use years of education for testing H2. For testing H3, the interactions between education and ethnic status, I use interaction terms of ethnic status and education to test their combined effect on occupational outcomes. To find out about gender differences, I use a male dummy to test H4. I use age in years to test H5. To measure effects from geographic location, I test H6 and H7. In order to disentangle ethnic and village effects, I only consider areas with mixed ethnic populations. I use county dummies to test main geographic effects with H6. During my field observations, however, I found that village size, i.e., number of residents in a village, is an important determinant for capturing employment chances as larger villages are generally better connected than smaller villages and, thus, facilitate access to *NA* employment; therefore, I test whether village size influences occupational outcomes with H7.

In chapter four I estimate occupational outcome models for the years 2004, 2000 and 1997 with variations in dependent variables (discrete set of occupations with

major focus on comparisons of agricultural and non-agricultural sectors). I also explain the CHNS dataset and the data sample used in my study. I further provide descriptive statistics and capture the dynamics of the effects over those years.

In the qualitative portion of my study, I use participant observation (see table 2.4). Participant observation is characterized by three different phases with different levels of accuracy in the observations: descriptive (D), focused (F) and selective (S) phases. In line with the underlying hypotheses, I apply different levels of accuracy in observing the key variables which influence occupational outcomes: ethnic status, years of education, gender, age and geographic location.

I concentrate on capturing effects of ethnic status and of education on occupational differences. All three phases D, F and S are, thus, used to test H1, H2 and H3. This means that I not only capture the complexity of the field in relation to these variables, but also identify specific problems and gather additional information through informal conversations and unstructured interviews with locals and key persons. I determine whether occupational outcomes differ between gender and age groups in two phases, D and F. I observe these two variables more generally and identify some specific problems related to gender and age. While travelling through Guizhou I observe the geographic particularities as postulated in H6. This approach actually served to formulate H7 as an additional explanatory variable in the occupational outcome model. I basically observed that in Guizhou larger villages are much better connected than remoter villages. In chapter three I explain in detail my qualitative approach of participant observation, provide field observation results and supporting photos. The results from occupational outcome models and from participant observations combined are much more reliable for better understanding ethnic differences in occupational outcomes than are single empirical approaches.

2.6 Conclusions

In this chapter I linked and classified theoretical and empirical concepts to measure ethnic differences in occupational outcomes and wages, which served to make assumptions and testable hypotheses as a basis for a rigorous empirical application of combined methodologies. I interlinked human capital and group differences theories, behavioral theories, labor market discrimination theories, occupational choice theories, farm household theories and non-farm rural employment theories based on the benchmark model of JOHNSON and STAFFORD (1998) to get a comprehensive theoretical framework for analyzing ethnic differences in occupational outcomes in rural areas. I arranged the theoretical concepts in a diamond of theories based on four major theoretical approaches: employer discrimination, institutional discrimination, abilities and preferences. These four major theoretical approaches for analyzing ethnic differences in occupations and wages are linked to each other with several causal directions. They can be combined with other theoretical

approaches such as farm household models, demand-pull/distress-push concepts, occupational outcome models and the sustainable livelihood framework.

While all theoretical approaches for understanding the phenomenon of ethnic differences in occupational outcomes and wages are important, investigating them empirically is challenging since employer discrimination is forbidden in China. The literature on China suggests that in XUAR and TAR the Uyghurs and Tibetans, respectively, face statistical discrimination. In contrast in Guizhou and Yunnan it is actually Han who face taste-based discrimination. The case of the Hui varies. As there are 55 classified ethnic minorities alongside the Han people in China, more empirical applications are definitely needed.

The application of econometric modeling to analyze ethnic differences in occupational outcomes and wages with occupational outcome models or wage equations is, however, often affected by omitted variable bias (the absence of significant variables), overfitting (too many insignificant variables) or endogeneity problems (unclear causal relationships). Qualitative methodologies such as audit studies, interviews or participant observation face shortcomings in biased selections, in the inability to capture evolution over time, in the inability to generalize research findings and in researcher bias. The empirical analysis of labor market discrimination, therefore, requires rigorous assumptions to clarify results.

I use the diamond of theoretical principles as theoretical foreknowledge to explain ethnic differences in occupations. This theoretical foreknowledge was then adjusted to my case study as relevant information evolved during field work and from additional literature at later stages in the research process. I made two assumptions to narrow down the theoretical concepts: first all individuals regardless of their ethnic affiliation prefer to work in the non-agricultural sector, and second there is only employer discrimination and no institutional discrimination (discrimination by law) in China. Theoretical explanations of ethnic differences in occupational outcomes in rural Guizhou are, hence, based on differences in abilities and employer discrimination, which are the core links between the theoretical framework and the empirical application.

Based on these two theoretical concepts and the importance of geographic location, I postulated that ethnic minorities face three constraints for accessing employment in the non-agricultural sector: 1) individual abilities are not adequate to perform non-agricultural work, 2) employers in the non-agricultural sector discriminate against ethnic minority workers and 3) non-agricultural work is not available in the area. In line with previous studies, I use human capital factors and geographic location to capture these effects in the empirical application. I use triangulation of quantitative and qualitative methodologies to get more reliable results. I use participant observation in the qualitative portion (chapter three) and apply occupational outcome models in the quantitative portion (chapter four) to analyze ethnic differences in occupational outcomes in rural Guizhou.

3 QUALITATIVE APPROACH AND EVIDENCE

In this chapter I present the qualitative approach applied in my study and evidence from the field. The results serve to better understand ethnic differences in Guizhou's rural labor market and provide supporting evidence for the occupational outcome analysis in the next chapter. In section 3.1 I describe my qualitative methodology, present pretest results and selected study areas. In section 3.2 I describe my experiences and observations obtained in the selected areas in Guizhou and show supporting pictures. In subchapter 3.3 I draw conclusions.

3.1 Participant observation

In section 3.1.1 I explain the qualitative approach of participant observation used in this study. I apply this research tool to better understand the rural labor market situation in Guizhou, to provide some supporting evidence for the occupational outcome analysis in the next chapter and to refine the hypotheses of this study. The pre-test results are presented in section 3.1.2 and the selected study areas in section 3.1.3.

3.1.1 The approach used in this study

I was not required to do a full survey in the area as there is secondary data information freely available from the China Health and Nutrition Survey (CHNS). Depending on the research focus and area, it is possible to apply the three phases of participant observation. The field observation mainly functions to allow better understanding quantitative model outcomes. I conducted the field observation from March 27 to April 29, 2010 in Guizhou. I made "ad-hoc" observations of the field and had informal conversations with locals. I acquired additional knowledge regarding my investigations through conversations with foreign and Chinese scholars of Guizhou University during winter term 2010 (August 24 to December 23), when I was studying Chinese language at Guizhou University thanks to a Confucius Institute scholarship and IAMO funding. Throughout the next chapter I include the knowledge I acquired through these informal conversations and when appropriate, experiences I had during that time.

I used the following approach to conduct participatory observation of rural labor markets in Guizhou. First, I formulated a travel plan through rural areas of the province (cf., subchapter 3.1.3). I particularly investigated those areas where Bouyei and Miao groups are living because I found in primary estimations of secondary data that these two ethnic minority groups have lower probabilities of working in the non-agricultural sector than do Han; results for the Tujia were not significant

and, thus, not the primer focus of my investigations. In most of the cases I resided in the district towns and visited the ethnic minority villages daily.

Second, based on the three phases of participant observation, I formulated some questions on which I focused my observations during the different phases. These questions served only as a guideline but are not directly referred to when presenting my results. I compiled field notes based on the ethnographic action guidelines put forth by TACCHI et al. (2003), which include notes from observations and from informal conversations with locals, researchers and students. I additionally include pictures, several samples of print media and other artifacts from the areas of investigation. The final goal in the interpretation of the field notes is to allocate dominant factors of ethnic groups and to relate them to their occupational outcomes.

3.1.1.1 Descriptive Phase

The descriptive phase serves to get a general understanding of the field. The goal is to analyze the ethnic groups and household members considered in respect to the following general questions in around two days in each study area.

- Is there a local residential segregation?
- Are there poorer and richer households? In case there are, what are the observed differences?
- Which occupations are available in the villages?
- Where do the agricultural households specialize?
- How many kindergartens, schools etc. are available?

3.1.1.2 Focused Phase

During the focused phase, which considers specific problems, processes and persons, I considered the following questions in around three days in each study area.

- What are the daily tasks of the household members?
- Which occupations do household members work in?
- Who works outside in the fields?
- What are the occupations of landless peasants?
- What does inter-household labor cooperation look like? Are there traditional norms of labor division between the household members?
- What kind of community work has to be done and who does it?[16]

[16] For example BELLÉR-HANN (1997, 1998) finds that in southern Xinjiang, community work is unorganized and unpopular because households prefer to follow an income generating

- What are the differences in age, social and economic position?
- To which ethnic group do the policemen belong?
- How many hours and working days per week are the household members engaged in income generating activities?

3.1.1.3 Selective Phase

The selective observations serve to gather additional information of already identified patterns and forms of behavior of local actors. My preliminary estimation results suggested that education is an important factor in determining occupational outcomes. For example individuals with more years of education have on average higher probabilities of working in non-agricultural positions than do individuals with fewer years of education. In the selective phase I, thus, seek to get more knowledge about which role education plays in accessing different occupations and whether or not there are differences between the ethnic groups considered.

3.1.2 Pre-test to field study

I conducted participatory observation of foreign merchants at the daily farmer's market in the city center of Halle (Saale) on the afternoons of March 1 and March 3, 2010. This pre-test provided a better understanding of the methodology of a distant and detached observer. I managed to analyze the foreign merchants and to identify hypotheses as a basis for further studies.

With the method of a distant and detached observer and in particular through informal conversations, it was possible to discover market regulation problems, which served for identifying hypotheses for explaining the underrepresentation of foreign merchants and the competitive advantages of established merchants compared to new merchants at the daily farmer's market in Halle (Saale).[17]

3.1.3 Selection of study areas

In Guizhou there are two autonomous prefectures and three autonomous counties where the Bouyei and the Miao are residing together (see table 3.1). Qiannan Bouyei-Miao Autonomous Prefecture is located 152 km southeast of the provincial capital Guiyang and Qianxinan Bouyei-Miao Autonomous Prefecture is located 330 km southwest of Guiyang. The three autonomous counties are all located around 150 km southwest of Guiyang. A crucial income source for ethnic minorities as pointed out by GUSTAFSSON and LI (2003) is the tourism industry in Guizhou. The observation, thus, concentrates on the above-mentioned autonomous prefectures and counties as well as the major touristic focal points of the province, which are located in Qiandongnan Miao-Dong Autonomous Prefecture and Anshun Prefecture.

activity; poorer men (often young boys) get paid to do the community work for those who are better off.

[17] The results of the pre-test are available upon request.

Table 3-1: Bouyei-Miao autonomous prefectures and counties

Guizhou Province – Autonomous Prefectures
Qiannan Bouyei-Miao Autonomous Prefecture 黔南布依族苗族自治州 (Qiánnán Bùyīzú Miáozú Zìzhìzhōu) Founded Aug. 8, 1956, Area km² 26.193, Population (thousand): 3.790.1 Ethnic minority proportion: 55.28 % Capital: Duyun City 都匀市
Qianxinan Bouyei-Miao Autonomous Prefecture 黔西南布依族苗族自治州 (Qiánxī'nán Bùyīzú Miáozú Zìzhìzhōu) Founded May 1, 1982, Area km² 16.804, Population (thousand): 3.016.2 Ethnic minority proportion: 42.94 % Capital: Xingyi City 兴义市
Guizhou Province – Autonomous Counties
Zhenning Bouyei-Miao Autonomous County 镇宁布依族苗族自治县 (Zhènníng Bùyīzú Miáozú Zìzhìxiàn) Founded Sept. 11, 1963, Area km² 1.721, Population (thousand): 334.6 Ethnic minority proportion: 58.61 %
Ziyun Miao-Bouyei Autonomous County 紫云苗族布依族自治县 (Zǐyún Miáozú Bùyīzú Zìzhìxiàn) Founded Feb. 11, 1966, Area km² 2.284, Population (thousand): 322.4 Ethnic minority proportion: 68.44 %
Guanling Bouyei-Miao Autonomous County 关岭布依族苗族自治县 (Guānlíng Bùyīzú Miáozú Zìzhìxiàn) Founded Dec. 31, 1981, Area km² 1.468, Population (thousand): 320.0 Ethnic minority proportion: 58.99 %

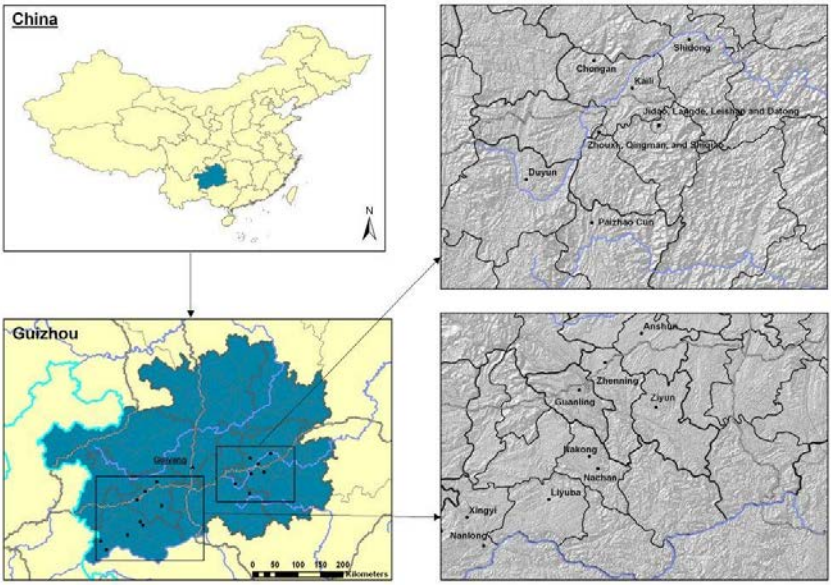
Source: CHINA.ORG.CN (2005).

The touristic focal points are selected based on information from travel guides. One guide clearly points out that the government seeks to overcome poverty and tries to increase development by promoting ethnic minority cultures as a local attraction.

Map 3-1 depicts the destinations in Guizhou. The provincial capital Guiyang, indicated with the triangle, is the starting point of the field observation. From here I first visited Qiannan Bouyei-Miao Autonomous Prefecture and Qiandongnan Miao-Dong Autonomous Prefecture, which are shown in the upper right map.

The city of Kaili is the center of Miao silver culture and the entry to neighboring ethnic minority villages. The next step is to visit the autonomous counties Zhenning, Ziyun and Guanling before going to Qianxinan Bouyei-Miao Autonomous Prefecture. These areas are shown in the lower right quadrant in map 3-1. The first stop on this route is Anshun, which is an important commercial city in western Guizhou. The Huangguoshu Falls and Longgong caves, which are important tourist attractions of Guizhou, are easily accessible from here. These two touristic sites and

Map 3-1: Areas considered in Guizhou province



Source: Author, illustration Jens Frayer.

the Sunday market in Anshun are particularly interesting for my research as ethnic minority traders sell their handicrafts there. The final destinations of my route are Xingyi, the capital of Qianxinan Bouyei-Miao Autonomous Prefecture and the neighboring ethnic minority villages.

3.2 Evidence from the field

The results of the aforementioned approach of using participant observation to analyze the field are provided in this subchapter. I conducted the field observation in Guizhou from March 27 to April 29, 2010. Additional knowledge was acquired in the winter term 2010 (August 24 to December 23), when I was studying Chinese language at Guizhou University. As already pointed out, the goal of participant observation is to get progressively involved with the field and the people and to develop the analysis from a broader observation of the field to more concrete attention to the research questions (FLICK, 1995). I show the differences between occupations in the agricultural (*A*) and non-agricultural (*NA*) sectors and particularly look at observable ethnic differences in the areas considered. Sometimes the results are much broader than the focus of this monograph, so that this subchapter also provides many insights for important future studies. The results in this subchapter are closely linked to the areas shown in map 3-1, which means that

the experiences in each area are given progressively in this subchapter. The links to other information sources as well as comparisons among the regions are also identified.

I start with general findings about using the approach of participant observation with focus on the identification of ethnic minorities and occupations. Then I analyze the findings in each area: first the findings in the Qiandongnan Miao-Dong Autonomous Prefecture and Qiannan Bouyei-Miao Autonomous Prefecture, second the findings in the three autonomous counties, Zhenning Bouyei-Miao Autonomous County, Ziyun Miao-Bouyei Autonomous County and Guanling Bouyei-Miao Autonomous County and finally the Qianxinan Bouyei-Miao Autonomous Prefecture. The locations of all areas are shown in map 3-1. In most of the cases I resided in the district towns and visited the villages daily.

In general with observation techniques I could only distinguish the ethnic minorities from their traditional clothing, which is very commonly worn on a daily basis by women, but not by men. Some women use simple towels as headdress, which makes the accurate identification of ethnic status difficult. I must point out that checking the identity card of each individual is the only way to be absolute certain that ethnic status is accurately recorded. This fact is very important to keep in mind while reading this chapter as I will not constantly point it out in the following.

In all areas I visited in Guizhou, agriculture (*A*) was done using traditional methods. The major reason is that the mountainous topography of the province makes it impossible to employ modern technologies to do the work more efficiently. For example, SCHEIN (2000) writes about the hard manual work during the rice harvest in Xijiang and concludes that:

> Xijang peasants and development consultants alike saw no alternative to this method of harvesting. The rice terraces were small and scattered through the mountains; there was no way for any vehicles to reach them. Even mechanical threshers could not be transported to the fields because of the narrow, steep, and muddy paths that defied anything on wheels. (SCHEIN, 2000, pp. 161-162)

This situation does not only prevail during rice harvest, but for all other crops and plants, which are set on the same fields. Pictures 3-1 show traditional agricultural methods used throughout Guizhou. The occupations available in the non-agricultural (*NA*) sector in the research area can broadly be distinguished between construction sector and service sector and corresponding subcategories of both sectors. Pictures of work in the *NA* sector are provided throughout this subchapter as they differ between the regions.

One major subcategory of the service sector is tourism, which is particularly important for ethnic minorities of the province as their traditional lifestyle is the target of governmental support. Since the opening reform in 1978 the ethnic tourism industry of Guizhou has been increasingly popular and, therefore, has enhanced

Pictures 3-1: Agriculture (1)

Source: Author.

job opportunities for ethnic minorities, particularly for women as stressed by SCHEIN (2000).[18] I also find that a touristic infrastructure focusing on ethnic minority culture, festivals, handicrafts and embroidery was implemented in the province. The most popular touristic area in Guizhou is the Qiandongnan Miao-Dong Autonomous Prefecture with the prefectural capital, Kaili. On my way from the provincial capital Guiyang to Kaili, I shared a seat with a 30-40 years-old man with Miao status. He informed me that the travel time between the two cities had been reduced tremendously; while five years ago the trip would take seven hours, today it is reduced to 2.5 hours. This indicates the development of infrastructure in the region. Generally I observed that by travelling independently, the locals of the touristic villages were often not prepared for my arrival, so that traditional performances were not shown. I observed that in most of the villages only registered travel groups receive special attention when, for example, local festivals and performances take place. For example when I participated in the "Photo China original conference" in mid-September 2010 in the city of Tongren in Guizhou, an excursion to mountain Fanjing and a neighboring Tujia village was part of the program. As we were a large group of Chinese and international photographers, and as the excursions were organized by a local agency, we received a huge welcome ceremony with Tujia people playing their traditional instruments in front of the mountains and Tujia people who gave an address of welcome at the village entrance (pictures 3-2).

In the village traditional Tujia food was served and a traditional wedding ceremony was reenacted. From these contrasting experiences I learned that tourism may for many locals not be the main daily activity and income source, but rather provides an additional income when particular festivals or conferences take place in the region.

[18] In other areas, such as Yunnan, the Bouyei have job advantages from the tourism industry (BHALLA and QUI, 2006, p. 68), while in Tibet most of the jobs in the *NA* sector, even in the tourism industry, are taken by non-Tibetans (HILLMAN, 2008, p. 9).

Pictures 3-2: Welcome ceremony and address of welcome by Tujia people

Source: Author.

3.2.1 Qiandongnan Miao-Dong Autonomous Prefecture

The first autonomous prefecture I visited in Guizhou was Qiandongnan Miao-Dong Autonomous Prefecture. Starting in the capital Kaili, I visited several different places in the prefecture: Shidong, Chongan, Jidao, Langde, Leishan, Datong, Zhouxi, Qingman and Shiqiao. I visited all villages except Chongan with a hired car and driver. In this area both Miao and Dong people were working in tourism; this indicated the good job chances for ethnic minorities in the tourism sector in this prefecture.

One Miao informant said that traditional clothing, local accents, dialects, languages and other particularities can change from one village to the next. This is also shown on the CITS (China International Travel Institute) website:

> During the course of their migrations the Miao diversified into subgroups, known as Black Miao, Red Miao, White Miao, Long-Horned Miao, and Flower Miao, after their style of dress. This is, however, largely a Han classification and few Miao use these names among themselves. In order to make these subgroups clear, we name them according to the regions they are living today… (CITS, 2010).

On the website they distinguish the Miao groups into eight main regional groups, which are Southeast Guizhou, South Guizhou, Anshun region, Bijie region, Zunyi region, Liupanshui region, Southwest Guizhou and Guiyang region; however, as yet not all regions have been investigated and opened for tourism; this list is far from complete. The embroidery and silver jewelry in these eight main Miao regions are further distinguished in subgroups on the website, with each subgroup having their own distinctive styles (see table 3.2). It is further pointed out on the website that by their style of dress, people are able to identify the village or region of a person. During festivals Miao women also wear silver jewelry, such as neck rings, chains, chest locks and multiple headdresses. I was informed that it is, however, difficult to preserve the embroidery culture as many young girls prefer western clothes and are unwilling to spend much time embroidering.

Table 3-2: Traditional clothing of the Miao in Guizhou

Region	Number of different traditional dresses
Southeast Guizhou	39
South Guizhou	10
Anshun region	12
Bijie region	10
Zunyi region	6
Liupanshui region	5
Southwest Guizhou	4
Guiyang region	7

Source: CITS (2010).

According to an older Miao woman in Shidong village, who offered embroidery and silver jewelry to tourists, the dresses, furthermore, identify the marital status of women, making distinctions between single, married or widowed.

In a conversation about Miao-Han relationships, informants said that for historical reasons the relationships are not favorable. It was, furthermore, pointed out that even Miao groups do not get along very well with each other, which was explained by different moral standards between Miao subgroups. For example the Miao from Shidong are considered to be very good at business and are stereotypically called the Jews of the Miao people, while other Miao people in southeastern Guizhou are more concerned about relationships and traditions. When I asked about how Han do in business, the informant replied with an obvious smile that Han are even better at business people than are the Miao of Shidong.[19] In future studies it would, thus, be interesting to consider personality traits of ethnic groups in occupational outcome analysis as implemented in the research of HAM et al. (2009). Another example of ethnic relationships concerns one informant, who is Dong, and his wife, who is Han and originally from Sichuan. They decided that their daughter should receive the Dong instead of the Han status, so that one day, given the additional points which ethnic minorities receive on the entrance examination to access higher education, her chance to enter university would be higher. It can be said that this family chose the ethnic status of their daughter for practical reasons rather than based on ethnic allegiances.

On the main road from Kaili to Shidong many individuals were engaged in agricultural activities. Pictures 3-3 show the major agricultural activities being conducted at that particular time. The picture on the left shows a peasant plowing a field with the help of a water buffalo. The picture on the right shows three women hoeing the field. From their headdresses I inferred that the women belong to the

[19] It is interesting to note that PILLSBURY (1973) cited in GLADNEY (2004, p. 167-168) finds that the Hui are stereotypically called the "Jews of China" for their good entrepreneurial skills. In one interview (GLADNEY, 2004, p. 293) he was even informed that: "The Hui are good at doing business; the Han are too honest and can´t turn a profit. Han are good at planting, Hui at trade."

Pictures 3-3: Agriculture (2)

Source: Author.

Miao group. The man in the picture on the left, however, wears average non-traditional clothes. It was, therefore, not possible to identify his ethnicity. One conclusion that can be drawn is that with observation, it was impossible to accurately ascertain ethnicity.

The country roads had no sidewalks (for pedestrians) in Guizhou; people, therefore, had to walk in the road. Particularly school children commuting from home to primary or middle schools located close to the main road and individuals engaged in agriculture were using the roads to go to their destinations. I perceived the mixture of fast driving vehicles and pedestrians on the same road as very dangerous. All drivers just honked before coming around a curve with the expectation that pedestrians would let them pass. Many pedestrians were, however, not very concerned about the vehicles and moved to the side of the road only when they were directly confronted with vehicles.

Shidong is a town located at a river in the northern part of Qiandongnan Miao-Dong Autonomous Prefecture. It is famous for Miao silver culture, Miao embroidery and the dragon boat festival. I met a young female English teacher of the local middle school. She informed me that she teaches two classes with around 48 pupils in every class. As middle schools are only located in towns and not in villages, I asked her about facilities for pupils from remote villages. The teacher waved towards the back of the school to indicate the newly constructed dormitories for children from distant villages.

She additionally informed me of the importance of the dragon boat festival for Miao people of the area. She said that every village has its own team and boat for the competition. The winner is highly lauded and receives a lot of homemade gifts and food, which spectators hang at the bow of the boat. At the dragon boat festival Miao people wear special traditional clothes.

The teacher said that, although life is difficult and income is low for everyone, the purchase of this particular traditional clothing worn on the festival day is very

important, to the extent that people who cannot afford the traditional clothing cannot take part in the festival. This indicates that the traditional clothing worn during festivals is an important share of household assets, which is worth considering in household surveys. In Shidong I, moreover, was offered the chance to see Miao jewelry in the house of a Miao woman. Her house is not only her home but also her business studio, as she sells the jewelry to tourists there.

I also visited Chongan to identify the vendors at the local market by ethnicity. The market was, however, so crowded that the exact location of each stand with information about the vendors and buyers, as gathered in the pre-test conducted in Halle (Saale), was impossible to capture (pictures 3-4). All kinds of traditional and modern products were sold at the market. Some of the female vendors and buyers wore traditional clothes, yet there were several distinct headdresses visible, I was unsure whether or not they were all Miao women or belonged to other local ethnic minority groups, such as Dong, Gelao or Shui groups. On the return trip to Kaili, an English-language movie with Sylvester Stallone was shown in English to the passengers; this made me think that the influence of western media and commodities on local ethnic communities is also a very important topic for future studies.

The next day I went with the same local guide from Kaili to Jidao, Langde, Leishan and Datong. It was very interesting to see the amount of construction that was underway, whether to build new or to improve existing roads or to build houses. Both women and men were doing the work, while tasks which particularly demanded strength were solely done by men. In all the areas I visited in Guizhou the plowing of fields with water buffalos was solely done by men.

In pictures 3-5 it is clear from the headdresses of the women that they belong to the Miao group. From this observation I, thus, deduced that Miao women are working

Pictures 3-4: Market scenes in Chongan

Source: Author.

Pictures 3-5: Miao women working in the construction sector


Source: Author.

Pictures 3-6: Brick production


Source: Author.

Pictures 3-7: Paper production


Source: Author.

in the construction sector; unfortunately observation techniques could not help me determine whether or not the women were engaged in assigned community work or in income generating activity. BELLÉR-HANN, for example, finds that in southern villages in XUAR: "Communal work has survived into the reform era and is the obligation of landholding peasants to work for the township in the construction and maintenance of irrigation canals, roads and schools, and in opening up wasteland to cultivation" (BELLÉR-HANN, 1998, p. 8).

The next day I visited Zhouxi, Qingman and Shiqiao. People were cutting stones and selling gravel along the roadside. I assume that there is a quarry nearby. Pictures 3-6 show that both women and men work in brick-making activities. Zhouxi, Qingman and Shiqiao are all tourist villages, yet in April 2010 there were only a few other people besides me visiting the villages. This indicates that tourism is probably a sideline activity during festivals and holidays and not the major income source of the locals. Shiqiao is famous for traditional paper making, and many households were engaged in this work. Pictures 3-7 show part of the traditional paper-making process. Shiqiao was the last village I visited in Qiandongnan Miao-Dong Autonomous Prefecture.

3.2.2 Qiannan Bouyei-Miao Autonomous Prefecture

The second area of investigation as indicated in map 3-1 is the Qiannan Bouyei-Miao Autonomous Prefecture, the district town of which is Duyun. The two cities, Kaili and Duyun, are connected with a highway, which shows again the improved infrastructure development in the area. The two young (20-30 years old) female receptionists in the hotel in Duyun were very friendly and asked me about my country of origin. They informed me that my skin is very white and beautiful. As they were both employees in a hotel, a particularly fair skin color may have been an important attribute for them to work in this sector. I asked the receptionists if they belong to either the Bouyei or Miao ethnic groups, and they replied with a little hesitation that they are both Miao. In this area the women dress in Western clothes on a daily basis. An informant from Kaili said that people of Qiannan Bouyei-Miao Autonomous Prefecture are already sinicized to a large extent. The informant further pointed out that the name Bouyei actually implies that they "follow the emperor", while Miao is related to inferior meanings, which were not specified. This statement can mean that in the present the Bouyei tend to follow the mainstream Han culture and language, rather than the Miao group. This is also shown on the CITS website:

> The Miao, or Hmong as they prefer to be called, are thought to have migrated 3000 years ago from an area north of the Yellow River to South Yangtze River, and migrated into Guizhou 2000 years ago. …. Many Miao continued their migration beyond China, into Laos, Thailand and Vietnam where they are known as the Hmong (or Mong). The Miao have a reputation as independent-minded and rebellious highlanders. Many Miao joined the armed uprising against the Qing government, from 1840 to 1870, which became known as the Miao Rebellion. Numerous Lao Hmong worked covertly for the US government

> during the American (Vietnam) War and settled in the USA after the fall of Saigon. (CITS, 2010)

SCHEIN (2000, p. 281), however, points out that the Miao are not accepted by the Hmong in the West as their own people. She added that the Ge of southeastern Guizhou also refused to be categorized as Miao. Ethnic classification is, therefore, an interesting topic for further investigation.

After I arrived in Duyun I visited a bookstore. I asked the female salesperson (30-40 years old) about her ethnic origin, and she informed me that she is Bouyei but that she cannot speak Bouyei language because in school she was only taught in Mandarin. I inferred from that that with her family she also used Mandarin. Around Duyun I only visited the hills in the surrounding area where tea is grown (pictures 3-8), because with pure observation it was impossible to identify ethnic differences here.

In this area I was, nevertheless, able to participate in ancestor worshipping in Paizhao village. John, who is a former university classmate of one of my informants, and his girlfriend Diana took me to Paizhao village, John's hometown. SCHEIN explains that "[i]n April, timed with the Han festival of Qingming, descendants tended their ancestor's graves, or in special cases, they erected new gravestones in a ceremony involving shamanic offerings, fire-crackers, and animal sacrifice" (SCHEIN, 2000, p. 214).

Paizhao is a Bouyei village. To reach Paizhao, we could only take a public train from Duyun to Dushan and had then to rely on a villager, who used his own driving services from Dushan to Paizhao. I was informed that an alternative to go to the village was by motorbike taxi. John said that today almost every household has a motorbike and that the dream of the young people is to study at a university or to go to work in one of the large cities in the coastal areas of China. Work-related migration to coastal areas is called 打工 (*dagong*) and can be translated as "to job". One researcher of Guizhou University informed me that workers mainly leave Guizhou to job in Guangdong, which is considered "the factory of the world." To

Pictures 3-8: Tea harvesting

Source: Author.

prevent workers from leaving the companies in Guangdong to help with the harvest at home, labor contracts normally last from one spring festival to the next.[20] Although there was a lot of work to do on the farm, young peasants preferred to job for some time in the coastal cities. The major reason was the better pay and lifestyle there. John said that many migrants were able to build new houses in Paizhao after jobbing for some time in Guangdong. In Guangdong they earned the necessary money and learned the necessary skills for building new houses at home. John added that roofs of traditional houses leak, while the new houses made of concrete are much more comfortable in this respect. I observed that buildings are usually without heating systems, so that the interior of those buildings made of concrete is particularly cold during autumn and winter.

John also talked about the life of peasants in the village. He said that their most important income source is selling rice. A newly established policy encourages farmers to grow rice and guarantees a minimum governmental payment if the harvest cannot be sold directly on the market. During my visit the farmers were, however, facing hardship with a drought in the area and feared losing their harvest. In Paizhao some peasants irrigate their fields by pumping water from the nearby river; however, in other areas without natural water sources this was not an option, and in periods of drought, the harvest might be lost. On the evening of April 4, 2010, we watched the local television news together. In Guizhou, Guangxi and Yunnan the Premier Wen Jiabao, visited people who were suffering from water shortages. The local television news, moreover, announced that famous Chinese actors were initiating a donation campaign to help the victims; however, it appears that the support was not sufficient and that only some of the victims actually received help.

John explained that land and cattle are the most important resources for village peasants who produce rice. He emphasized his point by giving me an example about a farmer who committed suicide because he accidentally lost his only cattle in the mountain slopes. On another occasion we talked about John and Diana's future prospects and plans. John and his family belong to the Bouyei group. His girlfriend Diana comes from rural Hubei and belongs to the Han group. They are planning to marry and will be allowed to have two children instead of one because John is Bouyei. He was very worried about the expenses of a wedding. He earns 2,500 CNY (roughly 290 EUR) monthly as a teacher at a university. When he marries, he has to buy a new apartment, pay for the wedding ceremony and pay the bride price to Diana's parents. To cover all the expenses, he has to take out a loan from a bank which generally has very high interest rates. He continued that later expenses for the tuition of the children will be another burden, so that he will finally be 60 before he can start enjoying life without monthly financial duties. He added that

[20] Spring festival (Chinese New Year) is the most important festival in China and families usually celebrate together at home.

he, thus, tries to enjoy every day without worrying about the future. John also said that he would like to travel the world to get to know different societies and that he envied me for being able doing that. His dream, he added, is to visit a traditional village in Africa.

John once wanted to improve his parents' life and offered to let them live with him in the city. His parents, however, preferred to stay in Paizhao with other family members and friends. John's father also refused to have additional money invested in the house, which is old enough to have deteriorated. The father only asked for a little sofa to sit on, as his knees hurt when he was sitting on the very small stools that are traditional seating in China. I was informed by a researcher familiar with the topic that in rural Guizhou it is traditional for peasants to pass their houses to the youngest sons. Although John is the youngest son in the family, he lives and works in the city. This might be another reason that the father was not willing to invest further in the house. John's brother is a farmer in Paizhao and already constructed a new house there, where he lives with his family. A researcher told me that before the economic opening of China, elder sons of poor households in Dong communities did not have the money necessary to invest in new houses; the parental house was, therefore, extended with a separate space for each son. The researcher added that with job possibilities in tourism and *dagong*, peasants' income has been increasing, so that older sons started constructing their own houses again, which started a construction wave in Dong communities.

In another conversation with John and Diana, we talked about differences in traditions and university access of ethnic groups. John said that the Shui ethnic group still strictly follows their traditional lifestyle. I asked why there have been differences between the Shui and the Bouyei, and John replied that in Qiannan Bouyei-Miao Autonomous Prefecture the Bouyei have been largely influenced by Han people. He said this in a humorous way so as not to offend his girlfriend, Diana, who is Han and originally from Hubei, and softly pinched her back. She smiled, a little embarrassed. This is in line with the view of another informant from Kaili, who also said that the Bouyei of Qiannan Bouyei-Miao Autonomous Prefecture are highly sinicized; traditional Bouyei culture is better observed in the autonomous county Zhenning close to Anshun.

In another conversation with John and Diana, Diana informed me that she would have liked to study English language and literature in Hubei, but that she was not accepted at a university in Hubei. She said that ethnic minorities, who receive higher test scores by law, get study places instead. Diana received a study place at a university in Duyun. The tuition fees in Duyun were lower, which reduced the financial burden of her parents, so that she finally considered studying at Duyun University her best option.

After three days in Paizhao, I went by public bus back to Duyun and from Duyun to Anshun in order to investigate autonomous counties in Anshun prefecture.

3.2.3 Autonomous counties in Anshun Prefecture

In Anshun prefecture I observed the three autonomous counties, Zhenning Bouyei-Miao Autonomous County, Ziyun Miao-Bouyei Autonomous County and Guanling Bouyei-Miao Autonomous County. Additionally I found out about the Han people of Tunpu garrison and also made an investigation in this area.

In Zhenning Bouyei-Miao Autonomous County I first went by public bus to Shitou village, a Bouyei village. Traditional Bouyei houses in this region are of stone; in fact the name of the village, 石头 (shitou), actually means stone. Pictures 3-9 show Shitou village. In the picture on the right an elderly Bouyei woman with her traditional headdress is visible. From my personal view the village is not very touristic, although Huangguoshu Falls, the largest waterfall in China, is located nearby, and the area seemed much poorer than the touristic places I visited in Qiandongnan Miao-Dong Autonomous Prefecture. As I was not part of a tourist group, the people in Shitou village were not prepared for my arrival. People, therefore, followed their daily routines and also did not ask me to pay the entrance fee as indicated on a sign at the village square. For the villagers tourism is, thus, a sideline, probably only conducted when registered travel groups arrive in the village.

I strolled in the area and could observe many peasants engaged in traditional agricultural activities. Most of the women were not wearing traditional headdresses or clothes, so that I could not observe to which ethnic group they actually belonged. Peasants were using the same traditional method of plowing the fields as in all other areas I visited. There was less construction of new houses and roads in this area compared to Qiandongnan Miao-Dong Autonomous Prefecture. Many women of Shitou village were washing or dyeing clothes. One woman was not very eager for me to take a picture of her, while she was polluting a small stream with blue dye pouring out of a cloth. I observed, moreover, that garbage was thrown in the natural environment. For example during a train trip from Guiyang to Kunming

Pictures 3-9: Shitou village

Source: Author.

(capital of Yunnan province) in winter term 2010, I observed that many people accumulated garbage during the train ride and threw everything out of the window. On other occasions I observed that garbage was not disposed of in landfills. Analyzing people's awareness regarding environmental pollution is an important topic for future studies for Guizhou and for China as a whole.

There was a primary school located at the roadside close to Shitou. As in Qiandongnan Miao-Dong Autonomous Prefecture, many school children in the Shitou area used the major road for commuting to and from school. Vehicles and pedestrians commuting on the same road seems a very dangerous situation.

From Shitou village I walked to the next crossroad to catch the bus to Huang-guoshu Falls. While I was waiting for the bus, I was lucky to be joined by a group of young Chinese tourists from Beijing, who came by taxi from Anshun. Around Huangguoshu Falls, which is the touristic focal point of Guizhou, young women were posing in traditional costumes (pictures 3-10). In my view the costumes were, however, inauthentic, and I was not convinced that the young women truly belonged to local ethnic minority groups. The costumes looked factory-made and not homemade. Both the materials and the workmanship of the costumes looked inauthentic. In the left-hand picture it is, moreover, visible that one woman is wearing a jeans jacket on top of her costume, while in the right-hand picture a woman is carrying a modern handbag. Given the high degree of sinicization in the area, the young women could indeed belong to an ethnic minority group, or the job might have been taken over by Han as in TAR, where, HILLMAN (2008, p. 9) observes most of the jobs in the tourism industry are taken by non-Tibetans. Two examples make particularly clear how Tibetans are underrepresented in the tourism industry in Lhasa:

> … [W]hen I visited Lhasa's Potala Palace a few years ago, I was surprised to find a young Han Chinese man dressed in Tibetan costume selling tickets. When I queried him, he laughed and said, "tourists don't know the difference anyway." In another market, a Han Chinese woman passing off wheat flour pancakes as Tibetan barley cakes gave a similar response to my queries. Tourists mightn't know the difference, but Tibetans do, and daily experiences like these are sources of a deep and growing resentment. (HILLMAN, 2008, p. 9)

Pictures 3-10: Ethnic tourism at Huangguoshu Falls

Source: Author.

Back in Anshun I recognized women with particular headdresses (pictures 3-11). I asked some of the women to which ethnic minority group they belonged, and they all informed me that they were Han. This aroused suspicion and I asked in a travel agency about these women. The travel agent informed me that these women are indeed Han and that their ancestors migrated during Ming Dynasty (1368-1644) to today's Anshun. Since then they have been living for many generations in Tunpu and surroundings. In the travel agency I received a tourist map of the Anshun area and information about the Tunpu garrison fortress, which is located northeast of Anshun. In the area Han, Bouyei and Miao live together, but each group lives in separate villages. I visited the Tunpu Culture Museum and a Tunpu village named Ben Zhai. The museum presents information about the migration history and settlement of Han soldiers in the Anshun area. One information panel in the museum says that: "… To the native born nationality, they [the Tunpu people] are conquerors and occupiers; to the Han people…they are pioneers and exploiters" (FN, 2010, p. 40). Few scholars have studied the Tunpu people so far. It would be interesting to analyze the relationship between Han of Tunpu and the surrounding ethnic minority villages to find out whether ethnic minorities still consider the Tunpu as conquerors and occupiers and whether there are occupational differences.

A traditional village of the Tunpu settlers is Ben Zhai. In the fields around Ben Zhai, machinery is used to plow the fields; I did not observe anyone using a water buffalo to plow the fields. Pictures 3-12 show a peasant plowing a field with a gas-powered plow on the left and the entrance to Ben Zhai village on the right. The fortress offers jobs in the tourism industry. When I visited the museum, I was lucky that a group of Chinese tourists were also visiting the fortress. As part of their guided tour, Tunpu people performed the traditional Tunpu Dixi, an ancient opera (pictures 3-13). As the performance is part of the Tunpu garrison culture, the performers were possibly Han Chinese.

The next area I visited was Ziyun Bouyei-Miao Autonomous County. This time I travelled by public bus from Anshun to Ziyun, which took around 2.5 hours.

Pictures 3-11: Tunpu women

Source: Author.

Pictures 3-12: BenZhai village


Source: Author.

Pictures 3-13: Dixi opera


Source: Author.

In the suburbs of Anshun I noticed many women wearing traditional Tunpu headdresses as shown in pictures 3-11. On the way to Ziyun and also within Ziyun, I saw a lot of construction work and people were cutting stones and selling gravel along the roadside. Traditional agricultural activities were also visible in the surrounding fields. I saw wheat, corn, peach trees and tea hills as well as greenhouses on the way to Ziyun. Without access to the greenhouses I could not, however, identify the plants growing there.

Pictures 3-14: Ziyun county

Source: Author.

On the way to and from Ziyun, I again noticed road safety issues; many people were walking along the main road, among them many children, who commute to and from school (left-hand picture 3-14). In Ziyun aside from a lot of construction work there was not much to observe. I saw that women were working in the construction sector, yet whether or not this was actually community work rather than an income-generating activity is unknown (right-hand picture 3-14). Although the women wore headdresses, I could not accurately determine their ethnic status.

During lunch the waiter and his wife, who were both Miao and spoke fluent Mandarin, confirmed my impression. They informed me that there was not much to see in Ziyun because it is a backward town. At the entrance of Ziyun, there was a sign announcing a World Bank poverty reduction program in Southwest China.

The next area I visited was Guanling Bouyei-Miao Autonomous County. I travelled by public bus from Anshun to Guanling, the county capital. Along the roadside to Guanling, there were many small stores and businesses, for example, a hairdresser, a Sinopec gas station, a post office, some grocery stores and car repair shops (pictures 3-15). Some vendors sold tofu and others, sugar cane. The upper left picture 3-15 shows a hairdresser and the upper right picture a grocery store. The lower left picture shows a repair shop and the lower right picture a sugar cane vendor. I could not identify ethnic differences of the workers with observation techniques. At the time of my visit, there were not many clients using the services.

Guanling, however, seemed to be a very developed city with subsidiaries of China Tobacco and China Southern Power Grid. It was, moreover, striking to see that on the way to Guanling the characters of Guanling were carved Hollywoodstyle on a mountain slope.

Pictures 3-15: Guanling county

Source: Author.

3.2.4 Qianxinan Bouyei-Miao Autonomous Prefecture

I travelled from Anshun to Qianxinan Bouyei-Miao Autonomous Prefecture. There, based on a description of a travel guide dated 2005, I expected to encounter an underdeveloped area with poor roads. Instead the roads in the prefecture were developed, and Xingyi was a booming town, which means that during the last five years tremendous effort was put into infrastructure development.

In the hotel in Xingyi, I was provided with a special tourist map of the prefecture in Chinese and English, which was very useful for locating the four accessible ethnic minority villages: Nanlong Bouyei Minority Village, Liyuba Miao Minority Village, Nakong Bouyei Minority Village and Nachan Bouyei Minority Village. Nanlong Bouyei Minority Village was not accessible by public transport, so I had to take a taxi to get there. The closer we came to the village, the simpler were the roads. On the way there were several other villages with people engaging in agricultural and other domestic activities. As the traditional method of wheat threshing is labor intensive, some peasants were putting the wheat on the street, thus using the power of cars to separate the wheat from the chaff. The left-hand picture 3-16 shows this approach, while the picture on the right shows a woman cutting wheat in the field.

Pictures 3-16: Wheat harvest

Source: Author.

The taxi driver informed me that during rainy season (May to October) peasants of the area cultivate rice and that during dry season (October to April) they cultivate wheat. In most of the villages we passed through on the way to Nanlong, Bouyei and Han people lived together in mixed communities. I was unable to distinguish the people by their ethnic status, yet the taxi driver could. He was Han and had grown up in one of the villages of this region. He distinguished Han and Bouyei mainly on the basis of their accent and facial features. Sometimes he could even distinguish them by looking only at their clothes. For example he could identify the women in the left hand side of pictures 3-17.

He deduced from their facial features and clothes that the woman in the middle was Han and the women on either side were both Bouyei. A local family informed me that Nanlong has many visitors from China and abroad; they even provided me with a leaflet in Mandarin which showed a group of Chinese visitors and introduced the village culture, architecture and natural environment (pictures 3-17 on the right). During my visit in Nanlong, I was, however, the only tourist in the village,

Pictures 3-17: Bouyei and Han women, Nanlong village

Source: Author.

which indicated that, as in other villages, tourism in Nanlong was more a side-line for festivals and holiday season than a full-time employment.

The older man who gave me the leaflet was the father of a former classmate of the taxi driver, with whom the taxi driver was in middle school and with whom he lived together in the same dormitory. Over the years they had lost contact and had not met for around 20 years. The family was extremely happy that the driver was concerned about their son and invited us to stay for lunch. The son now lives in Xingyi and already has two children. The parents took around two hours to prepare the meal. They slaughtered a chicken and offered smoked pork meat along with vegetables and rice. It was very interesting to see their traditional way of cooking.

In the meantime another woman showed her handmade Bouyei style blankets. She and the driver were discussing the fact that nowadays very few Bouyei women actually make clothes themselves, but rather buy them on the market, which she and the driver do not think were as beautiful as the homemade ones. While we were waiting for the food, the father offered the driver a cigarette. He refused the cigarette, although he was a smoker, which I had noticed during the trip to the village. I asked him on our way back to Xingyi why he had refused the cigarette, which I thought may be an insult to the host, but he replied that he did not want to take from the poor old man's limited supply of cigarettes. This behavior may indicate that the driver saw the villagers as poorer and his income in the city as comparatively higher.

The next day I visited Liyuba Miao Minority Village. I went first to the city Xingren, where I took a taxi to reach the village. The driver was a male Han around 40 years of age. Liyuba is a Miao village and, according to the driver, was newly constructed five years ago. The driver told me that the old houses had been in decay and were rebuilt. At that time the local government probably saw a source to attract tourists to the village and, thus, encouraged the rebuilding of the village in a traditional Miao style rather than in a more modern style seen in Paizhao village. The Miao houses are shown in pictures 3-18. The houses have symbolic cow horns at all four corners and crossed cow horns in the middle of the roof. These horns have a symbolic meaning for the Miao people; this symbolism would be interesting to analyze in future research.

Agriculture was done using traditional methods as in all the other areas rather than Ben Zhai. There was one grocery store located in the village. The woman who sold groceries was wearing a traditional Miao headdress, so that she would be classified as a Miao woman working in NA. Pictures 3-19 show the woman selling groceries to villagers on the right and an elderly Miao woman of the village on the left. According to the Han classification, the elderly woman belongs to the Black Miao.

Pictures 3-18: Liyuba village

Source: Author.

Pictures 3-19: Miao women of Liyuba village

Source: Author.

It is typical for Miao villages to have a public square in the middle of the village, where all festivals take place. SCHEIN (2000, p. 192) notes that after decollectivization, modernization has become a huge concern of Miao people, to the extent that they would even give up those ancient customs that are not profitable. As festivals are touristic attractions, it would be interesting to learn whether or not the festivals would continue under people's own instigation in absence of government support.

Pictures 3-20: Bouyei women

Source: Author.

To visit the last two villages in the prefecture, I went from Xingren to Zhengfen to visit Nakong Bouyei Minority Village and Nachan Bouyei Minority Village. From Xingren to Zhengfen I took a public bus and went by taxi to the villages.

Around Zhengfen there were many Miao women wearing the same traditional clothing as Miao women in Liyuba village. The driver said that Miao people here live in mixed communities with Bouyei and Han. Nakong and Nachan were, however, peopled by Bouyei. The driver knew the way from Zhengfen to Nakong immediately, but he did not know the way to Nachan village. In Nakong we had to ask many inhabitants about the way to Nachan village until finally someone knew and explained the way. This indicated that Nachan village was a little visited tourist site. In both villages there was not an agglomeration of houses, but houses were spread along several small roads. Some houses were traditional, but most were modern. The major observable jobs in the area were in agriculture and in the construction sector. Agriculture was mainly done using traditional methods. I saw, moreover, some males using mechanized hand plows as shown in pictures 3-12.

In both villages many women wore traditional Bouyei clothes (pictures 3-20). The driver informed me that Bouyei women usually start wearing traditional clothes after marriage. Many young unmarried women, therefore, do not wear traditional clothing. In future studies it would be interesting to investigate cultural identities of young unmarried and married women in this area. It is unclear whether these customs will endure, so it might be interesting to conduct a longitudinal study and revisit the issue periodically over the next ten or twenty years.

3.3 Conclusions

In this chapter I have shared my experience as well as some of the evidence gathered during my fieldwork in Guizhou. The fieldwork was conducted in two autonomous prefectures and three autonomous counties of Bouyei and Miao groups, from March 27 to April 29, 2010. I focused on analyzing ethnic differences in

agricultural and non-agricultural employment. Additional knowledge was acquired during winter term 2010 (August 24 to December 23), while I was studying Chinese language at Guizhou University.

The major difficulty was to correctly identify ethnic status just by observing individuals. I could only identify ethnic minorities based on their traditional clothing, which on an everyday basis was observable for women, but not for men. Most of the Miao women kept their traditional clothing in all of the areas observed, while for the Bouyei only those women in the villages Nakong and Nachan located in Qianxinan Bouyei-Miao Autonomous Prefecture wore traditional clothing. Although the Miao are classified as one ethnic minority group, traditional clothing, local accents, dialects, languages and other particularities changed from one village to the next. Locals were also able to distinguish people of various ethnic groups by their facial features. Whether or not differences in traditional awareness are directly reflected in occupational distribution was unobservable.

In most of the areas the remoteness of villages and available natural resources determine occupational distributions and employment possibilities. These geographic factors can be covered with county dummies and village size information in the occupational outcomes models in chapter four. In the areas considered agricultural production was done using traditional methods due to the mountainous topography of the province. In all the areas I visited in Guizhou, the plowing of fields with water buffalos was solely done by men. The available non-agricultural occupations are mainly in the construction and in the service sectors and their corresponding subcategories. I could not observe major ethnic distinctions in the two sectors. Tourism as a major subcategory of the service sector is particularly important for ethnic minorities in Guizhou, as their traditional lifestyle is the target of governmental support. Since the economic opening of China, job possibilities in tourism and *dagong* in Guangdong have increased peasants' income, and some Bouyei and Miao households have been able to build new houses in their home villages from this income.

The educational levels of individuals are impossible to capture with observation techniques. I saw that there are many primary and middle schools distributed all over the province. I find that Bouyei and Miao with university degrees are able to work in non-manual jobs as university teachers or tourist guides. Han see preferential policies towards ethnic minorities in accessing university education as unfair because ethnic minorities are direct competitors in the entrance examinations.

4 QUANTITATIVE APPROACH AND EVIDENCE

The hypotheses developed in subchapter 2.5.1.2 are tested econometrically in this chapter. Subchapter 4.1 describes the database. Subchapter 4.2 provides information about the dependent and independent variables. Subchapter 4.3 gives descriptive statistics. Subchapter 4.4 provides the formal econometric model specifications and gives estimation results. Subchapter 4.5 concludes and shows developments between the years considered in my study.

4.1 The database

This subchapter gives a detailed description of the database I use to quantitatively test my research hypotheses. First, I give a general description of the China Health and Nutrition Survey (CHNS) database. Second, I describe the reduced dataset and the dependent and independent variables.

4.1.1 The CHNS database

I use the CHNS database[21] for the empirical analysis. The CHNS has been conducted by 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 since 1989. The database covers the years 1989, 1991, 1993, 1997, 2000, 2004 and 2006.[22] The CHNS uses a multistage, random cluster process to draw a sample of about 4,400 households with a total of 19,000 individuals in nine provinces. These nine provinces are in alphabetical order: Guangxi Zhuang Autonomous Region, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning and Shandong. The CHNS stratifies counties in each of the nine provinces by income (low-, middle- and high-income tertiles) based on per capita income figures from the State Statistical Office. The CHNS uses a weighted sampling scheme to randomly select four counties in each province (one low income, two middle income and one high income) and selects the provincial capital and lower income cities when it was possible. The CHNS randomly selects villages and townships within the counties, urban and suburban neighborhoods within the cities.

I consider answers from the adult, the household and the community questionnaires in my analysis. The original questionnaires for all surveys (1989-2006) are available

---

[21] The CHNS is available in the public domain HTTPS://WWW.CPC.UNC.EDU/PROJECTS/CHINA.
[22] The most recent results from 2009 will be available in the public domain shortly.

in Chinese and English in the public domain (HTTP://WWW.CPC.UNC.EDU/PROJECTS/CHINA/DATA/QUESTIONNAIRES).

4.1.2 The data used for this study

Given my topic of interest and the chosen methodology, I reduced the above mentioned sample size based on four major restrictions:

1) I focus solely on Guizhou. In the questionnaire the observation *T1* indicates the province. I, thus, restrict the sample to observations in Guizhou which is coded as (*T1=52*) in the dataset.
2) I only consider rural areas. A distinction between rural and urban areas is given with variable *T2* in the original dataset. The sample is, thus, restricted to observations from rural sites, which means to observations when (*T2=2*).
3) I only consider comparisons of the data in a cross-sectional structure for the years 2004, 2000 and 1997 rather than a panel-structure. The years 1993, 1991 and 1989 are not considered in the analysis as I am interested in more recent developments; moreover, there are some data inconsistencies for the years 1991 and 1989. The years are given by the variable *wave*. The community survey for the year 2006 is not available, so that I do not consider this year in the analysis. For the other years our research group signed a user agreement with CHNS to receive and use the community datasets. Although usually the same individuals take part in the survey, I do not consider a panel structure of the data because changes in employment from the *A* to the *NA* sectors cannot accurately be measured as annual data is not available. This means that survival analysis cannot be applied because assumptions about occupational outcomes for the unobserved years would be required; because this is rather arbitrary, I do not use this approach.
4) Individuals with missing observations are not considered in the analysis.

Regarding the data sample three drawbacks have to be kept in mind when interpreting and generalizing results. First, the random selection of individuals is not based on the ethnic composition of the provinces. To control for selection bias in the empirical analysis, standard errors that are robust to a clustered sample design are calculated (DEATON, 1997, pp. 73-78).[23] With this control I make the assumption that error terms are correlated within the communities. The unobserved utility of individuals living in the same community is, therefore, assumed to be correlated. Second, only communities where a mixed ethnic population is living are considered in the analysis. This is done to have the same environmental conditions for all ethnic groups and to disentangle ethnic and community effects. Third, families that migrate from one community to another are not followed in the survey (HTTP://WWW.CPC.UNC.EDU/PROJECTS/CHINA/PROJ_DESC/SURVEY). Shufa Du, a

[23] This was implemented by using Stata's "vce(cluster)" option.

research assistant professor from CHNS, moreover, informed me that CHNS scholars were not able to collect data from participants who were not at home during the data collection period (usually five to seven days in a village) because scholars needed to collect three-day diet data and measure height, weight and blood pressure of the participants. When survey participants moved out of the survey communities, they were no longer considered in the survey because of funding shortages. The reduced data sample, therefore, contains only individuals who live and work in rural communities, which result in an overrepresentation of individuals working in agriculture. If all household members despite their residence in other areas were considered in the analysis, the share of individuals working in *NA* would be comparatively higher.

4.2 The variables

In section 2.4.1.1 I pointed out the difficulty of correctly specifying dependent and independent variables in discrete choice models. This subchapter shows how I selected these variables.

4.2.1 Dependent variables

The dependent variable considers individuals' primary and secondary occupations. Both primary and secondary occupations are self-reported and contain 13 main categories with several subgroups in the original questionnaires. I use the variables *B4* and *B9* from the CHNS. The 13 main categories and corresponding subgroups for both primary and secondary occupations are (*1*) senior professional/technical worker (doctor, professor, lawyer, architect, engineer), (*2*) junior professional/ technical worker (midwife, nurse, teacher, editor, photo-grapher), (*3*) administrator/ executive/manager (working proprietor, government official, section chief, department or bureau director, administrative cadre, village leader), (*4*) office staff (secretary, office helper), (*5*) farmer, fisherman, hunter, (*6*) skilled worker (foreman, group leader, craftsman), (*7*) unskilled worker (ordinary laborer, logger), (*8*) army officer, police officer, (*9*) ordinary soldier, policeman, (*10*) driver, (*11*) service worker (housekeeper, cook, waiter, doorman, hairdresser, salesperson, launderer, child-care worker), (*12*) athlete, actor, musician, (*13*) other, (-*9*) unknown. Individuals who chose other or unknown occupations, categories (*13*) and (-*9*), respectively, are excluded from the analysis.

The first step is combining the remaining categories in a way that the data can be used in discrete choice analysis. I already pointed out that there are many different ways to form occupational categories (see table 2.3). The categories can be formed based on statistical and/or theoretical reasoning. Based on my theoretical discussion and hypotheses, I focus on sectors *A* and *NA*. Individuals working in sector *A* can easily be identified as any individual who is a farmer, a fisherman or hunter is in category (*5*). All individuals who are working in other occupations except category (*5*) can be classified as workers of the *NA* sector.

I am also interested looking at subgroups within *A* and *NA* sectors. Within sector *A* some individuals, for example, have a secondary occupation (*soc*). In field observations (chapter three) I found that many ethnic minorities working in *A* also devote some time to touristic activities, mainly when it is demanded by visitors to the village (e.g., for special festivals or conferences). A drawback of the CHNS dataset is that tourism is not a job category. It can, however, be assumed that with increasing tourism, the demand for other occupations such as in the service sector also increases. The category *A* is, therefore, divided into $A_{Primary}$ and $A_{Primary+soc.}$ $A_{Primary}$ includes those individuals who have agriculture as their primary occupation without a *soc*. $A_{Primary+soc}$ includes those individuals who work primarily in *A* and also reported to have a *soc*.

In the *NA* sector occupations can be divided most appropriately into blue-collar (*BC*) and white-collar (*WC*) positions. *BC* workers are manual laborers and *WC* workers are professionals and/or educated workers. From the original categories given above, the combined categories (*7, 9, 10 and 11*) form the *BC* group and the combined categories (*1, 2, 3, 4, 6 and 8*) form the *WC* group. Some of the individuals who work primarily in *NA* employment often have agriculture as a secondary occupation; however, I do not further distinguish between *NA* and *NA* plus *soc* because I am primarily interested in observing *soc* alongside *A*. Table 4.1 and figure 4-1 show the occupational outcomes for the years 1997, 2000 and 2004.

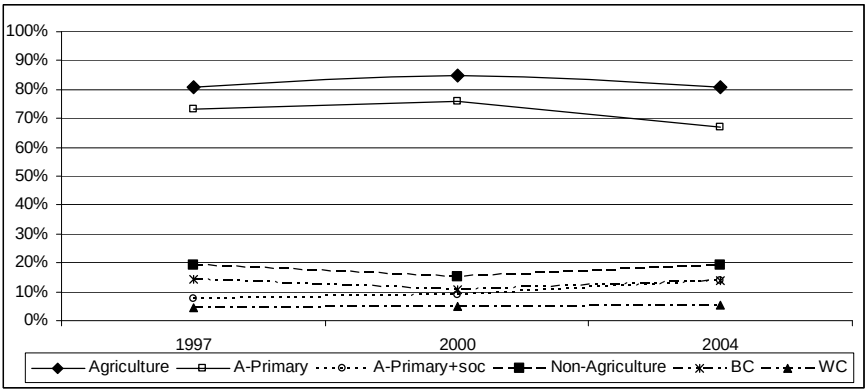
Table 4.1 shows that in all years (1997, 2000 and 2004) the number of individuals working only in $A_{Primary}$ is around 73 %, 76 % and 67 %, respectively; therefore, $A_{Primary}$ is the outcome with the highest number of workers in the samples considered (see figure 4-1). The comparison of 1997 and 2000 shows that the number of workers in $A_{Primary}$ increases by around three percentage points from around 73 % in 1997 to around 76 % in 2000. While in 2000 $A_{Primary}$ is around 76 %, it decreases by roughly nine percentage points in a period of four years to around 67 % in 2004. The number of workers in $A_{Primary+soc}$ increases from 1997 to 2004. While in 1997 only around 8 % of the considered individuals have a *soc*, in 2000 and 2004 the number increases to roughly 9 % and 14 %, respectively. This means that the number of individuals with a *soc* increases by around six percentage points in the time period considered.

Table 4-1: Frequency of occupational outcomes by year

Outcomes/Years	**1997 *N* (%)**	**2000 *N* (%)**	**2004 *N* (%)**
Agriculture	485 (80.97)	376 (84.69)	329 (80.84)
$A_{Primary}$	439 (73.29)	336 (75.68)	273 (67.08)
$A_{Primary+soc}$	46 (7.68)	40 (9.01)	56 (13.76)
Non-Agriculture	114 (19.03)	68 (15.32)	78 (19.17)
BC	86 (14.36)	47 (10.59)	56 (13.76)
WC	28 (4.67)	21 (4.73)	22 (5.41)
N	599 (100)	444 (100)	407 (100)

Source: Author's calculation based on CHNS sample.

Figure 4-1: Frequency of occupational outcomes by year



Source: Author based on CHNS sample.

Table 4.1 also shows the number of workers in the *NA* sector. For the *BC* sector there is no clear pattern (see figure 4-1). The highest number of workers in the *BC* sector with around 14 % is in the 1997 sample. The number decreases by around three percentage points to around 11 % in 2000. From 2000 to 2004 the share then increases again by around three percentage points to around 14 %. In the *WC* sector a slight upward trend is observable, yet it remains at around 5 % in all three years.

4.2.2 Independent variables

As shown in figure 2-5 and described in section 2.5.1.2, the core factors which influence occupational outcomes are ethnic status, years of education, gender, age and geographic location. Ethnic status is self-reported, and the original variable in the household survey is *nationality*. The subcategories of the variable in the samples considered include observations for Han, the Bouyei, the Miao and the Tujia groups. Likelihood ratio (LR) tests, AIC and BIC criteria[24] show that the inclusion of all ethnic groups leads to better model fit rather than does the use of only an ethnic dummy or the omission of ethnic differences. In this way dummy variables for the ethnic groups (Han – base, Miao, Bouyei and Tujia) are used in the analysis.

Human capital factors are included in the adult questionnaire. Years of education are calculated on the basis of variable *A11*, which gives information about how many years of formal education the respondent has completed in a regular school.

[24] The interpretation of the test results is straightforward. A comparatively lower value of AIC and BIC indicates a better model fit. The LR-test rejects the full model if the p-value is larger than the threshold value of 0.05. The results of all tests are available upon request.

Gender is represented by the variable *AA2a*, where *1* stands for male and *2* for female. I use the dummy form and name the gender variable male, which is *1* when the respondent is male and *0* when the respondent is female. Age in years is calculated based on the year of observation (*wave*) and the western date of birth (*AA3a*) of the respondent. I use this approach because the original dataset is missing many observations under the age category (*A3a*). With the variable age I also capture an individual's experience. I do not use an additional variable for experience, which is usually (age minus years of education minus five), because of the high degree of collinearity between age and experience. LR tests, AIC and BIC criteria suggest that the inclusion of years of education, gender and age all increase model fit. On theoretical as well as on statistical grounds, I, therefore, consider years of education, gender and age in the analysis.

To control for geographic location, I use county dummies and village size. In every province four counties are randomly selected and assigned the variable *T3*. Based on the fact that I only consider mixed ethnic communities, the first county will be omitted from the analysis as only Han are living there. In this way counties 2, 3 and 4 are considered. In the empirical analysis I use the values from 1 to 3 (county1 – base, county2 and county3). The community dataset provides information about the number of households in the village/neighborhood, variable *o0a*. I use this variable to capture the size of the village and to capture related occupational outcomes. I simply assume that larger villages offer a greater variety of occupations than smaller villages. Larger villages are, moreover, better connected to more developed regions than smaller villages. I found during fieldwork that aside from the cities and townships, larger villages were also connected with paved roads and public buses. In contrast smaller, remoter villages lacked this infrastructure. Another possible control variable for geographic location would be to consider the distance from the village to the next urban center. This information is, however, not captured with the available variables of the CHNS. Data which provide the exact GIS coordinates of villages can be purchased from CHNS; however, a lack of funding precluded this purchase.

In the literature other sub-relationships are also considered by some authors (see figure 2-5). I included, for example, the marital status of the individual to capture household characteristics. The overall model fit, however, deteriorated when this additional control was included; therefore, I did not consider marital status in the analysis. For other possible variables such as parental occupation and parental education, there were not enough available observations. ZANG (2008) finds that CCP membership is also an important factor which facilitates access to some occupations within the *NA* sector. In the CHNS dataset the variable *A15* indicates party membership. Unfortunately this information is not reported in the years 2004 and 2006. The main reason I did not consider this variable is, however, that in job category (*3*) administrative cadre is one subgroup. There is, therefore,

a perfect correlation between the variable cadre and job category (*3*); in the analysis cadre is considered a subcategory of non-agriculture.

The independent variables are, thus, ethnic status (Han – base, Bouyei, Miao, and Tujia), education (years), age (years), male (0/1), counties (county1 – base, county2 and county3) and number of households in the village. Based on the fact that additional education improves job access to the *NA* sector, I include interaction terms between ethnic status and years of education in the analysis. In this way it is possible to distinguish individuals not only by their ethnic status but also by their education.

4.3 Descriptive analysis for 2004, 2000 and 1997

The previous subchapter explained the dependent and independent variables; this chapter presents descriptive statistics of these variables in a cross-sectional approach for 2004, 2000 and 1997. I report the average values and statistical significance of the independent variables separately for each ethnic group and for the overall samples; moreover, I show distributions of ethnic groups by counties, occupational outcomes by counties and occupational outcomes by ethnic groups for each year under consideration.

4.3.1 Descriptive statistics 2004

Table 4.2 shows average values of the independent variables years of education, percentage of males and age for each ethnic group and for the overall sample of the year 2004. Altogether the sample includes 407 individuals. At 44.23 % the population share of the Bouyei is the highest, followed by Han at 22.6 %, the Tujia at 20.15 % and the Miao at 13.02 %.

The average education of this sample is 5.6 years. There is no statistically significant difference in education between Han and the ethnic minority groups Miao, Bouyei and Tujia. The average proportion of males in this sample is 52.8 %. I also find no statistically significant difference in this proportion between Han and ethnic minorities. The average age level in this sample is 46.2 years. There is a statistically significant age difference between Han and Tujia at the 10 % significance level. On average the Tujia are 4.1 years older than Han in this sample.

Table 4-2: Descriptive statistics by ethnicity 2004

Variables	Han	Miao	Bouyei	Tujia	Sample
Education mean (SD)	5.5 (3.9)	5.4 (3.9)	5.8 (3.7)	5.3 (3.9)	5.6 (3.8)
Male *%*	51.09	47.17	52.22	59.76	52.83
Age mean (SD)	45.2 (16.8)	48.3 (12.6)	44.6 (14.2)	49.3 (13.2)	46.2 (14.5)
N	92 (22.6)	53 (13.02)	180 (44.23)	82 (20.15)	407

Source: Author's calculation based on CHNS sample.

Table 4-3: Ethnic proportion in counties 2004

Variables	Communities	Han *(N, %)*	Miao *(N, %)*	Bouyei *(N, %)*	Tujia *(N, %)*	Sample *(N, %)*
County 1	522102	54 (43.2)	22 (17.6)	49 (39.2)	0 (0)	125 (30.71)
	522103					
	522104					
County 2	522201	14 (11.97)	21 (17.95)	0 (0)	82 (70.1)	117 (28.75)
	522202					
	522203					
	522204					
County 3	522301	24 (14.55)	10 (6.06)	131 (79.4)	0 (0)	165 (40.54)
	522302					
	522303					
	522304					
N		92 (22.6)	53 (13.02)	180 (44.23)	82 (20.15)	407 (100)

Communities 522101, 522401, 522402, 522403, and 522404 are not considered as only Han are living there.

Source: Author's calculation based on CHNS sample.

Table 4.3 shows the proportion of each ethnic group in the three counties. The second column from the left gives the community number of the CHNS dataset. The other columns to the right show the proportions of Han, Miao, Bouyei, Tujia and the overall sample for the three counties, respectively.

The 2004 sample includes 407 individuals; 30.71 % are living in county 1, 28.75 % in county 2 and 40.54 % in county 3. Han and the Miao reside in all three counties, the Bouyei only reside in county 1 and 3, and the Tujia reside only in county 2. This distribution fits with Guizhou's ethnic distribution. It particularly shows how unequally the Bouyei and Tujia groups are distributed in the province. The unequal ethnic distribution has to be kept in mind for the analysis of county dummies; significant county effects may have stronger effects on local ethnic groups than on other ethnic groups in Guizhou.

At 43.2 % the population share of Han is the highest in county 1, followed by the Bouyei at 39.2 % and the Miao at 17.6 %. At 70.1 % the population share of the Tujia is the highest in county 2, followed by the Miao at 17.95 % and Han at 11.97 %. At 79.4 % the population share of the Bouyei is the highest in county 3, followed by Han at 14.55 % and the Miao at 6.06 %.

Table 4.4 and 4.5 show the occupational outcomes in the counties. Table 4.4 distinguishes broadly between sectors *A* and *NA*. Table 4.5 shows occupational outcomes of sectors *A* and *NA* and their subcategories, in sector *A* the subcategory $A_{Primary+soc}$ and in sector *NA* the subcategories *BC* and *WC*. Table 4.4 shows that in all three counties the majority of individuals are working in *A*. The highest share is in county 1 with 92.8 %, followed by 79.4 % in county 3 and 70.1 % in county 2. The share of individuals working in *NA* is correspondingly the highest in

Table 4-4: Agriculture versus non-agriculture in counties 2004

Variables	Agriculture ($A_{Primary}$, $A_{Primary+soc}$) (N, %)	Non-Agriculture (BC, WC) (N, %)	Sample (N, %)
County 1	116 (92.8)	9 (7.2)	125 (100)
County 2	82 (70.1)	35 (29.9)	117 (100)
County 3	131 (79.4)	34 (20.6)	165 (100)
N	329 (80.8)	78 (19.2)	407 (100)

Source: Author's calculation based on CHNS sample.

Table 4-5: Occupational outcomes by counties 2004

Variables	$A_{Primary}$ (N, %)	$A_{Primary+soc}$ (N, %)	*BC* (N, %)	*WC* (N, %)	Sample (N, %)
County 1	81 (64.8)	35 (28)	7 (5.6)	2 (1.6)	125 (100)
County 2	72 (61.5)	10 (8.6)	28 (23.9)	7 (6)	117 (100)
County 3	120 (72.7)	11 (6.7)	21 (12.7)	13 (7.9)	165 (100)
N	273 (67.1)	56 (13.8)	56 (13.8)	22 (5.4)	407 (100)

Source: Author's calculation based on CHNS sample.

county 2 (29.9 %), followed by county 3 (20.6 %) and county 1 (7.2 %). The introduction of additional categories results in more variation within sectors *A* and *NA* (table 4.5).

Table 4.5 shows that the highest proportion of individuals working in $A_{Primary}$ is in county 3 (72.7 %), followed by county 1 (64.8 %) and county 2 (61.5 %). With the inclusion of the category $A_{Primayr+soc}$, I find a comparatively larger share of individuals with a *soc* in county 1 (28 %). County 2 has the highest share of *BC* jobs (24 %), followed by county 3 and county 1. Table 4.5, moreover, shows that the number of *WC* workers is the highest in county 3 (7.9 %) followed by county 2 (6 %) and county 1 (1.6 %). The inclusion of the additional subcategories makes the empirical analysis more accurate and will, therefore, be considered in the estimations.

Table 4.6 shows the occupational outcomes by ethnicity. Most of the individuals who work in $A_{Primary}$ are Bouyei (50.9 %), followed by Han (19.8 %), Tujia (18 %) and Miao (11.4 %). Most of the individuals who work in $A_{Primary+soc}$ are also Bouyei (35.7 %), followed by Miao (30.4 %), Han (21.4 %) and Tujia (12.5 %). In the *BC* sector I find that, with 37.5 %, Han and the Tujia share the position of having the highest share, followed by the Bouyei and the Miao with 21.4 % and 3.6 %, respectively. In the *WC* sector the Bouyei have the highest share (40.9 %), followed by Han and the Tujia with 22.7 % each and the Miao with 13.6 %.

Table 4-6: Occupational outcomes by ethnicity 2004

Variables	Han	Miao	Bouyei	Tujia	Sample
	(N, %)	*(N, %)*	*(N, %)*	*(N, %)*	*(N, %)*
$A_{Primary}$	54 (19.8)	31 (11.4)	139 (50.9)	49 (18)	273 (100)
$A_{Primary+soc}$	12 (21.4)	17 (30.4)	20 (35.7)	7 (12.5)	56 (100)
BC	21 (37.5)	2 (3.6)	12 (21.4)	21 (37.5)	56 (100)
WC	5 (22.7)	3 (13.6)	9 (40.9)	5 (22.7)	22 (100)
N	92 (22.6)	53 (13)	180 (44.2)	82 (20.2)	407 (100)

Source: Author's calculation based on CHNS sample.

4.3.2 Descriptive statistics 2000

Table 4.7 shows the average values of the independent variables years of education, age and percentage of males for each ethnic group and the overall sample in 2000. The sample includes 444 individuals; there are 37 more observations in 2000 than in 2004. With 41.4 % the Bouyei hold the highest share of the population, followed by the Tujia with 23.4 %, Han with 20.5 % and the Miao with 14.6 %. Between 2000 and 2004 the share of Han and the Bouyei increases by 2.1 and 2.9 percentage points, respectively, and the share of the Miao and Tujia decreases by 1.62 and 3.27 percentage points, respectively.

In 2000 the average education is 5.2 years. In 2004 individuals had 0.4 more years of education than they had been in 2000. In terms of education, in 2000 there is also no statistically significant difference between Han, the Miao, the Bouyei and the Tujia. In 2000 the average proportion of males is 51.8 %, and the average age level is 42.9 years. There is also no statistically significant difference in the proportion of males and age between Han and ethnic minorities.

Table 4.8 shows the ethnic proportion in 2000 in the three counties considered. The table is organized in the same way as table 4.3. In table 4.8 numerical changes in comparison to 2004 are additionally provided in italics below the 2000 values. The general ethnic representation in the counties is the same in 2000 and 2004. There are some changes in the number of observations between 2000 and 2004, which may indicate that in 2000 more household members were at home during the survey than in 2004.

Table 4-7: Descriptive statistics by ethnicity 2000

Variables	Han	Miao	Bouyei	Tujia	Sample
Education mean (SD)	5.5 (4.2)	5.5 (3.9)	5.3 (3.7)	4.5 (4.2)	5.2 (4)
Male *%*	56.04	52.31	50	50.96	51.8
Age mean (SD)	42.4 (16.1)	43.6 (14.1)	41.5 (14.6)	45.4 (15.3)	42.9 (15.1)
N	91 (20.5)	65 (14.6)	184 (41.4)	104 (23.4)	444

Source: Author's calculation based on CHNS sample.

Table 4-8: Ethnic proportion in counties 2000

Variables	Communities	Han *(N, %)*	Miao *(N, %)*	Bouyei *(N, %)*	Tujia *(N, %)*	Sample *(N, %)*
County 1	522102	54 (46.2)	30 (25.6)	33 (28.2)	0 (0)	117 (26.4)
	522103		*-8*	*16*		*8*
	522104					
County 2	522201	10 (7.4)	22 (16.2)	0 (0)	104 (76.5)	136 (30.6)
	522202	*4*	*-1*		*-22*	*-19*
	522203					
	522204					
County 3	522301	27 (14.14)	13 (6.8)	151 (79.1)	0 (0)	165 (43)
	522302	*-3*	*-3*	*-20*		*-26*
	522303					
	522304					
N		91 (22.6)	65 (14.6)	184 (41.4)	104 (23.42)	444 (100)
		1	*-12*	*-4*	*-22*	*-37*

The communities 522101, 522401, 522402, 522403, and 522404 are not considered as only Han are living there.

Source: Author's calculation based on CHNS sample.

In county 1 Han and the Tujia shares remain unchanged, the share of the Miao decreases between 2000 and 2004 by eight observations, and the share of the Bouyei increases by sixteen observations. In county 2 the proportions of Han, the Miao and the Tujia change between 2000 and 2004, the Han increase by four observations, the Miao and the Tujia decrease by one and twenty-two observations, respectively. In county 3 the shares of Han, the Miao and the Bouyei decrease between 2000 and 2004 by three, three and twenty observations, respectively. The overall sample decreases by 37 observations between 2000 and 2004. The exact reasons for the changes in observations are not reported in the CHNS.

Tables 4.9 and 4.10 show the occupational outcomes for 2000. Table 4.9 distinguishes broadly between sectors *A* and *NA*, and table 4.10 additionally considers the subcategories. The changes in the observations between 2000 and 2004 are given in italics. In 2000 most of the people are also working in *A*, with shares of 98.3 %, 80.9 % and 79.1 % in counties 1, 2 and 3, respectively. The comparison of 2000 and 2004 shows that observations in counties 2 and 3 decrease by 28 and 20 observations, respectively. This indicates that in counties 2 and 3 a greater share of individuals left sector A between 2000 and 2004. Considering the whole sample, there are 47 fewer individuals working in *A* in 2004 than in 2000. In the *NA* sector the corresponding shares are 1.7 %, 19.12 % and 20.9 % in counties 1, 2 and 3, respectively. In 2004 ten additional individuals are working in *NA* than in 2000. Between 2000 and 2004 the number of individuals who work in *NA* increases in county 1 and 2, and decreases in county 3.

Table 4-9: Agriculture versus non-agriculture in counties 2000

Variables	Agriculture *(A, A+soc)* *(N, %)*	Non-Agriculture *(BC + WC)* *(N, %)*	Sample *(N, %)*
County 1	115 (98.3) *1*	2 (1.7) *7*	117 (100) *8*
County 2	110 (80.9) *-28*	26 (19.12) *9*	136 (100) *-19*
County 3	151 (79.1) *-20*	40 (20.9) *-6*	191 (100) *-26*
N	376 (84.7) *-47*	68 (15.3) *+10*	444 (100) *-37*

Source: Author's calculation based on CHNS sample.

Table 4-10: Occupational outcomes in counties 2000

Variables	$A_{Primary}$ *(N, %)*	$A_{Primary+soc}$ *(N, %)*	BC *(N, %)*	WC *(N, %)*	Sample *(N, %)*
County 1	108 (92.3) *-27*	7 (6) *28*	1 (0.9) *6*	1 (0.9) *1*	117 (100) *8*
County 2	98 (72.1) *-26*	12 (8.8) *-2*	19 (14) *9*	7 (5.2) *0*	136 (100) -19
County 3	130 (68.1) *-10*	21 (11) *-10*	27 (14.1) *-6*	13 (6.8) *0*	191 (100) *-26*
N	336 (75.7) *-63*	40 (9) *16*	47 (10.6) *9*	21 (4.7) *1*	444 (100) *-37*

Source: Author's calculation based on CHNS sample.

Table 4.10 considers occupational outcomes including subcategories in 2000. In all three counties the number of individuals in $A_{Primary}$ is higher than in any other occupation, with the share at 92.3 %, 72.1 % and 68.1 % in counties 1, 2, and 3, respectively. In 2000 there are 27 more individuals in county 1, 26 more individuals in county 2 and 10 more individuals in county 3 in sector *A*. The share of individuals who work in sector *A* and have a *soc* is 6 %, 8.8 % and 11 % in counties 1, 2 and 3, respectively. While the observations in counties 2 and 3 decrease by two and ten observations between 2000 and 2004, respectively, there is an increase of twenty-eight observations in county 1 in the same period. Table 4.10, moreover, shows that counties 2 and 3 have shares in *BC* employment of around 14 %, but in county 1 of only 0.9 %. Between 2000 and 2004 there is an increase in *BC* employment by six and nine observations in counties 1 and 2, respectively, and a decline of six observations in county 3. In the *WC* sector the observations remain almost unchanged in 2000 and 2004, with low shares of 0.9 %, 5.2 % and 6.8 % in counties 1, 2 and 3, respectively. For the overall sample the number of individuals working in $A_{Primary}$ declines, while the number of individuals in all other sectors increases between 2000 and 2004. As in 2004 the inclusion of additional subcategories makes the analysis of 2000 more accurate and will, thus, be considered in the estimations.

Table 4-11: Occupational outcomes by ethnicity 2000

Variables	Han *(N, %)*	Miao *(N, %)*	Bouyei *(N, %)*	Tujia *(N, %)*	Sample *(N, %)*
$A_{Primary}$	61 (18.2) *-7*	52 (15.5) *-21*	151 (44.9) *-12*	72 (21.4) *-23*	336 (100) *-63*
$A_{Primary+soc}$	6 (15) *6*	5 (12.5) *12*	20 (50) *0*	9 (22.5) *-2*	40 (100) *16*
BC	17 (36.2) *4*	3 (6.4) *-1*	9 (19.2) *3*	18 (38.3) *3*	47 (100) *9*
WC	7 (33.3) *-2*	5 (23.8) *-2*	4 (19.1) *5*	5 (23.8) *0*	21 (100) *1*
N	91 (20.5) *1*	65 (14.6) *-12*	184 (41.4) *-4*	104 (23.4) *-22*	444 (100) *-37*

Source: Author's calculation based on CHNS sample.

Table 4.11 shows occupational outcomes by ethnicity for 2000. In $A_{Primary}$ the Bouyei have the highest share (44.9 %), followed by the Tujia (21.4 %), Han (18.2 %) and the Miao (15.5 %). In the 2004 sample the Bouyei also have the highest share in $A_{Primary}$ and the Miao the lowest share. The Tujia, however, have a lower share than Han in $A_{Primary}$ in 2004. In $A_{Primary+soc}$ the Bouyei also have the highest share (50 %), followed by the Tujia, Han and the Miao with shares of 22.5 %, 15 % and 12.5 %, respectively, in 2000. In the *BC* sector the Tujia have the highest share (38.3 %), followed by Han (36.2 %), the Bouyei (19.2 %) and the Miao (6.4 %) in 2000. These shares are similar to those of 2004. In the *WC* sector Han have the highest share (33.3 %), followed by the Miao and the Tujia (who both have a share of 23.8 %) and the Bouyei (19.1 %). In 2004 the Bouyei, however, have the highest share in the *WC* sector, followed by Han and the Tujia with the same shares and the Miao with the lowest share.

4.3.3 Descriptive statistics 1997

Table 4.12 shows the average values of the independent variables years of education, percentage of males and age for each ethnic group and the overall sample from 1997. The sample includes 599 individuals, 192 more individuals than in 2004. Based on the fact that the CHNS only contains information of individuals who were at home when the survey was conducted, the larger number of individuals in 1997 may indicate that more individuals migrated to other areas in 2004 than in 1997.

In 1997 the population share of the Bouyei is the highest at 41.2 %, followed by the Tujia (24.2 %), Han (22 %) and the Miao (12.5 %). The sample shares of Han, the Miao, and the Bouyei increase by 0.6, 0.5 and 3 percentage points, respectively, between 1997 and 2004. The share of the Tujia decreases by 4 percentage points between 1997 and 2004. The average education in 1997 is 5.2 years, the same as in 2000, but 0.4 years less than in 2004.

Table 4-12: Descriptive statistics by ethnicity 1997

Variables	Han	Miao	Bouyei	Tujia	Sample
Education mean (SD)	5.7 (4.1)	6 (4)	4.9 (3.6)	4.8 (4.6)	5.2 (4)
Male *%*	50.8	58.7	47.8	51.7	50.8
Age mean (SD)	38.5 (16.8)	37.9 (14.1)	38.4 (15.2)	43.5 (16.1)	39.6 (15.8)
N	132 (22)	75 (12.5)	247 (41.2)	145 (24.2)	599
	- 40	*-22*	*- 67*	*-63*	*-192*

Source: Author's calculation based on CHNS sample.

In 1997 there is no statistically significant educational difference between Han and the Miao. There are, however, statistically significant educational differences between Han and the Bouyei as well as between Han and the Tujia at the 5 % and 1 % significance levels, respectively. Han have on average 5.7 years of education, while the Bouyei and the Tujia only have on average 4.9 and 4.8 years of education, respectively. The test finds that Han have on average 0.8 years more education than the Bouyei and 0.9 years more education than the Tujia. The average percentage of males is 50.8 % in 1997. I find no statistically significant difference in the percentage of males between Han and ethnic minorities. The average age level is 39.6 years in 1997. As in 2004 there is a significant difference in age between Han and the Tujia. Han are on average 38.5 years and Tujia are on average 43.5 years. The test finds that Tujia are on average five years older than Han, which is significant at the 5 % level.

Table 4.13 shows the ethnic distribution in the three counties in 1997. Table 4.13 also shows in italics numerical changes between 1997 and 2004. The general ethnic representation in the counties is the same as in 2000 and 2004. There are larger

Table 4-13: Descriptive statistics counties 1997

Variables	Communities	Han *(N, %)*	Miao *(N, %)*	Bouyei *(N, %)*	Tujia *(N, %)*	Sample *(N, %)*
County 1	522102	68 (41.7)	34 (20.9)	61 (37.4)	0 (0)	163 (27.2)
	522103	*-14*	*-12*	*-12*		*-38*
	522104					
County 2	522201	19 (10.1)	24 (12.8)	0 (0)	145 (77.1)	188 (31.4)
	522202	*-5*	*-3*		*-63*	*-71*
	522203					
	522204					
County 3	522301	45 (18.2)	17 (6.9)	186 (75)	0 (0)	248 (41.4)
	522302	*-21*	*-7*	*-55*		*-83*
	522303					
	522304					
N		132 (22)	75 (12.5)	247 (41.2)	145 (24.2)	599 (100)
		-40	*-22*	*-67*	*-63*	*-192*

The communities 522101, 522401, 522402, 522403, and 522404 are not considered as only Han are living there.

Source: Author's calculation based on CHNS sample.

changes in the observations between 1997 and 2004 than between 2000 and 2004, which may indicate that in 1997 even more household members were at home when the survey was conducted. In all counties and for all ethnic groups the number of observations decreased from 1997 to 2004. For example in 2004 there are 63 fewer observations of the Tujia who live in county 2 than there are in 1997. Likewise in 2004 there are 55 fewer observations of Bouyei in county 3 than there are in 1997.

Tables 4.14 and 4.15 show the occupational outcomes for 1997. The changes in observations between 1997 and 2004 are given in italics. Table 4.14 distinguishes broadly between the sectors *A* and *NA*, and table 4.15 additionally considers subcategories. Most of the people are working in *A*, with shares of 97.6 %, 72.9 % and 76.2 % in counties 1, 2 and 3, respectively (table 4.14). Observations in sector *A* decrease between 1997 and 2004 in all three counties. In the entire sample 156 fewer individuals work in *A* in 2004 than in 1997. The shares of people working in *NA* are 2.6 %, 27.1 % and 23.8 % in counties 1, 2 and 3, respectively. There are 36 fewer individuals working in *NA* in 1997 than in 2004. There are fewer observations in the *NA* sector in counties 2 and 3, but five additional observations in *NA* in county 1 in 1997 than in 2004.

Table 4.15 considers occupational outcomes with subcategories. Among the occupations considered, in 1997 the shares of individuals in $A_{Primary}$ are the highest in counties 1, 2, 3, with 96.3 %, 63.8 % and 65.3 %, respectively. Compared to 2004 there are 76 more individuals in county 1, 48 more individuals in county 2 and 42 more individuals in county 3 working in sector *A* in 1997. When comparing the overall employment shares of $A_{Primary}$ between 1997 and 2004, there is a decrease of 31.5 percentage points and 2.3 percentage points in counties 1 and 2, respectively, but an increase of 7.4 percentage points in county 3. In sum there are 166 fewer individuals working in $A_{Primary}$ in 2004 than in 1997.

Table 4-14: Agriculture versus non-agriculture in counties 1997

Variables	**Agriculture** ($A_{Primary}$, $A_{Primary+soc}$) *(N, %)*	**Non-Agriculture** *(BC + WC) (N, %)*	**Sample** *(N, %)*
County 1	159 (97.6)	4 (2.6)	163 (100)
	-43	*+5*	*-38*
County 2	137 (72.9)	51 (27.1)	188 (100)
	-55	*-16*	*- 71*
County 3	189 (76.2)	59 (23.8)	248 (100)
	-58	*-25*	*-83*
N	485 (81)	114 (19)	599 (100)
	-156	*-36*	*-192*

Source: Author's calculation based on CHNS sample.

Table 4-15: Occupational outcomes in counties 1997

Variables	$A_{Primary}$ (N, %)	$A_{Primary+soc}$ (N, %)	*BC* (N, %)	*WC* (N, %)	Sample (N, %)
County 1	157 (96.3)	2 (1.2)	4 (2.5)	0 (0)	163 (100)
	-76	*+33*	*+3*	*+2*	*-38*
County 2	120 (63.8)	17 (9)	37 (19.7)	14 (7.5)	188 (100)
	-48	*-7*	*-9*	*-7*	*-71*
County 3	162 (65.3)	27 (10.9)	45 (18.2)	14 (5.7)	248 (100)
	-42	*-16*	*-24*	*-1*	*-83*
N	439 (73.3)	46 (7.7)	86 (14.4)	28 (4.7)	599 (100)
	-166	*+10*	*-30*	*-6*	*-192*

Source: Author's calculation based on CHNS sample.

In 1997 the shares of individuals who work in sector *A* and have a *soc* are 1.2 %, 9 % and 10.9 % in counties 1, 2 and 3, respectively. In total ten more individuals have a *soc* alongside *A* in 2004 than in 1997. It is particularly notable that in county 1 the share increases by 26.8 percentage points between 1997 and 2004, while in county 2 and 3 the shares of individuals who work in *A* and have a *soc* decreases by 0.4 percentage points and 4.2 percentage points, respectively. There are 30 fewer individuals working in *BC* positions in 2004 than in 1997. In county 1 the number of people working in *BC* increases by three individuals, but in county 2 and 3 the number of people working in *BC* decreases by nine and twenty-four individuals between 1997 and 2004, respectively. In the *WC* sector six fewer individuals are observed between 1997 and 2004, yet the overall share of *WC* employment increases from 4.7 % in 1997 to 5.4 % in 2004.

Table 4.16 shows the occupational outcomes by ethnicity for 1997. The Bouyei have the highest share in $A_{Primary}$ (46.5 %), followed by the Tujia (21.2 %), Han (18.9 %) and the Miao (13.4 %). The number of individuals who work solely in $A_{Primary}$ decreases for all ethnic groups between 1997 and 2004. The highest decrease is for the Bouyei with 65 observations and for the Tujia with 44 observations. Han and the Miao decrease by 29 and 28 observations, respectively. In $A_{Primary+soc}$ the Bouyei have the highest share (52.2 %), followed by the Tujia (28.3 %), the Miao (13 %) and Han (6.5 %). The share of individuals in $A_{Primary+soc}$ decreases for the Bouyei and the Tujia by four and six observations, respectively.

In contrast the share of individuals in $A_{Primary+soc}$ increases for Han and the Miao by nine and eleven observations, respectively. In the *BC* sector Han have the highest share (43 %), followed by the Tujia (30.2 %), the Bouyei (19.8 %) and the Miao (7 %). The number of workers in the *BC* sector has the greatest decrease (16 observations) for Han between 1997 and 2004. In the *WC* sector the shares are the highest for the Tujia (46.4 %), followed by Han (32.1 %), the Miao (14.3 %) and the Bouyei (7.1 %).

Table 4-16: Occupational outcomes by ethnicity 1997

Variables	Han (N, %)	Miao (N, %)	Bouyei (N, %)	Tujia (N, %)	Sample (N, %)
$A_{Primary}$	83 (18.9) *-29*	59 (13.4) *-28*	204 (46.5) *-65*	93 (21.2) *-44*	439 (100) *-166*
$A_{Primary+soc}$	3 (6.5) *+9*	6 (13) *+11*	24 (52.2) *-4*	13 (28.3) *-6*	46 (100) *+10*
BC	37 (43) *-16*	6 (7) *-4*	17 (19.8) *-5*	26 (30.2) *-5*	86 (100) *-30*
WC	9 (32.1) *-4*	4 (14.3) *-1*	2 (7.1) *+7*	13 (46.4) *-8*	28 (100) *-6*
N	132 (22) *-40*	75 (12.5) *-22*	247 (41.2) *-67*	145 (24.2) *-63*	599 (100) *-192*

Source: Author's calculation based on CHNS sample.

The number of Bouyei working in *WC* increases by seven individuals between 1997 and 2004; Han, the Miao and the Tujia decrease by four, one and eight individuals, respectively.

4.4 Econometric analysis

I follow a five-step procedure to analyze the data samples from each year:

1) I calculate the effects of the independent variables on the binary dependent variable *A* versus *NA* with linear probability and logit models. *A* combines the outcomes $A_{Primary}$ and $A_{Primary+soc}$ and takes the value zero. *NA* combines the outcomes *BC* and *WC* and takes the value one.

2) I estimate the effects of the independent variables on the subgroups; ($A_{Primary}$ versus $A_{Primary+soc}$), ($A_{Primary}$ versus *BC*) and ($A_{Primary}$ versus *WC*). $A_{Primary}$ has always the value zero and the other outcomes have the value one.

3) I test whether outcomes can be combined and then estimate multinomial logit models (MNL) and ordered logit models (OLOGIT) on the categories considered and compare the results. I additionally test the independence from irrelevant alternatives (IIA) assumption in the MNL setting.

4) I include interaction terms between ethnic status and education when the overall model fit improves. The first step is to make LR-tests between models without interaction terms (restricted model) and models with interaction terms (full model). I use the same binary comparisons: *A* versus *NA*, $A_{Primary}$ versus $A_{Primary+soc}$, $A_{Primary}$ versus *BC* and $A_{Primary}$ versus *WC*. I plot predicted probability of working in the *NA*, *BC* or *WC* sectors based on years of education for each ethnic group after logit estimations without interactions. The figures, therefore, serve to better evaluate the statistical significance of interactions between ethnic status and education. The vertical axis of the figures always shows predicted probability of working in the specified sector (*NA*, *BC* or *WC*), and the horizontal axis shows educa-

tion in years. The figures illustrate how probable it is for ethnic groups to work in each of the sectors in relation to years of education.

5) I choose the most appropriate model specification for each year and draw conclusions.

After introducing the econometric models in subchapter 4.4.1, I show the econometric analysis of 2004, 2000 and 1997 in subchapters 4.4.2, 4.4.3 and 4.4.4, respectively. I summarize the first three steps of the estimation procedure broadly in the section "Additive specifications" (subchapter 4.4.2), which are for 2004, 2000 and 1997 the sections 4.4.2.1, 4.4.3.1 and 4.4.4.1, respectively. I summarize the next two steps of the estimation procedure in the section "Interaction terms" (subchapter 4.4.3) for each year. The years 2004, 2000 and 1997 correspond to the sections 4.4.2.2, 4.4.3.2 and 4.4.4.2, respectively. The last subchapter provides interim conclusions, which are shown in sections 4.4.2.3, 4.4.3.3 and 4.4.4.3 for 2004, 2000 and 1997, respectively. In the last section I relate conclusions of the 2004, 2000 and 1997 econometric analyses in a dynamic way; moreover, I relate econometric model outcomes to field observations and determine the accuracy of the hypotheses of subchapter 2.5.1.

4.4.1 Econometric model specifications

On the basis of the five-step estimation approach, this subchapter provides the formal econometric model specifications for those models applied in this monograph, which are the linear probability model in section 4.4.1.1, the logit model in section 4.4.1.2, the multinomial logit model in section 4.4.1.3 and the ordered logit model in section 4.4.1.4. The major source for these model specifications is GREENE (2008). The parameters of the binary logit model, of the MNL model and of the OLOGIT model are estimated with maximum likelihood using Stata 11. For the linear probability model least square estimation is used. Marginal effects are calculated using the delta method.

There is a wide range of available discrete choice models for empirical estimations; therefore, I focus solely on those models which I actually use in this study. The reader who is interested in a comprehensive overview of available discrete choice models is referred to, for example, GREENE (2008, p. 770-859), HENSHER et al. (2005) and TRAIN (2009).

4.4.1.1 The linear probability model

I assume in the econometric application of the linear probability model that individuals, depending on the occupational outcomes in the years considered, work either in A ($Y=0$) or NA ($Y=1$); in $A_{Primary}$ ($Y=0$) or $A_{Primary+soc}$ ($Y=1$); in $A_{Primary}$ ($Y=0$) or BC ($Y=1$) and in $A_{Primary}$ ($Y=0$) or WC ($Y=1$). I assume that the main factors, ethnic status, education, gender, age, community location and village size, given in vector x influence the occupational outcomes. The probability of working in occupation ($Y=1$) is

$$\Pr ob(Y = 1 \mid \mathrm{x}) = F(\mathrm{x}, \beta) \tag{4-1}$$

$$\Pr ob(Y = 0 \mid \mathrm{x}) = 1 - F(\mathrm{x}, \beta)\,, \tag{4-2}$$

where β is a set of parameters and stands for the coefficients of each factor considered (GREENE, 2008, p. 772). I am particularly interested in measuring the marginal effects of the factors considered on the probability of working in *A* or *NA*. GREENE uses a linear regression model, $E[y \mid \mathrm{x}] = F(\mathrm{x}, \beta)$, so that

$$y = E[y \mid \mathrm{x}] + (y - E[y \mid \mathrm{x}]) = x'\beta + \varepsilon \tag{4-3}$$

(GREENE, 2008, p. 772). The linear probability model is, however, inappropriate for measuring discrete choice settings (GREENE, 2008, p. 772-773). The most striking shortcoming of the linear probability model is that $x'\beta$ cannot be constrained between zero and one, which implies that the estimated probability is incorrectly specified and that the variance can be negative.

The linear probability model is, therefore, not used for estimating discrete outcomes. When analyzing discrete outcomes, researchers use the linear probability model "as a basis for comparison to some other more appropriate models" (GREENE, 2008, p. 773). The linear probability estimations can be compared to nonlinear estimations such as the binary logit model and the MNL model.

4.4.1.2 The logit model

Instead of probit models, I use logit models because the MNL model, the workhorse model of discrete choice analysis (HENSHER et al., 2005), uses a logit approach. There is, however, no guideline researchers can apply to choose the most appropriate model specification unless they know about β (GREENE, 2008, p. 774). The logit model uses the logistic distribution, and the probit model, the standard normal distribution. Both distributions are symmetric. The calculated probability mainly differs in the tails (in the extreme values) of the distributions. In the case of small values for $x'\beta$, the logistic distribution is likely to have a larger probability for outcome (Y=1) compared to the normal distribution. In the case of large values for $x'\beta$, the logistic distribution is likely to have a smaller probability for outcome (Y=1) compared to the normal distribution. Not using an exact distribution based on β has the disadvantage that correlations of residuals are constrained, which means that relationships between occupational outcomes (left-hand side variable) and explanatory factors (right-hand side variables) can be incorrect.

In the econometric application of the binary logit model, I assume that individuals either work in *A* (*Y=0*) or *NA* (*Y=1*), either in $A_{Primary}$ (*Y=0*) or $A_{Primary+soc}$ (*Y=1*), either in $A_{Primary}$ (*Y=0*) or *BC* (*Y=1*) and either in $A_{Primary}$ (*Y=0*) or *WC* (*Y=1*). I assume that the main factors given in vector x, ethnic status, education, gender, age, community location and village size, have an influence on the occupational outcomes, so that the probability of working in occupation (*Y=1*) is

$$\Pr ob(Y=1\,|\,\mathrm{x}) = \frac{e^{x'\beta}}{1+e^{x'\beta}} = \Lambda(x'\beta)\,, \tag{4-4}$$

where $\Lambda(\cdot)$ denotes the logistic cumulative distribution function (GREENE, 2008, p. 773). In the econometric application I use the "logit" command implemented in Stata 11 for estimating all logit models in this monograph.

Unrelated to the distribution the probability model is

$$E[y\,|\,\mathrm{x}] = 0[1-F(x'\beta)] + 1[F(x'\beta)] = F(x'\beta) \tag{4-5}$$

(GREENE, 2008, p. 774).

Based on the nonlinearity of the model, the parameters can differ from marginal effects. Marginal effects are

$$\frac{\partial E[y\,|\,\mathrm{x}]}{\partial\,\mathrm{x}} = \left\{\frac{dF(x'\beta)}{d(x'\beta)}\right\}\beta = f(x'\beta)\beta\,, \tag{4-6}$$

where $f(\cdot)$ is the density function which matches the cumulative distribution $F(\cdot)$ (GREENE, 2008, p. 774). In the logit model the marginal effects are, thus, calculated as

$$\frac{\partial E[y\,|\,\mathrm{x}]}{\partial\,\mathrm{x}} = \Lambda x'\beta[1-\Lambda(x'\beta)]\beta \tag{4-7}$$

(GREENE, 2008, p. 775). I calculate average marginal effects with the "margins" command of Stata 11.[25] This implies that for each observation the marginal effect with respect to an explanatory factor, averaged over the estimation sample, is computed (BAUM, 2010, p. 13).

AI and NORTON (2003) argue that marginal effects in non-linear models with interaction terms are incorrectly measured with statistical software packages. Marginal effects can be of opposite sign, and standard error for each needs to be calculated separately (AI and NORTON, 2003). In the software package Stata 11 the newly implemented "margins" command, however, controls for the drawbacks of previous commands and optimally calculates the marginal effects with the delta method (BAUM, 2010).

For dummy variables the "margins" command calculates a discrete change from the base level (*Y=0*). The formal specification for calculating marginal effects of dummy variables such as *d* is

$$\Pr ob\left[Y=1\,|\,\bar{x}_{(d)}, d=1\right] - \Pr ob\left[Y=1\,|\,\bar{x}_{(d)}, d=0\right], \tag{4-8}$$

where $\bar{x}_{(d)}$ stands for the means of all other variables in the model (GREENE, 2008, p. 775).

[25] Stata's "margins" command is constantly updated; please refer to KARACA-MANDIC et al. (2012) and WILLIAMS (2012) for current information.

There are several sources for additional information about logit and probit models, for example, ALDRICH and NELSON (1984), CAMERON and TRIVEDI (2009), GREENE (2008), JONES (2007), LONG (1997), LONG and FREESE (2003), PAMPEL (2000) or POWERS and XIE (2008). Researchers who prefer that estimates are presented as odds ratios rather than as coefficients can use logistic regressions instead of logit models, see, for example, GOULD (2000), KLEINBAUM and KLEIN (2002), HOSMER and LEMESHOW (2000) and PAMPEL (2000).

4.4.1.3 The multinomial logit model

In the econometric application of the multinomial logit model, I assume, depending on the occupational outcomes in each year, that individuals either work in $A_{Primary}$ (*Y=0*), $A_{Primary+soc}$ (*Y=1*) or *NA* (*Y=2*). In the cases where non-agricultural employment cannot be collapsed into a single *NA* category, I distinguish between *BC* and *WC* occupations. The occupational outcomes are given with the letter *j* in GREENE's formal model specification (2008, p. 843). As the zero-outcome ($A_{Primary}$) is not counted in *j*, in total there are *j+1* outcome categories. The zero-outcome is crucial for normalizing the model. The outcome with the most observations, which in this case is $A_{Primary}$, is set to zero by default.

The main factors, ethnic status, education, gender, age, community location and village size, are represented by w_i for each individual *i* (GREENE, 2008, p. 844). The MNL model estimates a set of coefficients α_j, which correspond to each outcome, and measures changes of all considered outcomes relative to the base outcome. The obtained probabilities are

$$\Pr ob(Y_i = j \mid w_i) = P_{ij} = \frac{\exp(w_i'\alpha_j)}{1+\sum_{k=1}^{J}\exp(w_i'\alpha_k)}, \quad j = 0,1,...,J, \quad \alpha_0 = 0 \tag{4-9}$$

(GREENE, 2008, p. 844). The MNL model reduces to the binary logit model when *j=1*. It is, moreover, possible to calculate *j*-odd ratios with the MNL model,

$$\ln\left[\frac{P_{ij}}{P_{ik}}\right] = w_i'(\alpha_j - \alpha_k) = w_i'\alpha_j \quad \text{if } k = 0 \tag{4-10}$$

(GREENE, 2008, p. 844). The marginal effects δ_{ij} of the model are

$$\delta_{ij} = \frac{\partial P_{ij}}{\partial w_i} = P_{ij}\left[\alpha_j - \bar{\alpha}\right] \tag{4-11}$$

(GREENE, 2008, p. 845). The equations show that $\boldsymbol{\alpha}$ enters the marginal effects both through the probability in (4-10) and through the weighted average in (4-11). The signs of δ_{ij} and α_j need not be equal. In the econometric application I use the "mlogit" command of Stata 11 to estimate all MNL models in this monograph. I calculate average marginal effects with the "margins" command of Stata 11 as explained in the previous section.

Model specification test

The independence from irrelevant alternatives (IIA) assumption is one major assumption of the MNL model. The IIA assumption is based on the general assumption of the MNL model that error terms are independent and homoskedastic. The odds ratio in equation (4-10) implies that the chosen alternative (occupation) is independent from all other available alternatives. This implies that the decision to exit (or to remain in) sector A is not related to whether there is only one or several job alternatives available in the *NA* sector, which contradicts theoretical assumptions.

The IIA assumption can be tested. An irrelevant subset of the choice set has no systematic influence on parameter estimates when excluded from the model; therefore, if the irrelevant subset is omitted, the restricted model will give inefficient but consistent results (HAUSMAN and MCFADDEN, 1984). If the remaining odds ratios are not independent from the subset, the parameter estimates will be inconsistent. The Hausman specification test applies this logic to test the IIA assumption. The statistic used in the Hausman specification test is

$$\chi^2 = (\hat{\boldsymbol{\alpha}}_s - \hat{\boldsymbol{\alpha}}_f)'[\hat{\mathbf{V}}_s - \hat{\mathbf{V}}_f]^{-1}(\hat{\boldsymbol{\alpha}}_s - \hat{\boldsymbol{\alpha}}_f), \tag{4-12}$$

where subscript *s* denotes estimators of the restricted subset, subscript *f* denotes estimators of the full choice set and $\hat{\mathbf{V}}_s$ and $\hat{\mathbf{V}}_f$ denote estimates of the asymptotic covariance matrices (GREENE, 2008, p. 847). The statistic has a chi-squared distribution with *K* degrees of freedom.

For example among the three possible occupational outcomes, $A_{Primary+soc}$ seems to be a closer substitute to $A_{Primary}$ than to *NA* employment. This means that excluding $A_{Primary+soc}$ from the model is expected to affect the remaining occupational outcomes as $A_{Primary+soc}$ may be correlated with $A_{Primary}$ and, therefore, may violate the IIA assumption. To conduct the Hausman test, I first calculate the consistent estimator $\hat{\theta}_1$ with the originally specified MNL model and second I exclude one outcome from the model to obtain the efficient estimator $\hat{\theta}_2$. The null hypothesis states that the estimator $\hat{\theta}_2$ is in reality a consistent and efficient estimator of the true parameters. If the null hypothesis is true, there should be no systematic difference between the two estimators. The assumptions of the efficient estimator must, otherwise, be doubted. The classical Hausman test, however, cannot be used if observations come from a clustered sample; therefore, I use a variant implemented as the SUEST test (seemingly unrelated estimation) in Stata 11 for testing the IIA assumption.

Some authors provide additional information about the MNL model, for example, GREENE (2008, p. 843-845), HOSMER and LEMESHOW (2000, p. 260-287), LONG and FREESE (2003, chapters six and seven) and TREIMAN (2009).

4.4.1.4 The ordered logit model

The econometric setting of the ordered logit model requires that the occupational outcomes in the years considered, $A_{Primary}$ ($Y=0$), $A_{Primary+soc}$ ($Y=1$) and NA ($Y=2$), are categorical and ordered. This implies that the outcome with the lowest value ($Y=0$) is the least preferred and that the outcome with the highest value is ($Y=2$) the most preferred outcome. In the estimation not only the β coefficients are estimated, but also cut-points μ, which define the different categories of the model. The following probabilities are estimated

$$\Pr ob(y=0 \mid x) = \Lambda(-x'\beta)\,, \tag{4-13}$$

$$\Pr ob(y=1 \mid x) = \Lambda(\mu_1 - x'\beta) - \Lambda(-x'\beta)\,, \tag{4-14}$$

$$\Pr ob(y=2 \mid x) = \Lambda(\mu_2 - x'\beta) - \Lambda(\mu_1 - x'\beta)\,, \tag{4-15}$$

$$\vdots$$

$$\Pr ob(y=J \mid x) = 1 - \Lambda(\mu_{J-1} - x'\beta)\,, \tag{4-16}$$

where $\Lambda(\cdot)$ denotes the logistic cumulative distribution function (GREENE, 2008, p. 832). x is the same vector as in the other model specifications, including the factors ethnic status, education, gender, age, community location and village size. The marginal effects for an OLOGIT model with three categories are

$$\frac{\partial \Pr ob(y=0 \mid x)}{\partial x} = -\Lambda(x'\beta)\beta\,, \tag{4-17}$$

$$\frac{\partial \Pr ob(y=1 \mid x)}{\partial x} = \left[\Lambda(-x'\beta) - \Lambda(\mu - x'\beta)\right]\beta\,, \tag{4-18}$$

$$\frac{\partial \Pr ob(y=2 \mid x)}{\partial x} = \Lambda(\mu - x'\beta)\beta \tag{4-19}$$

(GREENE, 2008, p. 833). I use the "OLOGIT" command of Stata 11 for estimating all ordered logit models in this monograph. More information about the ordered logit model is provided by, for example, LONG and FREESE (2003, chapter 5) and CAMERON and TRIVEDI (2005, chapter 15).

4.4.2 Econometric analysis 2004

I follow the five-step procedure shown at the beginning of this subchapter to analyze the data sample from 2004. All interim results tables and figures for 2004 are given at the end of each section.

4.4.2.1 Additive specifications

First, I calculate the effects of the independent variables on the binary dependent variable A versus NA with linear probability and logit models. The estimation results are shown in table 4.17. The order of the columns is the same for all results tables in this subchapter. The first column on the left gives the independent variables

used in the analysis. The second column from the left shows OLS coefficients, which are estimated with a linear probability model. The third column from the left shows the coefficients after logit estimation, and the last column shows the marginal effects after logit estimation.

I find that both in the linear probability and in the logit models the variables Miao and age have significantly negative effects on the probability of working in *NA*. The marginal effects after logit indicate that being Miao rather than Han decreases the probability of working in *NA* by 7.14 %.[26] Each additional year of age from the average decreases the probability of working in *NA* by 0.6 %. With additional education the probability of working in *NA* increases in both models. The marginal effects after logit estimation show that each additional year of education from the average increases the probability of working in *NA* by 1.9 %. Residence in larger villages also increases the probability for *NA* employment. The variable male and county 2, moreover, have significantly positive effects on the probability of working in *NA* in the logit model, but not in the linear probability model. Men have a 5.8 % higher probability of working in *NA* than women in the logit model. Individuals residing in county 2 have a 19.9 % higher probability of working in *NA* than individuals in county 1. The Tujia's larger share in *BC* (25.6 %) in county 2 can partly be explained by the fact that they make up 70.1 % of the population of that county.

Second, I analyze the subrelationships of *A* and *NA* and compare the outcomes $A_{Primary}$ and $A_{Primary+soc}$ (see table 4.18). The results of the linear probability and the logit models have the same significant effects except for the variable Bouyei. Bouyei only has a positive significant effect on the probability of having a soc in the linear probability model, but not in the logit model. The magnitudes of the other variables are the same in the linear probability and logit models. The variables Miao, Tujia and male have statistically significant positive effects on the probability of having a *soc*, while the variables age, county 2 and county 3 have statistically significant negative effects on the probability of having a *soc*. The marginal effects after logit show that being Miao or Tujia rather than Han increases the probability of having a *soc* by 26.1 % and 15.6 %, respectively. I also find that being male rather than female increases the probability of having a *soc* by 18.8 %. The average marginal effects for age are, however, not statistically significant. In counties 2 and 3 the probability of having a *soc* is lower by 17.5 % and 17.3 % than in county 1, respectively.

I continue analyzing the subrelationships within sectors *A* and *NA* and compare the outcomes $A_{Primary}$ and *BC* (see table 4.19). The magnitudes of the estimation results are the same in the linear probability and logit models. Being Miao and being older decrease the probability of working in *BC* positions, while having more

[26] The marginal effects are based on the assumption that other effects are held constant; therefore, I do not state the *ceteris paribus* assumption for each marginal effect.

years of education, being male and living in larger villages increase the probability of working in *BC* positions. The marginal effects after logit estimation show that being Miao rather than Han decreases the probability of having a *BC* position by 9 %. If age increases by one year from the average level, the probability of working in *BC* decreases by 0.6 %. In contrast each additional year of education from the average increases the probability of working in *BC* by 1.6 %.

The next step is comparing the outcomes $A_{Primary}$ and *NA*. *NA* combines the outcomes *BC* and *WC*. I do not compare $A_{Primary}$ and *WC* because of the comparatively low number of observations in *WC* and the related computational drawbacks. The aggregations of the *BC* and *WC* sectors result in the same magnitudes as in the previous model, where only *BC* was considered (see table 4.20). As in the previous model, I find that Miao status and older age result in a lower probability of working in *NA* both in the linear probability and in the logit model, while more years of education, male status and residence in larger villages increase the probability of working in *NA*. The marginal effects after logit show that Miao have a 6.7 % lower probability of working in *NA* than Han. An additional year of age decreases the probability of working in *NA* by 0.6 %, while each additional year of education increases the probability of working in *NA* by 1.8 %. Men, moreover, have a 9.5 % higher probability of working in *NA* than women.

Third, I consider all outcomes in one single model. I use two different approaches, the MNL model and the OLOGIT model. As already pointed out, the MNL model assumes that the outcome categories are unordered, and the OLOGIT model assumes that the outcome categories are ranked. Outcomes with lower levels are, therefore, inferior to outcomes with higher levels in OLOGIT models. I first apply Wald tests to find out whether outcomes can be collapsed. I then estimate the linear probability model, the MNL model and the OLOGIT model. For the MNL model I have to conduct further tests. The MNL model follows the IIA assumption, which hypothesizes that the outcomes considered are independent from irrelevant alternatives. The IIA assumption, therefore, has to be tested after the estimation. While the linear probability model only serves as a comparison to the nonlinear model, the overall model fit and the prediction power of the MNL and the OLOGIT models must be compared to choose the model which best fits the data sample.

To determine which outcome categories can be collapsed, I use Wald tests (LONG and FREESE, 2003, p. 204). The possible outcome categories are $A_{Primary}$ (0), $A_{Primary+soc}$ (1), *BC* (2) and *WC* (3) (see table 4.21). In table 4.21 the first column from the left shows the pairs of tested categories, the following columns show the chi^2 values, the degrees of freedom and the p-values, respectively. The high p-value of the comparison of categories 2 and 3 suggests that these two outcomes can be collapsed into one single category.

Because *BC* and *WC* can be collapsed in the multinomial models, I consider three outcomes $A_{Primary}$ (0), $A_{Primary+soc}$ (1) and *NA* (2) (see table 4.22). In the linear probability and OLOGIT models, ethnic status has no significant influence on

occupational outcomes. In the linear probability model having more years of education, being male and living in larger villages increase the probability of working in *NA*, while being older decreases the probability of working in *NA*.

The MNL results provide two equations. One equation compares the outcomes $A_{Primary}$ (0) and $A_{Primary+soc}$ (1); the other equation compares the outcomes $A_{Primary}$ (0) and *NA* (2). I find that by comparing outcomes $A_{Primary}$ (0) and $A_{Primary+soc}$ (1), the MNL coefficients (table 4.22) have the same magnitude as the logit coefficients (table 4.18), except for village size, which has a statistically significant positive effect in the MNL model but is not statistically significant in the logit model. The MNL results show that the Miao and the Tujia both have a higher probability of having a *soc* than do Han. Men, moreover, have a higher probability of having a *soc* than women. Older age decreases the probability of having a *soc*. Residence in larger villages increases the probability of having a *soc*. In counties 2 and 3 the probability that individuals have a *soc* is lower than in county 1. By comparing the outcomes $A_{Primary}$ (0) and *NA* (2), ethnic status has no influence on the probability of working in *NA*. The major factors which increase the probability of working in *NA* are more years of education, male status and residence in larger villages. In contrast older age decreases the probability of working in *NA*. The MNL results of the second equation differ from the logit results (table 4.20). The Miao coefficient is statistically significant positive in the logit estimation, but the Miao coefficient is not statistically significant in the MNL estimation. In the OLOGIT estimation ethnic status has no significant influence on occupational outcomes. In the OLOGIT model having more years of education, being male and living in larger villages increase the probability of working in *NA*, while being older and living in county 3 rather than in county 1 decrease the probability of working in *NA*.

The results after MNL and OLOGIT estimation, however, have to be considered with caution. After testing the IIA assumption, the test results show that the IIA assumption is violated (table 4.23). This means that the estimation results are not independent from irrelevant alternatives. If one outcome category is excluded from the analysis, the coefficients of the remaining outcome categories will change, which violates the IIA assumption. An alternative to the MNL model is the OLOGIT model. It is possible to use the OLOGIT model instead of the MNL model only if the OLOGIT model better predicts the data sample than does the MNL model. The prediction power of the MNL and OLOGIT models can be compared by computing the predicted probability for both models, plotting the predicted probability for both models and analyzing the correlations between MNL and OLOGIT coefficients (LONG and FREESE, 2003, p. 211-212).

The categories $A_{Primary}$ after MNL and $A_{Primary}$ after OLOGIT, *NA* after MNL and *NA* after OLOGIT have correlations of 0.98 and 0.94, respectively. The category $A_{Primary+soc}$ after MNL and $A_{Primary+soc}$ after OLOGIT, however, has a correlation of only 0.61 (figure 4-2). Like LONG and FREESE (2003, p. 212) I also find that the predictions after OLOGIT are suddenly truncated, which may not reasonably

explain the real data sample. This implies that the MNL model has better prediction power than the OLOGIT model. Although the last test suggests that the MNL is preferable, the violation of the IIA assumption also casts doubt on MNL results. Multinomial models which do not violate the IIA assumption are, therefore, required, yet these models require alternative-specific variables.

Alternative-specific variables are variables which are specific to the outcome categories and which are not specific to the individual. An appropriate alternative specific variable is, for example, the wage rate for each occupation. The CHNS is, however, mainly based on individual data and, therefore, does not provide wage rates for all occupational outcomes. A possible solution would be to calculate the average wage rate for each occupation based on individual data and to use the average values as alternative-specific wage rates for each occupation. When I included these alternative-specific wage rates into the regression, the variation of wage differences between sectors was, however, not significant; the inclusion of these variables gave no additional explanation about different occupational outcomes; therefore, I do not use alternative-specific models and draw conclusions from estimation results of the binary logit models.

Table 4-17: Estimation results 2004 (A versus NA)

	OLS Coefficients *A (0), NA (1)*	**Logit Coefficients** *A (0), NA (1)*	**Logit Marginal Effects** *A (0), NA (1)*
Bouyei (0/1)	-0.122	0.789	0.055
	(0.144)	(1.157)	(0.075)
Miao (0/1)	-0.152***	-1.185***	-0.071***
	(0.040)	(0.373)	(0.022)
Tujia (0/1)	-0.035	-0.361	-0.025
	(0.073)	(1.021)	(0.066)
Education (years)	0.022**	0.265***	0.019***
	(0.008)	(0.049)	(0.003)
Male (0/1)	0.057	0.830**	0.058**
	(0.034)	(0.377)	(0.028)
Age (years)	-0.005***	-0.079***	-0.006***
	(0.001)	(0.015)	(0.001)
Households (*N*)	0.001***	0.008***	0.001***
	(0.001)	(0.003)	(0.001)
County 2 (0/1)	0.087	2.293*	0.199*
	(0.091)	(1.259)	(0.111)
County 3 (0/1)	0.112	0.357	0.026
	(0.126)	(0.726)	(0.052)
Constant	0.090	-5.463***	
	(0.107)	(1.856)	
Observations	407	407	407
R^2	0.4316		
Log ps. Likelihood		-96.590874	
Pseudo R^2		0.5143	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-18: Estimation results 2004 (A-Primary versus A-Primary+soc)

	OLS Coefficients $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*	**Logit Coefficients** $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*	**Logit Marginal Effects** $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*
Bouyei (0/1)	0.048*	0.474	0.049
	(0.025)	(0.311)	(0.032)
Miao (0/1)	0.255***	1.958***	0.261***
	(0.039)	(0.263)	(0.032)
Tujia (0/1)	0.180***	1.313**	0.156*
	(0.053)	(0.633)	(0.085)
Education (years)	0.009	0.097	0.010
	(0.008)	(0.070)	(0.007)
Male (0/1)	0.183***	1.928***	0.188***
	(0.032)	(0.449)	(0.037)
Age (years)	-0.004**	-0.034***	-0.004
	(0.001)	(0.010)	(0.001)
Households (*N*)	0.001	0.003	0.001
	(0.001)	(0.002)	(0.001)
County 2 (0/1)	-0.280***	-2.009***	-0.175***
	(0.069)	(0.419)	(0.034)
County 3 (0/1)	-0.205***	-1.753***	-0.173***
	(0.044)	(0.382)	(0.030)
Constant	0.168	-2.534**	
	(0.151)	(1.027)	
Observations	329	329	329
R^2	0.227		
Log ps. Likelihood		-109.9263	
Pseudo R^2		0.268	

Note: Standard errors are adjusted to 9 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-19: Estimation results 2004 (A-Primary versus BC)

	OLS Coefficients	**Logit Coefficients**	**Logit Marginal Effects**
	$A_{Primary}$ (0), BC (1)	*$A_{Primary}$ (0), BC (1)*	*$A_{Primary}$ (0), BC (1)*
Bouyei (0/1)	-0.144	0.027	0.002
	(0.136)	(0.874)	(0.057)
Miao (0/1)	-0.193***	-1.826***	-0.090***
	(0.036)	(0.559)	(0.026)
Tujia (0/1)	-0.027	0.213	0.014
	(0.081)	(0.996)	(0.068)
Education (years)	0.015**	0.243***	0.016***
	(0.006)	(0.063)	(0.004)
Male (0/1)	0.101**	1.239***	0.081***
	(0.034)	(0.406)	(0.029)
Age (years)	-0.006***	-0.094***	-0.006***
	(0.001)	(0.020)	(0.001)
Households (*N*)	0.001***	0.008***	0.001***
	(0.001)	(0.003)	(0.001)
County 2 (0/1)	0.057	1.331	0.098
	(0.083)	(1.134)	(0.090)
County 3 (0/1)	0.058	0.132	0.009
	(0.110)	(0.804)	(0.052)
Constant	0.204**	-4.183**	
	(0.088)	(1.754)	
Observations	329	329	329
R^2	0.432		
Log ps. Likelihood		-70.335808	
Pseudo R^2		0.5314	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-20: Estimation results 2004 (A-Primary versus NA)

	OLS Coefficients $A_{Primary}$ *(0), NA (1)*	**Logit Coefficients** $A_{Primary}$ *(0), NA (1)*	**Logit Marginal Effects** $A_{Primary}$ *(0), NA (1)*
Bouyei (0/1)	-0.143	0.636	0.046
	(0.140)	(0.920)	(0.063)
Miao (0/1)	-0.146***	-1.013**	-0.067***
	(0.025)	(0.393)	(0.022)
Tujia (0/1)	-0.013	-0.123	-0.009
	(0.085)	(1.205)	(0.087)
Education (years)	0.021**	0.234***	0.018***
	(0.008)	(0.050)	(0.003)
Male (0/1)	0.106**	1.250***	0.095***
	(0.041)	(0.402)	(0.032)
Age (years)	-0.006***	-0.087***	-0.006***
	(0.001)	(0.015)	(0.001)
Households (*N*)	0.001***	0.008***	0.001***
	(0.001)	(0.003)	(0.001)
County 2 (0/1)	0.034	1.692	0.143
	(0.103)	(1.161)	(0.103)
County 3 (0/1)	0.084	0.029	0.002
	(0.121)	(0.686)	(0.051)
Constant	0.203*	-4.574***	
	(0.094)	(1.598)	
Observations	351	351	351
R^2	0.461		
Log ps. Likelihood		-85.750287	
Pseudo R^2		0.539	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-21: Wald tests for combining outcome categories 2004

Ho: All coefficients except intercepts associated with given pair of outcomes are 0 (i.e., categories can be collapsed).

Categories tested	**Chi 2**	**df**	**P>chi2**
0-1	52.397	9	0.000
0-2	62.547	9	0.000
0-3	46.797	9	0.000
1-2	39.009	9	0.000
1-3	26.075	9	0.002
2-3	11.105	9	0.269

Source: Author's calculation based on CHNS sample, test commands and results based on LONG and FREESE (2003, p. 204).

Table 4-22: Multinomial Estimation Results 2004

	OLS Coefficients	MNL Coefficients		OLOGIT Coefficients
	$A_{Primary}$ (0), $A_{Primary+soc}$ (1), NA (2)	$A_{Primary}$ (0), $A_{Primary+soc}$ (1)	$A_{Primary}$ (0), NA (2)	$A_{Primary}$ (0), $A_{Primary+soc}$ (1), NA (2)
Bouyei (0/1)	-0.236	0.338	0.828	0.465
	(0.251)	(0.207)	(1.127)	(0.779)
Miao (0/1)	-0.048	1.816***	-0.612	0.465
	(0.118)	(0.180)	(0.457)	(0.408)
Tujia (0/1)	0.082	1.323*	-0.049	0.265
	(0.144)	(0.723)	(1.128)	(0.669)
Education(years)	0.045**	0.083	0.288***	0.188***
	(0.014)	(0.071)	(0.050)	(0.038)
Male (0/1)	0.254***	1.890***	1.192***	1.338***
	(0.073)	(0.412)	(0.430)	(0.320)
Age (years)	-0.013***	-0.034***	-0.088***	-0.057***
	(0.002)	(0.010)	(0.016)	(0.006)
Households (*N*)	0.001***	0.003**	0.009***	0.007***
	(0.000)	(0.001)	(0.003)	(0.003)
County 2 (0/1)	-0.109	-2.001***	1.653	0.114
	(0.124)	(0.463)	(1.154)	(0.458)
County 3 (0/1)	0.039	-1.723***	-0.152	-0.794*
	(0.212)	(0.356)	(0.721)	(0.466)
Constant	0.473***	-2.506***	-5.197***	
	(0.144)	(0.821)	(1.833)	
Cut 1				3.229**
				(1.289)
Cut 2				4.509***
				(1.397)
Observations	407	407	407	407
R^2	0.4319			
Log ps. Likelihood		-206.901		-232.21965
Pseudo R^2		0.4071		0.3345

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

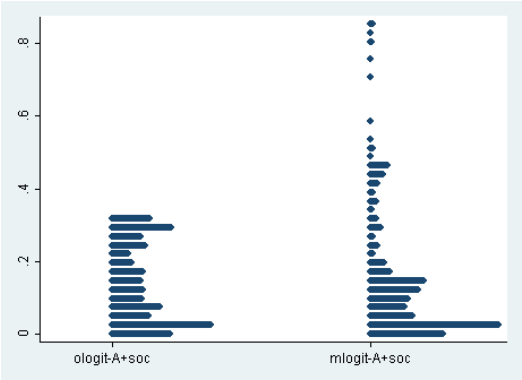
Source: Author's calculation based on CHNS sample.

Table 4-23: Wald test after Suest 2004

H_0 = no difference between estimators of full and restricted model *H_1 = difference between estimators of full and restricted model*		
	A+soc	***NA***
Model Restrictions		
Exclusion *A+soc*		chi2(10) = 6.4e+06 Prob > chi2 = 0.001
Exclusion *NA*	chi2(9) = 82.29 Prob > chi2 = 0.001	

$A_{Primary}$ is base group. Standard errors are based on cluster robust variance estimator.
Source: Author's calculation based on CHNS sample.

Figure 4-2: Plot to compare predicted probabilities of sector A-Primary+soc in MNL and OLOGIT models 2004



Source: Author's calculation based on CHNS sample.

4.4.2.2 Interaction terms

The next step is to include interaction terms between ethnic status and education into the regression. The results of linear probability and logit models suggest that education has a significantly positive effect on the probability of working in *NA*. I am, therefore, interested in answering the question of whether interactions of Bouyei*Education, Miao*Education and Tujia*Education have effects on occupational outcomes. Does the probability of working in *NA* increase, if the Miao, the Bouyei and the Tujia have more years of education?

In the first model (*A* versus *NA*), LR-tests suggest that including Miao*Education improves model fit. The plot of predicted probabilities of the restricted model (without interaction terms) illustrates the probability of working in *NA* on the y-axis and years of education on the x-axis (figure 4-3). The data are separately plotted for each ethnic group and for the sample average. The probability of working in *NA* is, in most of the cases, below average for the Bouyei and the Miao; Han and the

Tujia have above average probabilities in most of the cases. The general trend shows that with increasing years of education, there is also an increasing probability of working in *NA*.

Table 4.24 provides estimation results of models with the interaction term Miao*Education. The magnitudes of the estimation results are the same as in the restricted model (table 4.17), except for the marginal effect of county 2, which is no longer statistically significant. The interaction term Miao*Education is not statistically significant in the linear probability model, but it has a statistically significant positive value in the logit model. The marginal effects after logit show that individuals with Miao status rather than with Han status increase their probability of working in *NA* by 5.4 % with each additional year of education over the average. The Miao effect, however, decreases the probability of working in *NA* by 20.9 %. Better educated Miao, therefore, have a higher probability of working in *NA* than have less educated Miao.

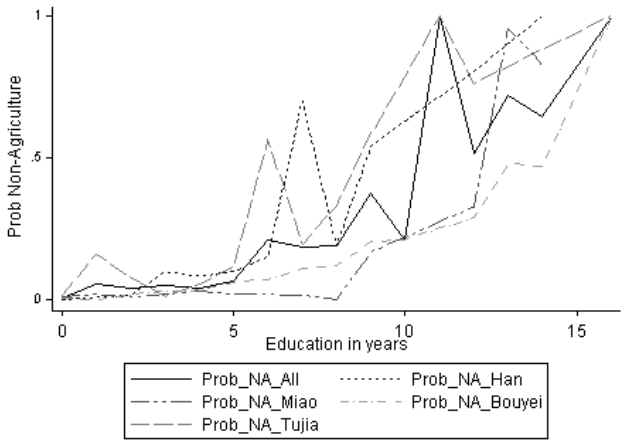
The next step is to analyze interactions between ethnic status and education by comparing the subsectors. I first compare $A_{Primary}$ and $A_{Primary+soc}$. In this comparison the probability of having a *soc* increases with years of education (figure 4-4). The predicted probability of working in *NA* and education plot shows that from zero to five years of education, the probability of having a *soc* is around zero for all ethnic groups, except for the Miao. The Miao have a higher probability of having a *soc* than the average in all educational cohorts, except from ten to twelve years of education, where the probability of having a *soc* was lower than the average. The Tujia, moreover, increase their probability of having a *soc* from eight to twelve years of education. In contrast the Bouyei have a below average probability of having a *soc* at all levels of education, yet the probability of having a *soc* does increase with years of education. The LR-tests indicate that including the interaction term Bouyei*Education significantly increases the model fit (table 4.25). The marginal effects after logit show that each additional year of education from the average increases the probability of having a *soc* by 2.4 % for the Bouyei rather than for Han. With the inclusion of the interaction term Bouyei*Education, the Bouyei coefficient is, however, no longer significant in the linear probability model. By comparing the logit results, the magnitudes of the coefficients are the same in both models with or without interaction terms.

The next step is to compare the sectors $A_{Primary}$ and *BC*. The plot of predicted probabilities and education shows the general trend that with increased years of education, there is an increased probability of working in *BC* (figure 4-5). In most cases the Bouyei and the Miao have below average values. In contrast Han and the Tujia have mainly above average values. The overall model fit, however, only improves when the interaction term Miao*Education is included in the estimation (table 4.26). The interaction term Miao*Education is statistically significant in the logit model but is insignificant in the linear probability model. The marginal effects after logit indicate that each additional year of education from the average

increases the probability of the Miao of working in *BC* positions by 1.9 %. The magnitudes of the coefficients are the same in the linear probability and logit estimations with and without interaction terms.

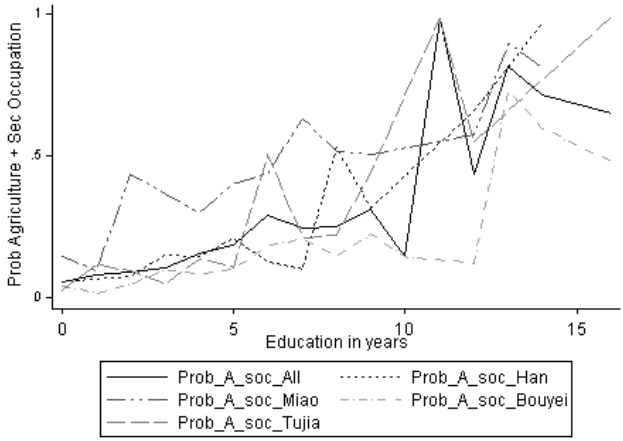
The final step in the 2004 analysis is to compare sectors $A_{Primary}$ and *NA*. The plot of predicted probability of working in *NA* and education shows that with increasing years of education, in most of the cases there is also an increasing probability of working in *NA* (figure 4-6). As in the previous comparison, the Bouyei and the Miao generally have a below average probability of working in *NA*. In contrast Han and the Tujia have in most of the cases an above average probability of working in *NA*. The LR-tests, however, indicate that even in the consideration of the entire *NA* sector, only the interaction term Miao*Education is statistically significant (table 4.27). The interaction term Miao*Education is not statistically significant in the linear probability model, but it is statistically significant in the logit model. The marginal effects after logit estimation indicate that with an additional year of education, Miao people have a 5.2 % higher probability of working in *NA*. The inclusion of the interaction term Miao*Education leaves the magnitudes of the other coefficients unchanged in the linear probability and logit models.

Figure 4-3: Plot of predicted probabilities and years of education by ethnicity for A versus NA 2004



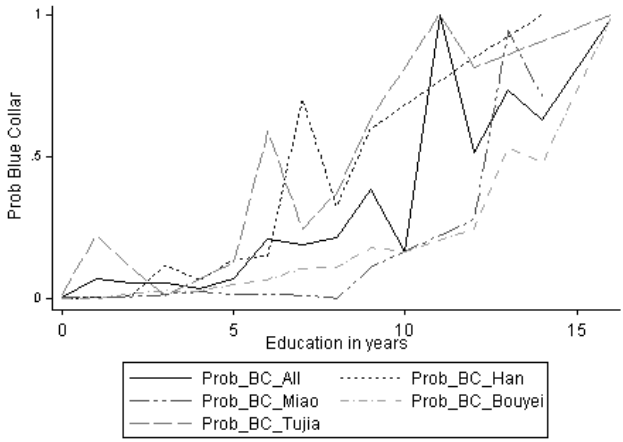
Source: Author's calculation based on CHNS sample.

Figure 4-4: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus A-Primary+soc



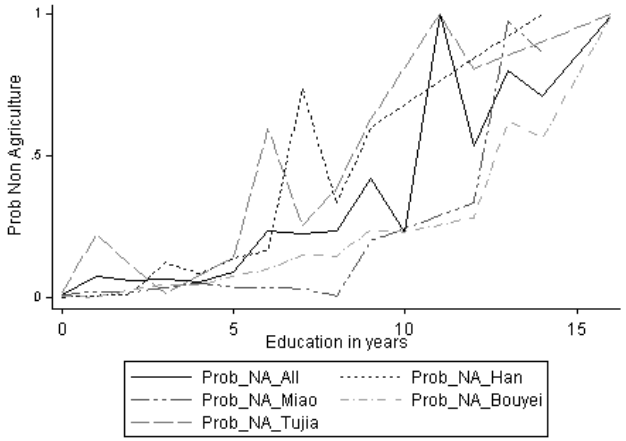
Source: Author's calculation based on CHNS sample.

Figure 4-5: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus BC 2004



Source: Author's calculation based on CHNS sample.

Figure 4-6: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus NA 2004



Source: Author's calculation based on CHNS sample.

Table 4-24: Estimation results with interactions 2004 (A versus NA)

	OLS Coefficients *A (0), NA (1)*	**Logit Coefficients** *A (0), NA (1)*	**Logit Marginal Effects** *A (0), NA (1)*
Bouyei (0/1)	-0.122	0.767	0.052
	(0.143)	(1.175)	(0.074)
Miao (0/1)	-0.191**	-8.227***	-0.209***
	(0.080)	(1.904)	(0.03)
Tujia (0/1)	-0.035	-0.253	-0.018
	(0.075)	(0.959)	(0.064)
Miao*Education	0.007	0.751***	0.054***
	(0.010)	(0.160)	(0.016)
Education (years)	0.020**	0.225***	0.016***
	(0.007)	(0.051)	(0.003)
Male (0/1)	0.057	0.804**	0.056**
	(0.034)	(0.357)	(0.026)
Age (years)	-0.005***	-0.084***	-0.006***
	(0.001)	(0.013)	(0.001)
Households (*N*)	0.001***	0.008***	0.001***
	(0.001)	(0.003)	(0.001)
County 2 (0/1)	0.089	2.140*	0.182
	(0.090)	(1.293)	(0.112)
County 3 (0/1)	0.113	0.327	0.023
	(0.125)	(0.700)	(0.049)
Constant	0.104	-4.871**	
	(0.112)	(1.908)	
Observations	407	407	407
R^2	0.4321		
Log ps. Likelihood		-94.856922	
Pseudo R^2		0.5230	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-25: Estimation results with interactions 2004 (A-Primary versus A-Primary+soc)

	OLS Coefficients $A_{Primary}$ *(0)*, $A_{Primary + soc}$ *(1)*	**Logit Coefficients** $A_{Primary}$ *(0)*, $A_{Primary + soc}$ *(1)*	**Logit Marginal Effects** $A_{Primary}$ *(0)*, $A_{Primary + soc}$ *(1)*
Bouyei (0/1)	0.021	-0.945	-0.095
	(0.065)	(0.985)	(0.095)
Miao (0/1)	0.256***	1.975***	0.254***
	(0.040)	(0.277)	(0.031)
Tujia (0/1)	0.177**	1.255**	0.144*
	(0.056)	(0.627)	(0.079)
Bouyei*Education	0.006	0.231*	0.024*
	(0.012)	(0.138)	(0.014)
Education (years)	0.006	0.011	0.001
	(0.011)	(0.081)	(0.008)
Male (0/1)	0.185***	2.082***	0.199***
	(0.036)	(0.569)	(0.045)
Age (years)	-0.004**	-0.036***	-0.004***
	(0.001)	(0.009)	(0.001)
Households (*N*)	0.001	0.003	0.001
	(0.001)	(0.002)	(0.001)
County 2 (0/1)	-0.278***	-1.935***	-0.172***
	(0.072)	(0.469)	(0.039)
County 3 (0/1)	-0.205***	-1.832***	-0.173***
	(0.045)	(0.408)	(0.032)
Constant	0.174	-2.235**	
	(0.152)	(1.079)	
Observations	329	329	329
R^2	0.2277		
Log ps. Likelihood		-107.99675	
Pseudo R^2		0.2805	

Note: Standard errors are adjusted to 9 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-26: Estimation results with interactions 2004 (A-Primary versus BC)

	OLS Coefficients	**Logit Coefficients**	**Logit Marginal Effects**
	$A_{Primary}$ (0), BC (1)	*$A_{Primary}$ (0), BC (1)*	*$A_{Primary}$ (0), BC (1)*
Bouyei (0/1)	-0.143	-0.002	-0.001
	(0.136)	(0.890)	(0.058)
Miao (0/1)	-0.147**	-4.228***	-0.147***
	(0.057)	(1.129)	(0.023)
Tujia (0/1)	-0.029	0.216	0.014
	(0.080)	(0.997)	(0.068)
Miao*Education	-0.010	0.298**	0.019**
	(0.008)	(0.137)	(0.008)
Education (years)	0.017**	0.231***	0.015***
	(0.006)	(0.064)	(0.004)
Male (0/1)	0.099**	1.267***	0.083***
	(0.034)	(0.393)	(0.028)
Age (years)	-0.006***	-0.096***	-0.006***
	(0.001)	(0.020)	(0.001)
Households (*N*)	0.001***	0.008***	0.001***
	(0.001)	(0.003)	(0.001)
County 2 (0/1)	0.057	1.318	0.097
	(0.082)	(1.151)	(0.091)
County 3 (0/1)	0.056	0.149	0.01
	(0.110)	(0.798)	(0.052)
Constant	0.188*	-4.016**	
	(0.089)	(1.783)	
Observations	329	329	329
R^2	0.4330		
Log ps. Likelihood		-70.145895	
Pseudo R^2		0.5327	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-27: Estimation results with interactions 2004 (A-Primary versus NA)

	OLS Coefficients	**Logit Coefficients**	**Logit Marginal Effects**
	$A_{Primary}$ (0), NA (1)	$A_{Primary}$ (0), NA (1)	$A_{Primary}$ (0), NA (1)
Bouyei (0/1)	-0.142	0.592	0.042
	(0.139)	(0.941)	(0.063)
Miao (0/1)	-0.203***	-7.554***	-0.207***
	(0.056)	(1.789)	(0.023)
Tujia (0/1)	-0.014	0.017	0.001
	(0.087)	(1.103)	(0.082)
Miao*Education	0.011	0.705***	0.052***
	(0.009)	(0.169)	(0.016)
Education (years)	0.019**	0.189***	0.014***
	(0.008)	(0.056)	(0.004)
Male (0/1)	0.107**	1.257***	0.094***
	(0.041)	(0.361)	(0.029)
Age (years)	-0.007***	-0.094***	-0.007***
	(0.001)	(0.015)	(0.001)
Households (*N*)	0.001***	0.008***	0.001***
	(0.001)	(0.003)	(0.001)
County 2 (0/1)	0.036	1.532	0.127
	(0.101)	(1.196)	(0.104)
County 3 (0/1)	0.084	0.021	0.002
	(0.120)	(0.657)	(0.049)
Constant	0.224**	-3.940**	
	(0.098)	(1.730)	
Observations	351	351	351
R^2	0.4623		
Log ps. Likelihood		-84.09496	
Pseudo R^2		0.5477	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

4.4.2.3 Interim conclusions 2004

After I described the 2004 sample, I followed the five-step procedure described at the beginning of subchapter 4.4 to analyze the data. The effects of the independent variables on the binary dependent variable (*A* versus *NA*) were calculated with linear probability and logit models. The linear probability model served as comparison to the non-linear models. I analyzed the sectors *A* and *NA* and considered the subsectors of *A*, $A_{Primary}$ and $A_{Primary+soc}$, and the subsectors of *NA*, *BC* and *WC*; moreover, I tested whether or not outcome categories could be combined, then estimated MNL and OLOGIT models on the categories considered and compared the results. I additionally tested the IIA assumption in the MNL setting. Finally I included interaction terms between ethnic status and education when the overall model fit improved.

The average education in the 2004 sample is 5.6 years. I find with two-tailed ttests that there is no statistically significant difference in education between Han and the ethnic minorities under consideration, the Miao, Bouyei and Tujia. The marginal effects after logit for all sectors considered in the previous estimations are shown in table 4.28. The marginal effects after logit estimations show that education is crucial for accessing employment in the *NA* sector in general and in particular for accessing employment in the subsectors *BC* and *WC*. Education is, however, not statistically significant for having a *soc*. Yet for the Bouyei an additional year of education from the average increases the probability of having a *soc*. The Miao have generally a higher probability of having a *soc*, but the Miao have much lower probabilities than Han of working in *NA*, *BC* and *WC*. If the Miao, however, have more years of education, they have also a higher probability of working in *NA*.

The results additionally suggest that men have a higher probability of working in *NA* and of having a *soc* than do women. In contrast older individuals have a higher probability of working in sector *A* and a lower probability of finding a *soc* than do younger individuals. The results further suggest that *NA* employment is more probable in larger than in smaller villages; moreover, individuals in counties 2 and 3 have a lower probability of having a *soc* than individuals in county 1.

The MNL results do not support the IIA assumption, and the MNL results are, therefore, not considered. I estimate an OLOGIT model as an alternative to the MNL model, yet the post estimation tests indicated that the data sample was not optimally supported by the OLOGIT model setup; therefore, I did not consider OLOGIT results. The most reliable results for 2004 are, hence, those shown in table 4.28.

Table 4-28: Marginal effects after logit with interactions for all sectors 2004

	Logit Marginal effects *A (0), NA (1)*	**Logit Marginal effects** $A_{Primary}$ *(0),* $A_{Primary+soc}$ *(1)*	**Logit Marginal effects** $A_{Primary}$ *(0), BC (1)*	**Logit Marginal effects** $A_{Primary}$ *(0), NA (1)*
Bouyei (0/1)	0.052	-0.095	-0.001	0.042
	(0.074)	(0.095)	(0.058)	(0.063)
Miao (0/1)	-0.209***	0.254***	-0.147***	-0.207***
	(0.03)	(0.031)	(0.023)	(0.023)
Tujia (0/1)	-0.018	0.144*	0.014	0.001
	(0.064)	(0.079)	(0.068)	(0.082)
Bouyei*Education		0.024*		
		(0.014)		
Miao*Education	0.054***		0.019**	0.052***
	(0.016)		(0.008)	(0.016)
Education (years)	0.016***	0.001	0.015***	0.014***
	(0.003)	(0.008)	(0.004)	(0.004)
Male (0/1)	0.056**	0.199***	0.083***	0.094***
	(0.026)	(0.045)	(0.028)	(0.029)
Age (years)	-0.006***	-0.004***	-0.006***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Households (*N*)	0.001***	0.001	0.001***	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)
County 2 (0/1)	0.182	-0.172***	0.097	0.127
	(0.112)	(0.039)	(0.091)	(0.104)
County 3 (0/1)	0.023	-0.173***	0.01	0.002
	(0.049)	(0.032)	(0.052)	(0.049)
Observations	407	329	329	351

Source: Author's calculation based on CHNS sample.

4.4.3 Econometric analysis 2000

I follow the five-step procedure introduced at the beginning of this subchapter to analyze the 2000 sample. All tables and figures for 2000 are given at the end of each section.

4.4.3.1 Additive specifications

First, I estimate the effects of the independent variables on the binary dependent variable *A* versus *NA* with linear probability and logit models (table 4.29). The estimation results are quite different from those of 2004 (table 4.17); moreover, the results differ between linear probability and logit estimations. Looking at the marginal effects after logit, I find that ethnic status and education are not statistically significant in 2000. In 2000 the age level and geographic characteristics (village size and county of residence) are crucial for determining the probability of working in *A* or *NA*. An increase in the average age level by one year decreases the probability

of working in *NA* by 0.3 %; this is similar to the 2004 results with a marginal effect of 0.6 %. In 2004 and in 2000 residents of a village with more households have a higher probability of working in *NA*. In 2000 residing in counties 2 and 3 increases the probability of working in *NA*, but in 2004 only residing in county 2 increases the probability of working in *NA*.

Second, I consider subcategories and compare the outcomes $A_{Primary}$ and $A_{Primary+soc}$ (table 4.30). While in 2004 the Miao and the Tujia have a higher probability of having a *soc* (table 4.18), in 2000 ethnic status has no impact on the probability of having a *soc*. The linear probability and logit models, however, show that education and residence in county 3 increase the probability of having a *soc*. The marginal effects after logit indicate that in 2000 each additional year of education over the average increases the probability of having a *soc* by 1.9 %. In 2004 education has no impact on having a *soc*, but males and younger individuals have a higher probability of having a *soc*. Age and gender are, however, not statistically significant in 2000. While in 2004 the probability of having a *soc* is lower in county 3 than in county 1, in 2000 the probability of having a *soc* is higher in county 3 than in county 1.

I continue analyzing the subcategories and compare the outcomes $A_{Primary}$ and *BC* (table 4.31). In the linear probability model the independent variables have no statistically significant effects. In the logit model Bouyei, Miao and age have statistically significant negative effects on the probability of working in *BC*. In contrast residence in larger villages and in counties 2 and 3 increase the probability of working in *BC*. The marginal effects after logit are not significant for the variable Bouyei, but are significant for the variable Miao. The probability of working in *BC* decreases by 5.9 % for the Miao in comparison with Han; moreover, it is notable that with every year over the average age, the probability of working in the *BC* sector declines by 0.3 %. Additionally male status and education are more important for working in *BC* positions in 2004 than in 2000.

The next step is to compare $A_{Primary}$ and *WC* in the 2000 sample. Due to the comparatively small number of observations in *WC* and the related computational drawbacks, I combine, as in the 2004 estimation, *BC* and *WC* sectors and, therefore, consider the *NA* sector (table 4.32). The results are not coherent between the linear probability and logit models. In the linear probability model, education and county 2 have significant positive effects on the probability of working in *NA*. The logit results, however, show that particularly younger age, residence in larger villages and residence in counties 2 and 3 increase the probability of working in the *NA* sector. The marginal effects after logit indicate that for each additional year of age from the average, the probability of working in *NA* increases by 0.3 %.

In contrast the 2004 results of the $A_{Primary}$ and *NA* comparison show a statistically significant negative effect for Miao status and a statistically significant positive effect for education and male status but insignificant county effects (table 4.20). The individual characteristics, thus, have a stronger influence on the outcomes

$A_{Primary}$ and *NA* in 2004 than in 2000, but the geographic location is less important for working in *NA* in 2004 than in 2000.

The third step is to consider all outcomes in one single model. Again I use two different approaches, the MNL model and the OLOGIT model. The next step is to analyze the outcome categories and to apply Wald tests to find out whether outcome categories can be collapsed. I then estimate the linear probability model, the MNL model and the OLOGIT model. I also test the IIA assumption after the MNL estimation. Finally the overall model fit and the prediction power of the MNL and the OLOGIT models must be compared to choose the model which best represents the data.

I analyze which outcome categories, $A_{Primary}$ (0), $A_{Primary+soc}$ (1), *BC* (2) and *WC* (3) can be collapsed. As in the 2004 analysis, I use Wald tests to combine outcome categories (LONG and FREESE, 2003, p. 204) (see table 4.33). As in the 2004 sample, the high p-value for the comparison of the second and third category indicates that these two outcomes can be collapsed into one single category. In the MNL and OLOGIT estimations I, thus, consider three outcomes, $A_{Primary}$ (0), $A_{Primary+soc}$ (1) and *NA* (2), where *NA* combines the categories *BC* and *WC*.

The estimation results show that ethnic status is not statistically significant in any of the models (table 4.34). In 2000 the statistically significant variables in the linear probability model are education, county 2 and county 3, which all have positive effects. In 2004 education, gender, age and village size have statistically significant effects, while the counties have no statistically significant effect (table 4.22).

The MNL estimation in 2000 has two equations. One equation compares the outcomes $A_{Primary}$ and $A_{Primary+soc}$, and the other equation compares the outcomes $A_{Primary}$ and *NA*. The first equation indicates that more years of education and residence in county 3 increase the probability of having a *soc*. The magnitudes of the coefficients are the same as in the binary comparison (table 4.30). In the MNL estimations Miao, Tujia and male status have a statistically significant positive effect on having a *soc* in 2004, but not in 2000. Age changed from an insignificant value in 2000 to a statistically significant positive value in 2004. While county 3 increases the probability of having a *soc* in 2000, county 1 and larger villages increase the probability of having a *soc* in 2004. The other equation in the MNL model compares the outcomes $A_{Primary}$ and *NA* (table 4.34). The results indicate that in larger villages and in counties 2 and 3 the probability of working in *NA* increases. The results, moreover, show that younger individuals have a higher probability of working in *NA*. The magnitudes of the coefficients are the same as in the binary comparison (table 4.32). The 2000 OLOGIT results show that more years of education, lower age, residence in larger villages and residence in counties 2 and 3 all increase the probability of working in *NA* (table 4.34). In 2004 male status also significantly increases the probability of working in *NA*, but county 2 is insignificant (table 4.22).

Again the MNL results should be considered with caution. I find that the IIA assumption is violated (table 4.35). This means that estimation results are not independent from irrelevant alternatives in 2000. Following the estimation procedure, the next step is to compare the prediction power of the MNL and OLOGIT models (LONG and FREESE, 2003, p. 211-212). I find that the categories $A_{Primary}$ after MNL and $A_{Primary}$ after OLOGIT, *NA* after MNL and *NA* after OLOGIT have correlations of 0.98; however, the categories $A_{Primary+soc}$ after MNL and $A_{Primary+soc}$ after OLOGIT only have a correlation of 0.72. As in the 2004 sample I also find that the predictions after OLOGIT are suddenly truncated in the 2000 sample, which may not reasonably explain the real data sample (figure 4-7). The test results, thus, indicate that it is preferable to use the MNL rather than the OLOGIT model. The violation of the IIA assumption in the MNL model, however, requires the usage of alternative specific models. Alternative specific variables which are essential in alternative specific models are, however, not available in the CHNS dataset; therefore, in the 2000 analysis I cannot consider the multinomial models. The 2000 conclusions are drawn from binary logit estimations.

Table 4-29: Estimation results 2000 (A versus NA)

	OLS Coefficients *A (0), NA (1)*	**Logit Coefficients** *A (0), NA (1)*	**Logit Marginal Effects** *A (0), NA (1)*
Bouyei (0/1)	-0.061	-0.625	-0.037
	(0.197)	(0.574)	(0.036)
Miao (0/1)	-0.053	0.001	0.001
	(0.058)	(0.584)	(0.034)
Tujia (0/1)	0.072	0.761**	0.047
	(0.062)	(0.343)	(0.031)
Education (years)	0.021**	0.085	0.005
	(0.009)	(0.106)	(0.006)
Male (0/1)	0.008	0.621	0.035
	(0.037)	(0.522)	(0.032)
Age (years)	-0.001	-0.045***	-0.003***
	(0.001)	(0.008)	(0.001)
Households (*N*)	0.001	0.008***	0.001***
	(0.001)	(0.002)	(0.001)
County 2 (0/1)	0.244*	6.010***	0.485***
	(0.130)	(2.205)	(0.079)
County 3 (0/1)	0.402	7.645***	0.492***
	(0.251)	(2.357)	(0.039)
Constant	-0.364	-10.342***	
	(0.210)	(2.691)	
Observations	444	444	444
R^2	0.3527		
Log ps. Likelihood		-91.58616	
Pseudo R^2		0.5182	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-30: Estimation results 2000 (A-Primary versus A-Primary+soc)

	OLS Coefficients $A_{Primary}$ *(0)*, $A_{Primary + soc}$*(1)*	**Logit Coefficients** $A_{Primary}$ *(0)*, $A_{Primary + soc}$ *(1)*	**Logit Marginal Effects** $A_{Primary}$ *(0)*, $A_{Primary + soc}$ *(1)*
Bouyei (0/1)	-0.070	-0.890	-0.077
	(0.064)	(0.667)	(0.061)
Miao (0/1)	-0.028	-0.205	-0.017
	(0.082)	(1.042)	(0.082)
Tujia (0/1)	0.013	0.233	0.021
	(0.091)	(1.082)	(0.101)
Education (years)	0.019***	0.226***	0.019***
	(0.005)	(0.074)	(0.005)
Male (0/1)	0.049	0.596	0.049
	(0.040)	(0.410)	(0.036)
Age (years)	-0.000	-0.004	-0.001
	(0.001)	(0.010)	(0.001)
Households (*N*)	-0.000	-0.000	-0.001
	(0.000)	(0.001)	(0.001)
County 2 (0/1)	0.031	0.360	0.032
	(0.076)	(0.853)	(0.083)
County 3 (0/1)	0.115***	1.515***	0.142***
	(0.020)	(0.200)	(0.023)
Constant	-0.022	-4.051***	
	(0.067)	(0.915)	
Observations	376	376	376
R^2	0.0844		
Log ps. Likelihood		-110.37637	
Pseudo R^2		0.1338	

Note: Standard errors are adjusted to 10 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table-31: Estimation results 2000 (A-Primary versus BC)

	OLS Coefficients	**Logit Coefficients**	**Logit Marginal Effects**
	$A_{Primary}$ (0), BC (1)	*$A_{Primary}$ (0), BC (1)*	*$A_{Primary}$ (0), BC (1)*
Bouyei (0/1)	-0.099	-1.305*	-0.074
	(0.217)	(0.748)	(0.049)
Miao (0/1)	-0.096	-1.488**	-0.059**
	(0.072)	(0.605)	(0.024)
Tujia (0/1)	0.073	0.620	0.034
	(0.088)	(0.743)	(0.044)
Education (years)	0.014	0.019	0.001
	(0.009)	(0.089)	(0.005)
Male (0/1)	0.014	0.872	0.044
	(0.047)	(0.618)	(0.039)
Age (years)	-0.002	-0.051***	-0.003***
	(0.001)	(0.010)	(0.001)
Households (*N*)	0.000	0.007***	0.001***
	(0.000)	(0.002)	(0.001)
County 2 (0/1)	0.186	6.620**	0.488***
	(0.124)	(2.654)	(0.079)
County 3 (0/1)	0.356	8.750***	0.495***
	(0.241)	(2.974)	(0.039)
Constant	-0.200	-10.228***	
	(0.223)	(3.099)	
Observations	383	383	383
R^2	0.2771		
Log ps. Likelihood		-72.045949	
Pseudo R^2		0.4947	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-32: Estimation results 2000 (A-Primary versus NA)

	OLS Coefficients $A_{Primary}$ *(0), NA (1)*	**Logit Coefficients** $A_{Primary}$ *(0), NA (1)*	**Logit Marginal Effects** $A_{Primary}$ *(0), NA (1)*
Bouyei (0/1)	-0.073	-0.884	-0.056
	(0.197)	(0.630)	(0.042)
Miao (0/1)	-0.060	-0.229	-0.013
	(0.066)	(0.819)	(0.046)
Tujia (0/1)	0.066	0.498	0.032
	(0.073)	(0.583)	(0.041)
Education (years)	0.023**	0.105	0.006
	(0.009)	(0.110)	(0.006)
Male (0/1)	0.021	0.669	0.04
	(0.045)	(0.574)	(0.038)
Age (years)	-0.002	-0.044***	-0.003***
	(0.001)	(0.008)	(0.001)
Households (*N*)	0.001	0.007***	0.001***
	(0.000)	(0.002)	(0.001)
County 2 (0/1)	0.257*	5.945***	0.478***
	(0.132)	(2.213)	(0.086)
County 3 (0/1)	0.416	7.505***	0.501***
	(0.242)	(2.372)	(0.042)
Constant	-0.340	-9.820***	
	(0.206)	(2.569)	
Observations	404	404	404
R^2	0.3785		
Log ps. Likelihood		-86.207539	
Pseudo R^2		0.5292	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-33: Wald tests for combining outcome categories 2000

Ho: All coefficients except intercepts associated with given pair of outcomes are 0 (i.e., categories can be collapsed).

Categories tested	Chi 2	df	P>chi2
0-1	26.931	9	0.001
0-2	66.358	9	0.000
0-3	52.620	9	0.000
1-2	37.819	9	0.000
1-3	32.403	9	0.000
2-3	15.522	9	0.078

Source: Author's calculation based on CHNS sample, test commands and results based on LONG and FREESE (2003, p. 204).

Table 4-34: Multinomial estimation results 2000

	OLS Coefficients	MNL Coefficients		OLOGIT Coefficients
	$A_{Primary}$ (0), $A_{Primary+soc}$ (1), NA (2)	$A_{Primary}$ (0), $A_{Primary+soc}$ (1)	$A_{Primary}$ (0), NA (2)	$A_{Primary}$ (0), $A_{Primary+soc}$ (1), NA (2)
Bouyei (0/1)	-0.143	-0.843	-0.798	-0.802
	(0.343)	(0.719)	(0.536)	(0.577)
Miao (0/1)	-0.112	-0.282	-0.056	-0.033
	(0.139)	(1.057)	(0.754)	(0.723)
Tujia (0/1)	0.143	0.086	0.790	0.780
	(0.147)	(1.124)	(0.499)	(0.647)
Education(years)	0.051***	0.206***	0.121	0.184**
	(0.016)	(0.067)	(0.113)	(0.084)
Male (0/1)	0.064	0.655	0.720	0.527
	(0.084)	(0.399)	(0.531)	(0.366)
Age (years)	-0.003	-0.006	-0.044***	-0.021***
	(0.003)	(0.009)	(0.008)	(0.008)
Households (*N*)	0.001	-0.001	0.008***	0.005***
	(0.001)	(0.001)	(0.002)	(0.002)
County 2 (0/1)	0.488*	0.423	6.100***	3.544**
	(0.255)	(0.843)	(2.229)	(1.804)
County 3 (0/1)	0.813*	1.312***	7.921***	5.395***
	(0.441)	(0.172)	(2.360)	(2.025)
Constant	-0.642	-3.665***	-10.554***	
	(0.376)	(0.942)	(2.750)	
Cut 1				7.550***
				(2.305)
Cut 2				8.582***
				(2.367)
Observations	444	444	444	444
R^2	0.3651			
Log ps. Likelihood		-202.3302		-215.15876
Pseudo R^2		0.3628		0.3224

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

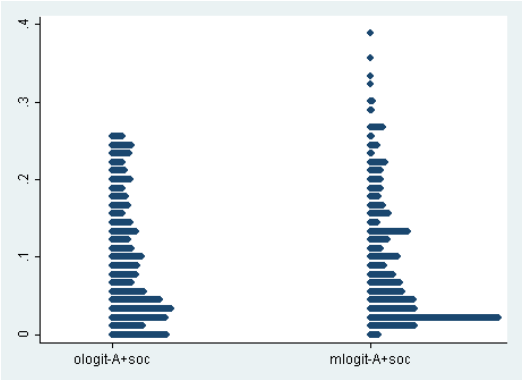
Table 4-35: Wald test after suest 2000

H_0 = no difference between estimators of full and restricted model *H_1 = difference between estimators of full and restricted model*		
	A+soc	***NA***
Model Restrictions		
Exclusion *A+soc*		chi2(10) = 122.74 Prob > chi2 = 0.001
Exclusion *NA*	chi2(10) = 3977.62 Prob > chi2 = 0.001	

Agriculture is base group. Standard errors are based on cluster robust variance estimator.

Source: Author's calculation based on CHNS sample.

Figure 4-7: Plot to compare predicted probabilities of sector A-Primary+soc in MNL and OLOGIT models 2000



Source: Author's calculation based on CHNS sample.

4.4.3.2 Interaction Terms

The next step is to include interaction terms between ethnic status and education in the binary models. I use the binary comparisons of *A* versus *NA*, $A_{Primary}$ versus $A_{Primary+soc}$ and $A_{Primary}$ versus *NA*. While there are some statistically significant interaction effects in 2004, there are only statistically insignificant interaction effects in 2000. The plots of the predicted probability and education, however, show that there are some differences among ethnic groups.

The inclusion of all three interaction terms between ethnic status and education deteriorates model fit when comparing the outcomes *A* and *NA*. The significant Tujia coefficient becomes insignificant when all interactions are included and when the interaction term Tujia*Education and Miao*Education are included. The major factors which determine a higher probability of working in *NA* in 2000 depend on younger age, residence in a larger village and residence in counties 2 and 3 (table 4.29). I postulate that the variable Tujia is nonlinear over observations as

the marginal effects are insignificant and as the inclusion of interaction terms result in changes of the coefficient. The interpretation of the Tujia effect should, therefore, be treated with caution.

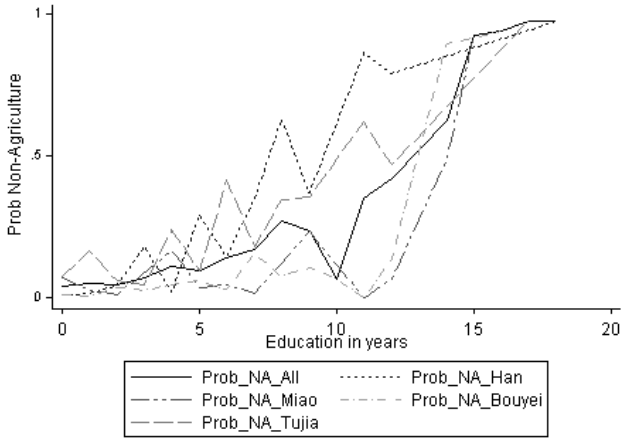
I observe that in all ethnic groups individuals with more years of education have a higher probability of working in *NA* (figure 4-8). For individuals with education between eight and fourteen years, ethnic status has a strong influence on the probability of working in *NA*. It can be seen that Han and Tujia have a higher probability of working in *NA* in these education cohorts, but that Miao and Bouyei have a lower probability of working in *NA* in these years.

The next step is to analyze interactions between ethnic status and education by comparing $A_{Primary}$ and $A_{Primary+soc}$. The inclusion of all three interaction terms deteriorates model fit when comparing the outcomes $A_{Primary}$ and $A_{Primary+soc}$. Regardless of ethnic status, individuals with more years of education have a higher probability of having a *soc* (figure 4-9). There are, however, some differences between ethnic groups after six years of education. The Miao have a lower probability of having a *soc* than the average from around six to fourteen years of education. Han are above average in these educational cohorts. The Bouyei have a lower probability of having a *soc* after ten years of education, and the Tujia are scattered around average levels.

The next step is to compare the sectors $A_{Primary}$ and *BC*. The plot of predicted probability and education shows that there are stark differences in the probability of working in *BC* depending on education and ethnic status (figure 4-10). It is clear that with more years of education, the probability of working in *BC* positions increases. Han and the Tujia have a higher probability of working in *BC* than the average in all educational cohorts, but the Bouyei and the Miao have a lower probability of working in BC than the average in most of the cases. The inclusion of interaction terms between ethnic status and education, however, also deteriorates model fit when comparing the sectors $A_{Primary}$ and *BC*.

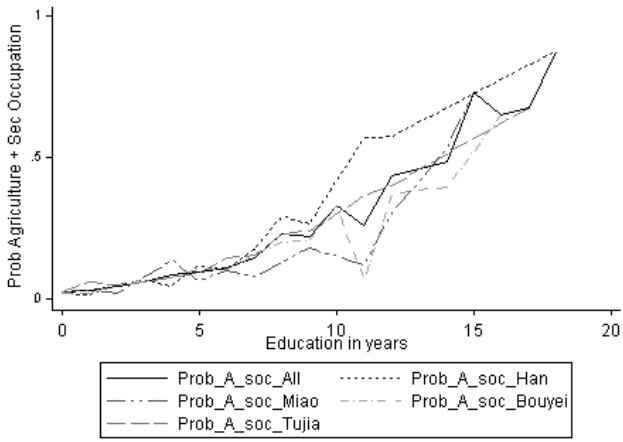
The final step in the 2000 analysis is to compare the sectors $A_{Primary}$ and *NA*. The inclusion of interaction terms between ethnic status and education also deteriorates model fit in this specification. It is, however, true that with more years of education, the probability of working in *NA* increases (figure 4-11). For some years of education, the Miao and the Bouyei have a below average probability of working in *NA*, while Han and the Tujia have an above average probability of working in *NA*. The estimation results after logit, however, show that neither ethnic status nor years of education is statistically significant (table 4.32).

Figure 4-8: Plot of predicted probabilities and years of education by ethnicity for A versus NA 2000



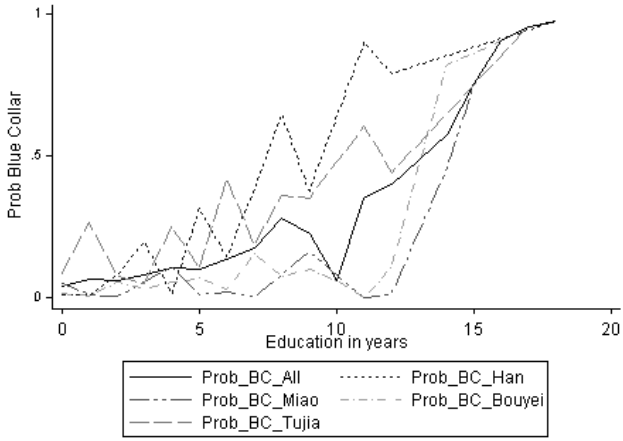
Source: Author's calculation based on CHNS sample.

Figure 4-9: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus A-Primary+soc 2000



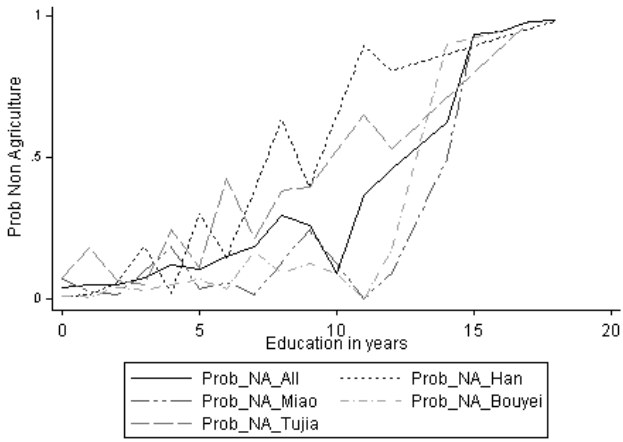
Source: Author's calculation based on CHNS sample.

Figure 4-10: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus BC 2000



Source: Author's calculation based on CHNS sample.

Figure 4-11: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus NA 2000



Source: Author's calculation based on CHNS sample.

4.4.3.3 Interim conclusions 2000

I analyzed the sample with the five-step procedure as described at the beginning of this chapter. As in the 2004 sample, the MNL results of the 2000 estimations also do not support the IIA assumption; thus, I do not consider the MNL model results. The post estimation tests after the OLOGIT estimation, moreover, indicate that the data sample is not optimally supported by the OLOGIT model setup; therefore, I do not consider the OLOGIT results of this sample. The most reliable results for the 2000 sample are, thus, the binary logit results (table 4.36).

In 2000 the average education is 5.2 years. I find no statistically significant difference in education between Han and ethnic minorities; therefore, interaction terms between ethnic status and education are also insignificant. The marginal effects of single coefficients, however, show some significant effects in 2000. The only significant ethnic effect is for individuals with Miao status. In the $A_{Primary}$ versus *BC* model being Miao has a statistically significant negative effect on the probability of working in *BC*. Education is, however, only statistically significant in the comparison of $A_{Primary}$ and $A_{Primary+soc}$; more years of education positively influence the probability of having a *soc* alongside work in agriculture. Younger individuals, moreover, have a higher probability of working in *NA* in general and *BC* in particular. It is crucial that in 2000 no gender differences

Table 4-36: Marginal effects after logit for all sectors 2000

	Logit Marginal Effects *A (0), NA (1)*	**Logit Marginal Effects** $A_{Primary}$ *(0),* $A_{Primary+soc}$ *(1)*	**Logit Marginal Effects** $A_{Primary}$ *(0), BC (1)*	**Logit Marginal Effects** $A_{Primary}$ *(0), NA (1)*
Bouyei (0/1)	-0.037	-0.077	-0.074	-0.056
	(0.036)	(0.061)	(0.049)	(0.042)
Miao (0/1)	0.001	-0.017	-0.059**	-0.013
	(0.034)	(0.082)	(0.024)	(0.046)
Tujia (0/1)	0.047	0.021	0.034	0.032
	(0.031)	(0.101)	(0.044)	(0.041)
Education (years)	0.005	0.019***	0.001	0.006
	(0.006)	(0.005)	(0.005)	(0.006)
Male (0/1)	0.035	0.049	0.044	0.04
	(0.032)	(0.036)	(0.039)	(0.038)
Age (years)	-0.003***	-0.001	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Households (*N*)	0.001***	-0.001	0.001***	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)
County 2 (0/1)	0.485***	0.032	0.488***	0.478***
	(0.079)	(0.083)	(0.079)	(0.086)
County 3 (0/1)	0.492***	0.142***	0.495***	0.501***
	(0.039)	(0.023)	(0.039)	(0.042)
Observations	444	376	383	404

Source: Author's calculation based on CHNS sample.

are observable. For the *NA* and *BC* sector the geographic characteristics, however, have significantly positive effects on working in *NA* and *BC*. Residence in larger villages and in county 2 and 3 increase the probability of working in *NA* and *BC*. Residence in county 3, moreover, results in a higher probability of having a *soc* alongside agricultural work.

4.4.4 Econometric analysis 1997

To analyze the 1997 sample I use the five-step estimation procedure introduced at the beginning of this subchapter. All intermediary results tables and figures are given at the end of each section.

4.4.4.1 Additive specifications

First, I calculate the effects of the independent variables on the binary dependent variable *A* versus *NA* with linear probability and logit models (table 4.37). The magnitudes of the coefficients are the same for the linear probability and for the logit model, except the county 2 coefficient, which is not significant in the linear model. The marginal effects after logit indicate that the Miao have a 4.1 % lower probability of working in *NA* than do Han. Each additional year of education above the average, however, increases the probability of working in *NA* by 0.9 %. The probability of working in *NA* is, moreover, higher in larger villages and in county 2.

In 2004 in addition to these variables, males and younger individuals have a higher probability of working in *NA* (table 4.17). In 2000 younger age, larger villages and counties 2 and 3 increase the probability of working in *NA* (table 4.29). The Miao have a lower probability of working in *NA* than do Han in 2004 and 1997, but not in 2000. Similarly education positively influences the probability of working in *NA* in 2004 and 1997, but not in 2000. Age, however, increases in importance in the 2000 and 2004 samples, but is insignificant in 1997. While there are no gender differences in 1997 and 2000, males have a significantly higher probability of working in *NA* in 2004. The probability of working in *NA* also significantly increases in larger villages and in county 2 in all samples; county 3, however, has a statistically significant positive value only in 2000.

Second, I consider subcategories and compare the outcomes $A_{Primary}$ and $A_{Primary+soc}$ (table 4.38). The magnitudes of the coefficients are the same in the linear probability model and in the logit model. While ethnic status is insignificant, individuals with more years of education, with male status, with lower age levels and individuals living in counties 2 and 3 have a higher probability of having a *soc*. Each additional year of education above the average increases the probability of having a *soc* by 1.5 %. Being male rather than female increases the probability of having a *soc* by 6.8 %. In contrast each additional year of age above the average decreases the probability of having a *soc* by 0.3 %.

In the 2000 results only more years of education and residence in county 3 increase the probability of having a *soc*. From 1997 to 2000 age and gender effects declined, yet in 2004 there are not only age and gender effects, but also ethnic effects. In 2004 both the Tujia and the Miao have a higher probability of having a *soc* than do Han. The county effects are the opposite in 1997 and 2004: in 1997 residence in counties 2 and 3 increases the probability of having a *soc* but in 2004 residence in counties 2 and 3 decreases the probability of having a *soc*.

I continue to analyze the subcategories within the sectors *A* and *NA* and compare the outcomes $A_{Primary}$ and *BC* (table 4.39). The magnitudes of the coefficients in the linear probability model and in the logit model differ in the variables education, age and county 2. Education has a statistically significant positive effect on the probability of working in *BC* positions in the linear probability model but is insignificant in the logit model. Younger age and residence in county 2 have a statistically significant positive effect on the probability of working in *BC* positions in the logit model but are insignificant in the linear probability model. In both models Bouyei status has a statistically significant positive effect on the probability of working in *BC* positions, while Miao status and residence in smaller villages have statistically significant negative effects on the probability of working in *BC* positions. The average marginal effects after logit are, however, insignificant for Bouyei but significant for Miao. The Miao have a 5.5 % lower probability of working in *BC* positions than do Han. Each additional year of age above the average decreases the probability of working in *BC* by 0.2 %.

The marginal effects are almost the same in 2000 (table 4.31). The only exception is county 3, which has a statistically significant positive effect on the probability of working in *BC* in 2000, but not in 1997. In 2004 more years of education and male status increase the probability of working in *BC* positions, while the county effects are insignificant. In 2004 Miao and older individuals have a lower probability of working in *BC* positions; in contrast better educated individuals, males and residents of larger villages have a higher probability of working in *BC* positions. The comparison of $A_{Primary}$ and *BC* also suggest that gender differences increase over time, but that county effects disappear over time. Education has an increasingly positive effect for working in *BC* positions between 1997 and 2004.

The next step is to compare $A_{Primary}$ and *WC*. As in the 2000 and 2004 estimations, because of the relatively low number of observations in *WC* and the related computational drawbacks, I combine *BC* and *WC*; therefore, I only consider the *NA* sector. The magnitudes of the estimation results show that the linear probability model and the logit model are the same for all coefficients except for county 2 (table 4.40). The results indicate that the Bouyei have a higher probability of working in *NA* than do Han. The average marginal effects for the Bouyei are, however, not statistically significant. In contrast the Miao have a lower probability of working in *NA* than do Han. Being Miao rather than Han decreases the probability of working in *NA* by 4.4 %. Education, however, has a positive effect on the

probability of working in *NA*. Each additional year of education above the average increases the probability of working in *NA* by 1 %. Additionally residence in larger villages and in county 2 has statistically significant positive effects on the probability of working in *NA*.

The 1997 results differ from the 2000 results. In 2000 ethnic effects and education are not statistically significant. Younger individuals, however, have a higher probability of working in *NA*. Residences in larger villages and in county 2 increase the probability of working in *NA* in both samples. In 2000 county 3, moreover, has a statistically significant positive effect on the probability of working in *NA*. Comparing the 1997 and 2004 samples, in both samples the Miao have a significantly negative probability of working in *NA* than do Han. More years of education and residence in larger villages, moreover, significantly increase the probability of working in *NA* in 1997 and 2004. In 1997 residence in county 2 has a statistically significant positive effect on working in *NA*, but in 2004 the county effects are insignificant. While in 1997 gender and age are not statistically significant, in 2004 male status positively influences the probability of working in *NA*, and older age negatively influences the probability of working in *NA*.

I analyze now which outcome categories can be collapsed (LONG and FREESE, 2003, p. 204). The outcome categories are again $A_{Primary}$ (0), $A_{Primary+soc}$ (1), *BC* (2) and *WC* (3). In this case the low p-values suggest that none of the outcomes can be collapsed into one single category (table 4.41); thus, I use all outcomes $A_{Primary}$ (0), $A_{Primary+soc}$ (1), *BC* (2) and *WC* (3) in the multinomial models. This means that there will be three instead of two equations in the MNL model. In the OLOGIT model I assume that *WC* is ranked higher than the other categories.

I observe that the magnitudes of the coefficients differ between the models (table 4.42). In the linear probability case Miao status has a statistically significant negative effect, and education and residence in larger villages have statistically significant positive effects on occupational outcomes. The 2000 results also indicate that education has a statistically significant positive effect on occupational outcomes (table 4.34). The 2000 results further show that the village effect is not statistically significant but that residence in both counties 2 and 3 has statistically significant positive effects on occupational outcomes. The 2004 results also show statistically significant positive effects for education and for residence in larger villages (table 4.22). The 2004 results, moreover, show a statistically significant positive effect for males and a statistically significant negative effect for age, but insignificant county effects.

The MNL results include three equations in 1997. First, I compare the outcomes $A_{Primary}$ (0) and $A_{Primary+soc}$ (1). I find that ethnicity has no statistically significant effect in the comparison of $A_{Primary}$ (0) and $A_{Primary+soc}$ (1). More years of education and male status, however, have a statistically significant effect on the probability of having a *soc* alongside agriculture. In contrast older age decreases the probability

of having a *soc* alongside agriculture. Additionally residence in counties 2 and 3 significantly increases the probability of having a *soc* alongside agriculture. In 2000 the same comparison shows that only education and residence in county 3 increase the probability of having a *soc* alongside agriculture. In the comparison of $A_{Primary}$ (0) and $A_{Primary+soc}$ (1) in the MNL setting in 2004, there are significantly positive ethnic effects for the Miao and the Tujia. This means that both the Miao and the Tujia have a higher probability of having a *soc* alongside agriculture than do Han. In line with the 2004 results, the 1997 results also show that male status has a statistically significant positive effect, and older age has a statistically significant negative effect on the probability of having a *soc* alongside agriculture. The county effects in 1997 and in 2004 are, however, the opposite. In 1997 residence in counties 2 and 3 significantly increase the probability of having a *soc* alongside *A*, but in 2004 residence in counties 2 and 3 significantly decrease the probability of having a *soc* alongside *A*.

The next MNL estimation compares the outcomes $A_{Primary}$ (0) and *BC* (1) in 1997. In 2000 and 2004 *BC* and *WC* are combined in the category *NA*. A direct comparison of the equations $A_{Primary}$ (0) versus *BC* (1) and $A_{Primary}$ (0) versus *WC* (1) is, therefore, not suitable among samples. There are significant effects for Bouyei and Miao in 1997: Bouyei have a statistically significant higher probability of working in *BC* than do Han, and Miao have a statistically significant lower probability of working in *BC* than do Han. There are, more-over, statistically significant age and village effects. Older age decreases the probability of working in *BC*, and residence in a larger village increases the probability of working in *BC*. Residence in county 2 also has a statistically significant positive effect on the probability of working in *BC* in comparison to residence in county 1. In contrast in the comparison of $A_{Primary}$ (0) and *WC* (1), there are no statistically significant ethnic effects. Having more years of education, living in larger villages and living in counties 2 and 3 each increase the probability of working in *WC*. The OLOGIT results show statistically significant positive ethnic effects. The Bouyei and the Tujia have a higher probability of working in *NA* positions than do Han. Having more years of education, living in larger villages and living in counties 2 and 3 each increase the probability of working in *NA* positions in the OLOGIT model.

The IIA assumption is violated in 1997, just as it was in 2000 and 2004, which indicated that MNL estimation results are not independent from irrelevant alternatives (table 4.43). The 1997 results after MNL estimation, therefore, have to be considered with caution. The next step is to compare the prediction power of the MNL and OLOGIT models (LONG and FREESE, 2003, p. 211-212) for the 1997 sample. The categories $A_{Primary}$ after MNL and $A_{Primary}$ after OLOGIT have a correlation of 0.98. The categories *BC* after MNL and *BC* after OLOGIT have a correlation of 0.93. The category $A_{Primary+soc}$ after MNL and $A_{Primary+soc}$ after OLOGIT, however, has a correlation of only 0.67 (figure 4-12). The category *WC* after MNL and *WC* after OLOGIT has a correlation of 0.83 (figure 4-13).

Like LONG and FREESE (2003, p. 212) I also find that predictions after OLOGIT are suddenly truncated for both categories $A_{Primary+soc}$ (figure 4-12) and *WC* (figure 4-13). The OLOGIT results, therefore, may not reasonably explain the real data sample.

The test results of 1997, thus, indicate that it is preferable to use the MNL model rather than the OLOGIT model. The violation of the IIA assumption in the MNL model, however, requires alternative specific models; unfortunately, the necessary alternative specific variables are not available in the CHNS dataset; therefore, as in 2004 and 2000, I do not consider multinomial models in 1997. I consider the estimation results of the binary logit models in 1997, in 2000, in 2004 and in the overall conclusions.

Table 4-37: Estimation results 1997 (A versus NA)

	OLS Coefficients *A (0), NA (1)*	**Logit Coefficients** *A (0), NA (1)*	**Logit Marginal Effects** *A (0), NA (1)*
Bouyei (0/1)	0.213	4.105	0.234
	(0.117)	(2.545)	(0.155)
Miao (0/1)	-0.120***	-1.165***	-0.041***
	(0.033)	(0.288)	(0.017)
Tujia (0/1)	-0.025	-0.151	-0.006
	(0.029)	(0.251)	(0.010)
Education (years)	0.014**	0.210***	0.009***
	(0.005)	(0.055)	(0.003)
Male (0/1)	-0.022	-0.453	-0.02
	(0.038)	(0.719)	(0.029)
Age (years)	0.000	-0.020	-0.001
	(0.001)	(0.021)	(0.001)
Households (*N*)	0.002***	0.017***	0.001***
	(0.000)	(0.004)	(0.001)
County 2 (0/1)	0.096	2.985**	0.196**
	(0.155)	(1.406)	(0.094)
County 3 (0/1)	0.115	1.266	0.062
	(0.131)	(2.094)	(0.113)
Constant	-0.545***	-11.415***	
	(0.149)	(2.930)	
Observations	599	599	599
R^2	0.639		
Log ps. Likelihood		-95.265711	
Pseudo R^2		0.673	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-38: Estimation results 1997 (A-Primary versus A-Primary+soc)

	OLS Coefficients $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*	**Logit Coefficients** $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*	**Logit Marginal Effects** $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*
Bouyei (0/1)	-0.036	0.002	0.001
	(0.025)	(0.449)	(0.031)
Miao (0/1)	0.009	0.673	0.053
	(0.030)	(0.737)	(0.063)
Tujia (0/1)	0.031	0.897	0.069
	(0.045)	(0.637)	(0.055)
Education (years)	0.015***	0.220***	0.015***
	(0.004)	(0.080)	(0.005)
Male (0/1)	0.065*	1.058*	0.068**
	(0.033)	(0.565)	(0.033)
Age (years)	-0.002*	-0.043***	-0.003***
	(0.001)	(0.015)	(0.001)
Households (*N*)	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
County 2 (0/1)	0.094***	2.086***	0.172***
	(0.019)	(0.541)	(0.049)
County 3 (0/1)	0.121***	2.606***	0.182***
	(0.019)	(0.590)	(0.044)
Constant	0.048	-4.874***	
	(0.061)	(0.934)	
Observations	485	485	485
R^2	0.1351		
Log ps. Likelihood		-111.41551	
Pseudo R^2		0.2675	

Note: Standard errors are adjusted to 9 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-39: Estimation results 1997 (A-Primary versus BC)

	OLS Coefficients	**Logit Coefficients**	**Logit Marginal Effects**
	$A_{Primary}$ (0), BC (1)	*$A_{Primary}$ (0), BC (1)*	*$A_{Primary}$ (0), BC (1)*
Bouyei (0/1)	0.233*	4.431*	0.233
	(0.126)	(2.371)	(0.15)
Miao (0/1)	-0.144***	-1.758***	-0.055**
	(0.028)	(0.440)	(0.023)
Tujia (0/1)	-0.060	-0.586	-0.022
	(0.036)	(0.420)	(0.016)
Education (years)	0.009*	0.136	0.005
	(0.005)	(0.085)	(0.003)
Male (0/1)	-0.016	-0.374	-0.015
	(0.037)	(0.735)	(0.027)
Age (years)	-0.001	-0.045**	-0.002*
	(0.001)	(0.021)	(0.001)
Households (*N*)	0.002***	0.017***	0.001***
	(0.000)	(0.004)	(0.001)
County 2 (0/1)	0.112	2.890**	0.189*
	(0.158)	(1.424)	(0.103)
County 3 (0/1)	0.108	1.152	0.050
	(0.124)	(1.926)	(0.088)
Constant	-0.489***	-10.278***	
	(0.152)	(2.674)	
Observations	525	525	525
R^2	0.607		
Log ps. Likelihood		-78.209668	
Pseudo R^2		0.666	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-40: Estimation results 1997 (A-Primary versus NA)

	OLS Coefficients $A_{Primary}$ *(0), NA (1)*	**Logit Coefficients** $A_{Primary}$ *(0), NA (1)*	**Logit Marginal Effects** $A_{Primary}$ *(0), NA (1)*
Bouyei (0/1)	0.219*	4.086*	0.230
	(0.117)	(2.425)	(0.145)
Miao (0/1)	-0.130***	-1.188***	-0.044***
	(0.030)	(0.282)	(0.017)
Tujia (0/1)	-0.037	-0.342	-0.015
	(0.026)	(0.308)	(0.013)
Education (years)	0.016***	0.224***	0.01***
	(0.005)	(0.049)	(0.003)
Male (0/1)	-0.016	-0.253	-0.011
	(0.036)	(0.661)	(0.028)
Age (years)	-0.000	-0.025	-0.001
	(0.001)	(0.020)	(0.001)
Households (*N*)	0.002***	0.016***	0.001***
	(0.000)	(0.004)	(0.001)
County 2 (0/1)	0.122	3.273**	0.225**
	(0.154)	(1.354)	(0.092)
County 3 (0/1)	0.120	1.404	0.073
	(0.128)	(2.012)	(0.113)
Constant	-0.541***	-11.220***	
	(0.146)	(2.747)	
Observations	553	553	553
R^2	0.647		
Log ps. Likelihood		-90.075565	
Pseudo R^2		0.6799	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

Table 4-41: Wald tests for combining outcome categories 1997

Ho: All coefficients except intercepts associated with given pair of outcomes are 0 (i.e., categories can be collapsed).

Categories tested	**Chi 2**	**df**	**P>chi2**
0-1	46.948	9	0.000
0-2	81.714	9	0.000
0-3	64.164	9	0.000
1-2	41.370	9	0.000
1-3	39.509	9	0.000
2-3	23.467	9	0.005

Source: Author's calculation based on CHNS sample, test commands and results based on LONG and FREESE (2003, p. 204).

Table 4-42: Multinomial estimation results 1997

	OLS Coefficients	MNL Coefficients			OLOGIT Coefficients
	$A_{Primary}$ (0), $A_{Primary+soc}$ (1), BC (2), WC (3)	$A_{Primary}$ (0), $A_{Primary+soc}$ (1)	$A_{Primary}$(0), BC (2)	$A_{Primary}$(0), WC (3)	$A_{Primary}$ (0), $A_{Primary+soc}$ (1), BC (2), WC (3)
Bouyei (0/1)	0.402	0.003	4.332*	3.089	1.257**
	(0.243)	(0.464)	(2.496)	(2.548)	(0.607)
Miao (0/1)	-0.204**	0.594	-1.162***	-0.562	-0.121
	(0.085)	(0.758)	(0.344)	(0.667)	(0.413)
Tujia (0/1)	0.059	0.727	-0.383	1.226	0.742**
	(0.077)	(0.608)	(0.273)	(1.130)	(0.327)
Education	0.056***	0.218***	0.141	0.553***	0.246***
(years)	(0.010)	(0.077)	(0.089)	(0.093)	(0.041)
Male (0/1)	-0.013	1.068*	-0.199	-0.811	0.027
	(0.067)	(0.560)	(0.749)	(0.678)	(0.334)
Age (years)	0.001	-0.041***	-0.040**	0.020	-0.012
	(0.003)	(0.015)	(0.019)	(0.014)	(0.017)
Households	0.003***	-0.001	0.017***	0.015***	0.010***
(N)	(0.001)	(0.001)	(0.004)	(0.004)	(0.002)
County 2	0.246	2.232***	3.134**	14.531***	2.370**
(0/1)	(0.313)	(0.557)	(1.409)	(1.508)	(1.181)
County 3	0.366	2.555***	1.200	13.909***	3.165**
(0/1)	(0.276)	(0.520)	(2.044)	(1.434)	
Constant	-1.244***	-4.813***	-10.555***	-28.699***	
	(0.333)	(0.820)	(2.840)	(3.100)	
Cut 1					8.582***
					(1.877)
Cut 2					9.731***
					(1.959)
Cut 3					13.528***
					(2.862)
Observations	599	599	599	599	599
R^2	0.6097				
Log ps. Likelihood			-251.22117		-291.63969
Pseudo R^2			0.5047		0.425

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

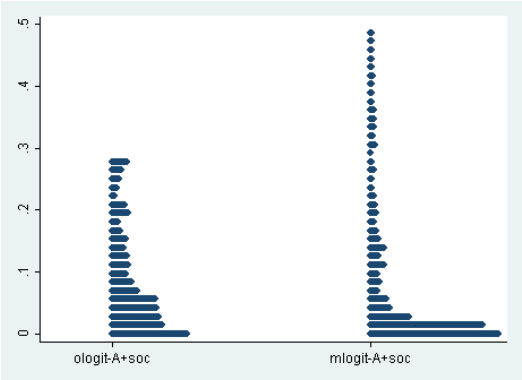
Table 4-43: Wald test after suest 1997

H_0 = no difference between estimators of full and restricted model
H_1 = difference between estimators of full and restricted model

	A+soc	***BC***	***WC***
Model Restrictions			
Exclusion *A+soc*		chi2(10) =148.92 Prob > chi2 = 0.001	chi2(8) = 1794.56 Prob > chi2 = 0.001
Exclusion *BC*	chi2(10) = 837.99 Prob > chi2 = 0.001		chi2(10) = 790.87 Prob > chi2 = 0.001
Exclusion *WC*	chi2(10) = 64.23 Prob > chi2 = 0.001	chi2(10) = 229.43 Prob > chi2 = 0.001	

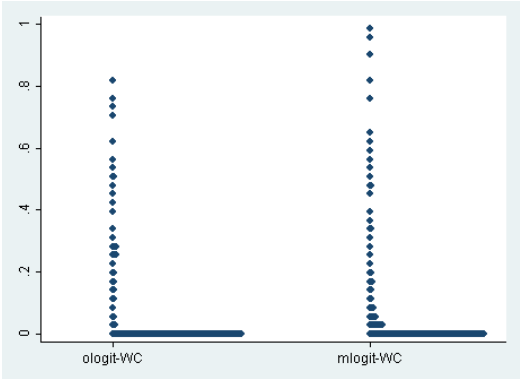
Agriculture is base group. Standard errors are based on cluster robust variance estimator.
Source: Author's calculation based on CHNS sample.

Figure 4-12: Plot to compare predicted probabilities of sector A-Primary+soc in MNL and OLOGIT models 1997



Source: Author's calculation based on CHNS sample.

Figure 4-13: Plot to compare predicted probabilities of sector WC in MNL and OLOGIT models 1997



Source: Author's calculation based on CHNS sample.

4.4.4.2 Interaction terms

The next step is to include interaction terms between ethnic status and education. I use the binary comparisons of *A* versus *NA*, $A_{Primary}$ versus $A_{Primary+soc}$, $A_{Primary}$ versus *BC* and $A_{Primary}$ versus *NA*. I also show the predicted probability plot and education for all occupational outcome comparisons.

In the comparison of the main sectors *A* and *NA*, there are no improvements of model outcomes after including interaction terms between ethnic status and education. There is a pattern in the sample which shows that from zero to five years of education, it is very unlikely for any ethnic group to work in the *NA* sector (figure 4-14). From five to twelve years of education there is no clear pattern, while after twelve years of education the probability of working in *NA* increases for all ethnic groups.

In the comparison of the sub-sectors $A_{Primary}$ and $A_{Primary+soc}$, there is an improvement in the model outcomes after including the interaction term Tujia*Education (table 4.44). There is a pattern in the sample which shows that for those who have from zero to six years of education, it is very unlikely to have a *soc* alongside *A* for all ethnic groups (figure 4-15). From six to thirteen years of education there is no clear pattern. After thirteen years of education, the probability of having a *soc* alongside agriculture, however, increases for all ethnic groups.

The magnitudes of the coefficients education, male, age, county 2 and 3 are the same as in the restricted model (see table 4.44 and table 4.38). With the inclusion of the interaction term Tujia*Education, there are statistically significant effects in the logit model for the interaction term Tujia*Education and also for Tujia status. The marginal effects after logit indicate that the Tujia have a 34.7 % higher probability of having a soc alongside agriculture than do Han. For each year

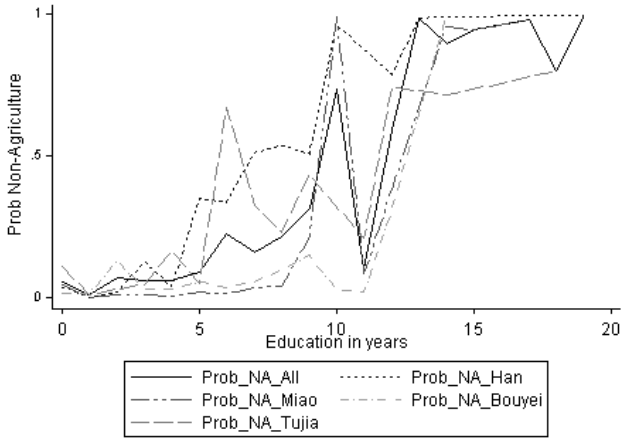
of education above the average, the Tujia, however, have a 2.9 % lower probability of having a *soc* than do Han. This indicates that better educated Tujia may leave agriculture completely to work full-time in the *NA* sector rather than having a *soc* alongside agriculture.

In the comparison of the sub-sectors $A_{Primary}$ and *BC*, there is an improvement in the model outcomes after including the interaction term Bouyei*Education (table 4.45). There is a trend in the sample in which individuals having from zero to five years of education are very unlikely to work in the *BC* sector regardless of ethnic group; from five to twelve years of education, there is no clear pattern observable; and after twelve years of education, the probability of working in *BC* increases for all ethnic groups (figure 4-16). Overall the figure shows that, regardless of educational cohort, Han have a higher probability of working in *BC* positions than do ethnic minorities.

In the original model without interaction terms, the magnitudes of the coefficients differ in both the linear probability and logit models (table 4.39). In the model with interaction terms, there are also differences between the linear probability and logit models (table 4.45). The interaction effect Bouyei*Education is only significant in the logit model. While in the original logit model Bouyei status has a statistically significant positive effect on the probability of working in *BC*, Bouyei status is insignificant in the logit model with interaction terms (see table 4.39 and table 4.45). The interaction term Bouyei*Education, however, has a statistically significant positive effect on the probability of working in *BC*. For each additional year of education above the average, the probability of working in *BC* increases by 1.2 % for the Bouyei. The other significant coefficients have the same magnitudes in both models.

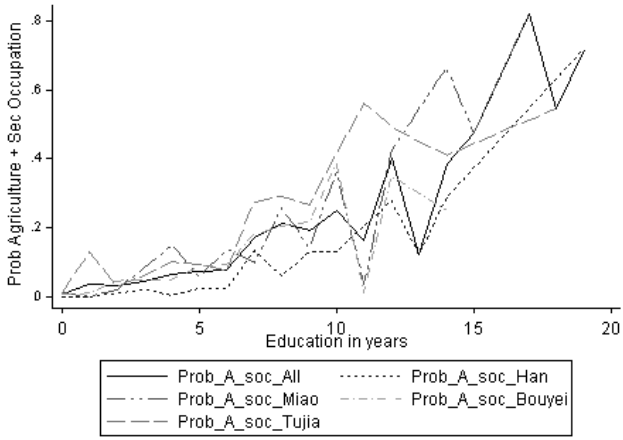
In the comparison of the sub-sectors $A_{Primary}$ and *NA*, there are no improvements in the model outcomes after including interaction terms between ethnic status and education. There is a pattern in the sample in which individuals with zero to five years of education are very unlikely to work in the *NA* sector regardless of ethnic group (figure 4-17). From five to twelve years of education there is no clear pattern, and after twelve years of education the probability of working in *NA* increases for all ethnic groups (figure 4-17). The figure, moreover, shows that the probability of working in *NA* is above average for Han.

Figure 4-14: Plot of predicted probabilities and years of education by ethnicity for A versus NA 1997



Source: Author's calculation based on CHNS sample.

Figure 4-15: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus A-Primary+soc 1997



Source: Author's calculation based on CHNS sample.

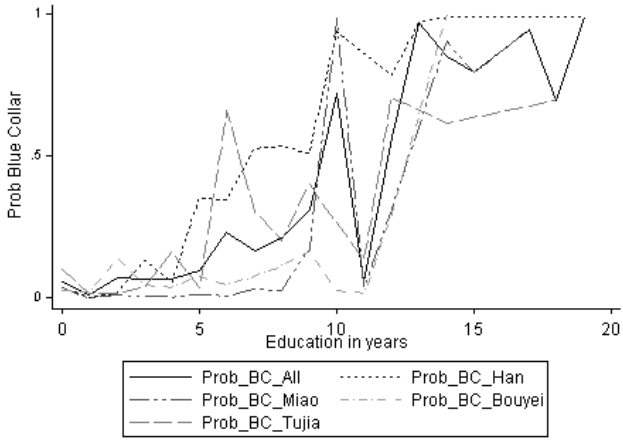
Table 4-44: Estimation results with interactions 1997 (A-Primary versus A-Primary+soc)

	OLS Coefficients $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*	**Logit Coefficients** $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*	**Logit Marginal Effects** $A_{Primary}$ *(0)*, $A_{Primary+soc}$ *(1)*
Bouyei (0/1)	-0.037	0.024	0.002
	(0.027)	(0.502)	(0.033)
Miao (0/1)	0.008	0.888	0.069
	(0.031)	(0.756)	(0.067)
Tujia (0/1)	0.061	4.330***	0.347***
	(0.052)	(0.676)	(0.064)
Tujia*Education	-0.008	-0.441***	-0.029***
	(0.012)	(0.088)	(0.007)
Education (years)	0.017**	0.421***	0.028***
	(0.005)	(0.044)	(0.004)
Male (0/1)	0.066*	1.224**	0.076**
	(0.033)	(0.599)	(0.032)
Age (years)	-0.002*	-0.053***	-0.004***
	(0.001)	(0.015)	(0.001)
Households (*N*)	-0.000	-0.001	-0.001
	(0.000)	(0.002)	(0.001)
County 2 (0/1)	0.093***	1.903***	0.132***
	(0.019)	(0.571)	(0.037)
County 3 (0/1)	0.121***	2.755***	0.205***
	(0.019)	(0.750)	(0.057)
Constant	0.042	-6.386***	
	(0.053)	(1.212)	
Observations	485	485	485
R^2	0.1368		
Log ps. Likelihood		-105.77617	
Pseudo R^2		0.3046	

Note: Standard errors are adjusted to 9 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

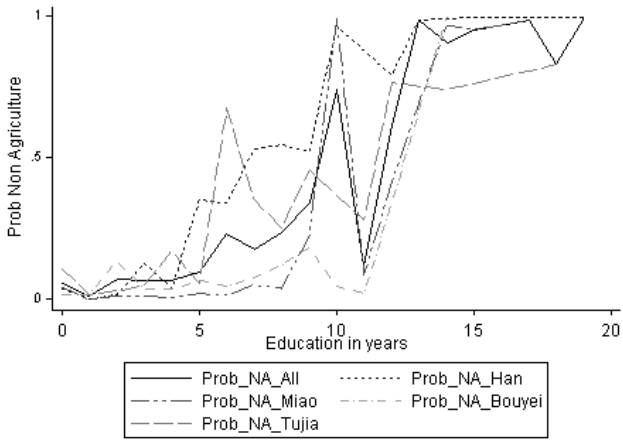
Source: Author's calculation based on CHNS sample.

Figure 4-16: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus BC 1997



Source: Author's calculation based on CHNS sample.

Figure 4-17: Plot of predicted probabilities and years of education by ethnicity for A-Primary versus NA 1997



Source: Author's calculation based on CHNS sample.

Table 4-45: Estimation results with interactions 1997 (A-Primary versus BC)

	OLS Coefficients	**Logit Coefficients**	**Logit Marginal Effects**
	$A_{Primary}$ (0), BC (1)	*$A_{Primary}$ (0), BC (1)*	*$A_{Primary}$ (0), BC (1)*
Bouyei (0/1)	0.192	2.340	0.107
	(0.135)	(2.201)	(0.134)
Miao (0/1)	-0.141***	-1.688***	-0.053***
	(0.028)	(0.372)	(0.021)
Tujia (0/1)	-0.062	-0.675	-0.025
	(0.035)	(0.471)	(0.018)
Bouyei*Education	0.011	0.309***	0.012***
	(0.008)	(0.069)	(0.004)
Education (years)	0.005	-0.017	-0.001
	(0.004)	(0.079)	(0.003)
Male (0/1)	-0.016	-0.311	-0.012
	(0.037)	(0.712)	(0.026)
Age (years)	-0.001	-0.052**	-0.002*
	(0.001)	(0.023)	(0.001)
Households (*N*)	0.002***	0.017***	0.001***
	(0.000)	(0.004)	(0.001)
County 2 (0/1)	0.111	2.898**	0.174*
	(0.159)	(1.442)	(0.094)
County 3 (0/1)	0.108	1.284	0.058
	(0.126)	(1.944)	(0.097)
Constant	-0.475**	-9.186***	
	(0.158)	(2.698)	
Observations	525	525	525
R^2	0.6096		
Log ps. Likelihood		-75.842515	
Pseudo R^2		0.676	

Note: Standard errors are adjusted to 11 community clusters, robust standard errors in parentheses, *** Significant at 1 %, ** Significant at 5 %, * Significant at 10 %.

Source: Author's calculation based on CHNS sample.

4.4.4.3 Interim conclusions 1997

I analyzed the 1997 sample with the five-step procedure described at the beginning of this chapter. As in other years, the 1997 MNL results do not support the IIA assumption; therefore, I do not give them further consideration. The post estimation tests after OLOGIT estimation indicate that the data sample is, moreover, not optimally supported by the OLOGIT model setup; therefore, I do not give further consideration to the OLOGIT results in 1997. The most appropriate results for 1997 are, thus, the estimation results after the binary logit estimations. An overview of marginal effects after binary logit estimations is shown in table 4.46.

Table 4-46: Marginal effects after logit with interactions for all sectors 1997

	Logit Marginal Effects *A (0), NA (1)*	**Logit Marginal Effects** $A_{Primary}$ *(0),* $A_{Primary+soc}$ *(1)*	**Logit Marginal Effects** $A_{Primary}$ *(0), BC (1)*	**Logit Marginal Effects** $A_{Primary}$ *(0), NA (1)*
Bouyei (0/1)	0.234	0.002	0.107	0.230
	(0.155)	(0.033)	(0.134)	(0.145)
Miao (0/1)	-0.041***	0.069	-0.053***	-0.044***
	(0.017)	(0.067)	(0.021)	(0.017)
Tujia (0/1)	-0.006	0.347***	-0.025	-0.015
	(0.010)	(0.064)	(0.018)	(0.013)
Bouyei*Education			0.012***	
			(0.004)	
Tujia*Education		-0.029***		
		(0.007)		
Education (years)	0.009***	0.028***	-0.001	0.01***
	(0.003)	(0.004)	(0.003)	(0.003)
Male (0/1)	-0.02	0.076**	-0.012	-0.011
	(0.029)	(0.032)	(0.026)	(0.028)
Age (years)	-0.001	-0.004***	-0.002*	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Households (*N*)	0.001***	-0.001	0.001***	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)
County 2 (0/1)	0.196**	0.132***	0.174*	0.225**
	(0.094)	(0.037)	(0.094)	(0.092)
County 3 (0/1)	0.062	0.205***	0.058	0.073
	(0.113)	(0.057)	(0.097)	(0.113)
Observations	599	485	525	553

Source: Author's calculation based on CHNS sample.

The average level of education is 5.2 years in 1997. There is a statistically significant difference in education between Han and the Bouyei and between Han and the Tujia. Han, the Bouyei and the Tujia have on average 5.7 years, 4.9 years and 4.8 years of education, respectively. The test results suggest that Han have on average 0.8 years more education than the Bouyei and 0.9 years more education than the Tujia.

The marginal effects after logit indicate that there are ethnic effects for all three ethnic minorities in 1997 (table 4.46). Better educated Bouyei have a higher probability of working in *BC* than do Han. In contrast the Miao have a lower probability of working in *NA* than do Han, and within the *NA* sector, in *BC* than do Han. The Tujia generally have a higher probability of having a *soc* alongside agriculture than do Han. This probability, however, decreases for the Tujia as years of education increase above the average. Tujia with more education may leave the *A* sector to work full time in the *NA* sector.

Additional years of education, moreover, increase the chances of having a *soc* alongside *A* and of working in *NA*. More years of education, however, have no significant effect on the probability of working in the *BC* sector. It can also be seen that men have a higher probability of having a *soc* alongside agriculture than do women. In contrast older individuals have a lower probability of having a *soc* or of working in *BC*. Living in a larger village increases the probability of working in *NA* and in the *BC* sector, but living in a larger village is insignificant in respect to having a *soc*. Living in county 2 rather than in county 1 increases the probability of working in *NA*, including a *soc*; living in county 3 rather than in county 1 increases the probability of having a *soc* alongside agriculture, though it does not change the probability of working in *NA* in general.

4.5 Conclusions

This subchapter provides the overall conclusions of the econometric analyses. In the first part (section 4.5.1) I show dynamic developments of the determinants of occupational differences between 1997, 2000 and 2004. In the second part (section 4.5.2) I answer the hypotheses about the determinants of occupational outcomes.

4.5.1 Dynamic developments in occupational outcomes

This subchapter combines the interim conclusions given in sections 4.4.1.3, 4.4.2.3 and 4.4.3.3 for 2004, 2000 and 1997, respectively. The goal of this subchapter is to show dynamic developments in the determinants of occupational outcomes over this timeframe based on the binary logit models. Table 4.47 summarizes the magnitudes of the marginal results after the binary logit estimations. There are some clear effects which are the same throughout the considered years, yet some effects change their significance levels and magnitudes throughout the years; there might be some unobservable effects related to the sample size because the sample size varies in each year. Regarding the dependent variables I draw the following conclusions:

Ethnic status

The Bouyei coefficient is insignificant in all model specifications and years. There are some significant effects for Bouyei with more years of education. In 1997 Bouyei with an additional year of education above the average are more likely to work in the *BC* sector than are Han. In 2004 Bouyei with an additional year of education above the average are more likely to have a *soc* alongside *A* than are Han.

The Tujia coefficient is only significant in the comparison of $A_{Primary}$ and $A_{Primary+soc}$. In 1997 and 2004 the Tujia have a higher probability of having a *soc* alongside *A*, with the exception of 1997, when Tujia with an additional year of education above the average have a lower probability of having a *soc* than do Han. The 2000 results are insignificant.

Table 4-47: Summary of marginal effects after logit with interactions for all sectors and years

	Model 1 *A (0), NA (1)*			***Model 2*** *$A_{Pr.}$ (0), $A_{Pr.+soc}$ (1)*			***Model 3*** *$A_{Pr.}$ (0), BC (1)*			***Model 4*** *$A_{Pr.}$ (0), NA (1)*		
Variables | Years	**97**	**00**	**04**	**97**	**00**	**04**	**97**	**00**	**04**	**97**	**00**	**04**
Bouyei	0	0	0	0	0	0	0	0	0	0	0	0
Miao	–	0	–	0	0	+	–	–	–	–	0	–
Tujia	0	0	0	+	0	+	0	0	0	0	0	0
Bouyei*Education						+	+					
Miao*Education			+						+			+
Tujia*Education				–								
Education	+	0	+	+	+	0	0	0	+	+	0	+
Male	0	0	+	+	0	+	0	0	+	0	0	+
Age	0	–	–	–	0	–	–	–	–	0	–	–
Households	+	+	+	0	0	0	+	+	+	+	+	+
County 2	+	+	0	+	0	–	+	+	0	+	+	0
County 3	0	+	0	+	+	–	0	+	0	0	+	0

Discrete model settings: A = agriculture (includes agriculture as primary or secondary occupation), *NA* = non-agriculture (includes all jobs except agriculture), $A_{Pr.}$ = agriculture is primary occupation, $A_{Pr.+soc}$ = alongside agriculture as primary occupation the individual also has a secondary occupation, *BC* = blue-collar jobs. (0) indicates that the occupation is the base group. (1) indicates that the occupation is the comparative group.

The estimation effects are 0, +, – which stand for no statistically significant effect, statistically significant positive effect, statistically significant negative effect, respectively. If the field is empty, then the interaction terms were based on likelihood ratio test results not included in the model.

Source: Author's calculation based on CHNS sample.

Miao status has a significant effect in all model specifications, but this effect is not the same for all years. In the general sector comparison (*A* versus *NA*), Miao status has a statistically significant negative effect on the probability of working in *NA* in 1997 and 2004. The interaction term Miao*Education and the probability plot, however, indicate that better educated Miao have a higher probability of working in *NA* than do Han in 2004. In the comparison of the subcategories of agriculture ($A_{Primary}$ and $A_{Primary+soc}$) I find that the Miao are more likely to have a *soc* alongside *A* than are Han in 2004. In the comparison of the subcategories ($A_{Primary}$ and *BC*, $A_{Primary}$ and *NA*), I find that the Miao have a lower probability of working in *BC* and in *NA* than do Han, with increasingly negative effects over time. In 2004 Miao with an additional year of education above the average, however, have a higher probability of working in the *BC* or *NA* sectors than do Han.

Education

Each additional year of education increases employment probability in all branches of the *NA* sector. The effects are, however, not statistically significant in all years. By comparing (*A* versus *NA*, $A_{Primary}$ versus *NA*) I find that in 1997 and in 2004 each additional year of education above the average improves access to employment in the *NA* sector. When comparing ($A_{Primary}$ and $A_{Primary+soc}$), I observe that each additional year of education above the average significantly increases the probability of having a *soc* in 1997 and in 2000. By comparing ($A_{Primary}$ and *BC*) the positive educational effect is only significant in 2004, but not in 2000 and in 1997.

Gender

Although the gender effect is not always statistically significant in all models and in all years, there is a clear pattern of men having a higher probability of working in *NA* than do women. This is particularly apparent in 2004, when men have a higher probability of working in *NA* in all model specifications. In contrast in 2000 there is no statistically significant gender effect. In 1997 there is a significant effect in the comparison of ($A_{Primary}$ and $A_{Primary+soc}$); in that same year men are also more likely than women to have a *soc* alongside *A*. This effect is much higher in 2004 at 19.9 % than at 7.6 % in 1997.

Age

Although the age effect is not always statistically significant in all years, there is a clear pattern that younger individuals have a higher probability than older individuals of working in *NA*. In the comparisons (*A* versus *NA*, $A_{Primary}$ versus *NA*) each additional year of age above the average decreases the probability of working in *NA* in 2000 and 2004. I also observe that younger individuals have a higher probability of having a *soc* than do older individuals in 1997 and 2004. In the comparison of ($A_{Primary}$ and *BC*), there are the same negative age effects with increasing magnitudes over time.

Geographic Location

Number of households in villages

There is a clear village effect in 1997, 2000 and 2004. The only exception is in the comparison of ($A_{Primary}$ and $A_{Primary+soc}$), where no statistically significant effect is observable. For all other model specifications, I observe that larger villages facilitate access to all types of *NA* employment considered in all years.

Counties

I consider three counties in the econometric analysis. The first county (county 1) is the base outcome. The second county (county 2) and the third county (county 3) are compared to the base county (county 1). The county effects are not always

significant and, moreover, do not have the same magnitudes in all years and in all model specifications. There is, thus, no clear pattern observable.

In the model (*A* versus *NA*) I find that in 1997 and 2000 working in the *NA* sector is much more probable in county 2 than in county 1, while in 2004 this effect was not statistically significant. In county 3 the probability of working in *NA* has a statistically significant increase only in 2000, but not in other years. In the model ($A_{Primary}$ and $A_{Primary+soc}$) the county effects are mainly statistically significant, but the county effects change with time. In 1997 county 2 positively affects the access to a *soc* alongside *A*, but in 2004 county 2 negatively affects the access to a *soc* alongside *A*. County 3 is significant in all three years. In 1997 and in 2000 county 3 has a statistically significant positive effect, but the effect diminishes between 1997 and 2000. In 2004 the county 3 effect even turns negative. In the model ($A_{Primary}$ and *BC*) I find that county 2 significantly improves access probability to the *BC* sector when compared to county 1 in 1997 and 2000; the effect is, however, not statistically significant in 2004. County 3, however, has a significantly positive effect only in 2000. In the model ($A_{Primary}$ and *NA*) I observe the same county effects as in the comparison of $A_{Primary}$ and *BC*. County 2 significantly increases the probability of working in *NA* compared to county 1 in 1997 and 2000, but county 2 is insignificant in 2004. In contrast county 3 only has a statistically significant positive effect in 2000.

4.5.2 Determinants of occupational outcomes

The conclusions of this chapter are related to the theoretical framework and refer to the corresponding hypotheses (H1-H7), which I developed in section 2.5.1.2. I test the hypotheses (H1-H7) and draw conclusions based on empirical evidence.

Ethnic status

> H1: Being an ethnic minority negatively influences access to non-agricultural employment.

I consider the occupational outcomes of the ethnic minorities Bouyei, Miao and Tujia and compare their occupational outcomes with the majority group Han. In the comparisons of occupational outcomes with logit models Bouyei status is insignificant. H1, thus, can be rejected for the Bouyei group. First, the Bouyei coefficient is insignificant in the occupational outcome analysis in the years considered. Second, Bouyei with an additional year of education above the average have a higher probability of working in *BC* in 1997 and of having a *soc* alongside *A* in 2004 than do Han. These results are somehow coherent with the outcomes of my field observation. Bouyei people of Qiannan Bouyei-Miao Autonomous Prefecture have already been sinicized to a large extent. In the visited areas in Qiannan Bouyei-Miao Autonomous Prefecture, Bouyei people speak Mandarin and some Bouyei even informed me that they are unable to speak the Bouyei language. In this area traditional clothing is, moreover, an exception for the Bouyei. As the

estimation results show, better educated Bouyei have a higher probability of working in *BC* than do Han in 1997 and have a higher probability of having a *soc* alongside agriculture in 2004. The positive institutional framework may, therefore, offset negative employer discrimination against better educated Bouyei.

The Miao coefficient is statistically significant. It is apparent that without controlling for education, the Miao have a lower probability of working in *NA* in 1997 and in 2004 and in the *BC* sector in all three years. I, thus, accept H1 for the Miao group, which means that being Miao negatively influences access to *NA* employment. These results are in accordance with the outcomes of my field observation. I find that prejudices about the inadequate culture and weaker Mandarin language skills are strongest against the Miao group, particularly in Qiandongnan Miao-Dong Autonomous Prefecture. I find that in all Miao areas considered in this research cultural markers such as traditional language and clothing have been sustained to a large extent. Better educated Miao, who usually speak Mandarin, however, have a higher probability of working in *NA* in general and in the *BC* sector in particular in 2004. In my field observation I find, for example, that a tourguide, who is Miao and has a university education, works in the ethnic tourism industry in Kaili. This indicates that better educated Miao can find jobs in the *NA* sector.

In the comparison of *A* with all branches of the *NA* sector, except $A_{Primary}$ versus $A_{Primary+soc}$, Tujia status is insignificant. This indicates that there is no difference in occupational outcomes between the Tujia and Han. H1 can, thus, be rejected for the Tujia group. In previous estimations I already observed that Tujia status is insignificant when comparing occupational outcomes. Due to time and funding constraints, I did not consider the Tujia in my field observation; therefore, I cannot directly compare the quantitative and qualitative results for the Tujia.

In 2004 the Miao, the Tujia and the better educated Bouyei have a higher probability of having a *soc* alongside *A*. The positive institutional framework, therefore, may offset negative employer discrimination against ethnic minorities. The ethnic tourism industry in Guizhou particularly offers *soc* for ethnic minorities, which increases ethnic minorities' employment probability in this sector. In 1997, however, a higher probability of having a *soc* alongside *A* is only statistically significant for the Tujia, but the probability of having a *soc* decreases for Tujia with more years of education. This indicates that Tujia with more years of education leave sector *A* for good and work full time in the *NA* sector.

Education

> H2: More years of education positively influence access to non-agricultural employment.

I analyze whether or not education influences occupational outcomes. In 1997 an additional year of education above the average increases the probability of working in the *NA* sector and of having a *soc* alongside *A*. In 2000 each additional educational

year above the average increases the probability of having a *soc* alongside *A*, but this additional year of education has insignificant effects on other outcomes. In 2004 an additional year of education above the average is crucial for working in *NA* in general and in the *BC* sector in particular. I, thus, accept H2 for all years considered in my study. More years of education positively influence access to *NA* employment. I conclude that from 1997 to 2004 education is essential for finding full employment in the *NA* sector; however, the importance of more education declines for finding a *soc* alongside *A*.

Ethnic status and education

> H3: The lower educational achievement of ethnic minorities negatively influences their access to non-agricultural employment.

In the descriptive statistics the average education in 1997 is 5.2 years, the same as in 2000, but 0.4 years less than in 2004. There are statistically significant differences in education between Han and the Bouyei and between Han and the Tujia in 1997, respectively. Han, Bouyei and Tujia individuals have on average 5.7 years, 4.9 years and 4.8 years of education in 1997, respectively. The test results indicate that Han have on average 0.8 more years of education than do the Bouyei and 0.9 more years of education than do the Tujia in 1997. In contrast in 2000 and 2004 there are no statistically significant differences in education between ethnic minorities and Han.

Differences in years of education by ethnicity are considered with interaction terms between ethnic status and education in the logit models and with probability plots. There are differences in education and in occupational outcomes between ethnic groups in 1997 and in 2004. Bouyei with more education have a higher probability of working in *BC* in 1997 and of having a *soc* alongside *A* in 2004. Better educated Miao have a higher probability of working in *NA* in general and in the *BC* sector in particular in 2004. I, thus, reject H3 for the Bouyei and for the Miao in the respective samples; better educated Bouyei and Miao have a higher probability of working in non-agricultural employment than do Han. The results for the Tujia do not follow a clear pattern. In 1997 and 2004 the Tujia have a higher probability of having a *soc* alongside *A*, with the exception of 1997, when Tujia with more years of education have a lower probability of having a *soc* than do Han. This indicates that Tujia with more years of education leave sector *A* for good and work full time in the *NA* sector. In 2000 I reject H3 because there are no statistically significant differences in education among ethnic groups. The positive institutional framework may offset negative employer discrimination against better educated Bouyei, Miao and Tujia. The results confirm that prejudices are not directly against the ethnic minority status, but are against an agglomeration of inadequate factors (see figure 2.5), such as less education, which is linked to lower Mandarin language skills, inadequate culture and other factors.

Gender

H4: Being female (male) negatively (positively) influences access to non-agricultural employment.

In 1997 men have a higher probability of having a *soc* alongside *A* than do women. I, thus, accept H4 in 1997. I conclude that being female negatively influences access to non-agricultural employment, and being male positively influences access to non-agricultural employment in 1997. The gender effect is, however, not statistically significant in 2000. I, thus, reject H4 in 2000, which indicates that there are no gender differences in accessing employment in the *NA* sector.

In 2004 the gender effect is, however, observable in the *NA* sector in general and in the *BC* sector in particular. I, thus, accept H4 in 2004. Being female negatively influences access to non-agricultural employment, and being male positively influences access to non-agricultural employment in 2004. It can be concluded that from 2000 to 2004 gender inequalities increased. In 2004 it is easier for men than for women to access full-time employment in the *NA* sector than in 2000 and in 1997. Over time gender inequalities in having a *soc* alongside *A*, however, disappear. In my field observation I found that particularly jobs which demand a lot of strength (e.g., plowing the fields with the help of a buffalo or heavy construction work) are done by men rather than by women.

Age

H5: Older age negatively influences access to non-agricultural employment.

In 1997 older age decreases the probability of working in *BC* and of having a *soc* alongside agriculture. In 2000 older individuals have a lower probability of working not only in the *BC* sector in particular but also in the *NA* sector in general. In 2004 older individuals even have a lower probability of working in all of the possible non-agricultural occupations. I, thus, accept H5 for all three years. I conclude that older age negatively influences access to non-agricultural employment with increasing effects over time.

Geographic Location

Geographic location is captured with county dummies and village size. H6 concerns differences in occupational outcomes depending on the counties, and H7 focuses on differences in occupational outcomes depending on the village size.

H6: There are on average differences in occupational outcomes depending on the counties considered.

To capture county effects in the estimations, I consider three counties with mixed ethnic populations. The results after logit show that in 1997, in contrast to residence in county 1, residence in county 2 increases the probability of working in both *NA* and *BC* and increases the probability of having a *soc* as well. In 2000 residence in county 2 also increases the probability of working in *NA* and *BC*;

however, residence in county 2 no longer has a statistically significant effect on the probability of having a *soc*. In contrast in 2004 residence in county 2 has a negative effect on the probability of having a soc; however, residence in county 2 no longer has a statistically significant effect on the probability of working in *NA* in general and in the *BC* sector in particular.

In 1997 residence in county 3 increases the probability of having a *soc* alongside *A* in contrast to residence in county 1; in 2000 residence in county 3 has a statistically significant positive effect on the probability of working in all kinds of *NA* employment; and in 2004 residence in county 3 has a statistically significant negative effect on the probability of having a *soc*; thus, I accept H6 for all three years. I conclude that there are on average differences in occupational outcomes depending on the counties considered. I find that county effects change their magnitudes over time; this could be a result of increasing infrastructure investments in counties 2 and 3 in contrast to county 1. In my field observation I was informed that highways in Guizhou were newly constructed around five years ago and that the travel time, for example, between Kaili and Duyun had decreased.

> H7: Larger villages positively influence access to non-agricultural employment.

In all three years residence in larger villages increases the probability of working in *NA* and in *BC*; thus, I accept H7 for the years considered. I conclude that residence in larger villages positively influences access to non-agricultural employment. I conclude that in Guizhou better developed infrastructure and larger local markets, which are both related to larger village size, improve job chances in *NA*.

5 CONCLUSIONS

This chapter presents some conclusions regarding the major findings of this study. Theoretical and empirical conclusions are presented in subchapters 5.1 and 5.2, respectively. Policy recommendations and an outlook for further research are presented in subchapters 5.3 and 5.4, respectively.

5.1 Theoretical conclusions

The leading content related research question of this monograph, are minorities, due to their ethnic affiliation, discriminated against in the rural labor market, serves as guideline in the theoretical discussion. To analyze ethnic differences in occupational outcomes in rural areas, I interlinked group differences theories, human capital theories, labor market discrimination theories, occupational choice theories, farm household theories and non-farm rural employment theories based on the benchmark model of JOHNSON and STAFFORD (1998). I arranged the theoretical concepts in a diamond of theories based on four reasons for occupational differences: differences in employer discrimination, in institutional discrimination, in abilities and in preferences (figure 2-4).

To analyze the Guizhou case, I made two assumptions to narrow down the four possible reasons for occupational differences among ethnic groups (see subchapter 2.5.1). First, all individuals regardless of their ethnic affiliation prefer to work in the non-agricultural sector because in Guizhou agricultural work is seen as inferior to non-agricultural work. Second, there is a preferential policy framework for ethnic minorities in China. This preferential policy framework consists of laws (institutions), which support ethnic minorities in many aspects, such as in education and in employment; therefore, I postulate that in Guizhou ethnic minorities are not discriminated against by institutions. Preferences and institutional discrimination are, therefore, held constant in the analysis. The two remaining explanations for occupational differences among ethnic groups are, hence, differences in employer discrimination and in abilities. Additionally I find that geographic location is an important determinant for explaining differences in occupational outcomes.

Based on these assumptions and findings, I postulate that ethnic minorities in rural Guizhou face three constraints for accessing employment in the non-agricultural sector: 1) individuals' abilities are not adequate to perform non-agricultural work, 2) employers in the non-agricultural sector discriminate against ethnic minority workers and 3) non-agricultural work is not available in the area. The theoretical concepts for explaining differences in abilities are group differences theories and human capital theories. In these theories educational attainment, which is influenced

by various pre-labor market factors, is an important determinant of differing occupational outcomes (figure 2-5). The theoretical concepts for explaining differences in employer discrimination are taste-based discrimination and statistical discrimination theories. Employers can influence occupational and wage distributions of ethnic minorities through at least five discriminatory practices: distaste (prejudice), negative belief (stereotype), information uncertainty about workers' productivity, negative signal about workers' abilities and language discrimination (subchapter 2.2.2).

The theoretical concepts which focus on rural areas are farm household models and non-farm rural employment theories. The major theoretical conclusion of my work is that farm household models and non-farm rural employment theories should be combined with employer discrimination theories. As a result of employer discrimination, ethnic minorities are in a distress-push situation and are forced either to work in badly paid non-agricultural jobs or to stay in agriculture, *ceteris paribus*. Employer discrimination, therefore, influences ethnic minorities' time allocation in farm household models. Aside from these theoretical explanations, non-agricultural employment in rural areas can actually only be obtained if there are available jobs. In remote rural villages there are often fewer non-agricultural jobs than in better connected and larger villages.

5.2 Empirical conclusions

Empirical investigations of ethnic differences in occupational outcomes are cumbersome because employer discrimination is forbidden and there is only limited data available; therefore, I posed the question to what extent discrimination is empirically measurable.

To analyze ethnic differences in occupational outcomes in rural Guizhou, I applied methodological triangulation by combining discrete choice modeling in the quantitative portion of my research and participant observation in the qualitative portion. I use discrete choice models (occupational outcome models) rather than wage equations or segregation indices as with occupational outcome models I can better reflect differences in job categories, most importantly between agricultural and non-agricultural sectors. I use participant observation rather than audit studies or structured interviews in the qualitative portion as field investigations about sensitive topics are restricted in China. I am, moreover, not required to do a full survey in the area as there is secondary data information conducted by the China Health and Nutrition Survey (CHNS) freely available. Through conversations with foreign and Chinese scholars of Guizhou University, I acquired additional qualitative information. The field observations and conversations allowed me to better understand modeling results and to refine hypotheses. The qualitative portion of my research is adopted as a "corrective" to complement and make up for the deficiencies of an exclusively quantitative approach.

Quantitative and qualitative methodologies both have shortcomings in the measurement of ethnic differences in occupational outcomes, particularly in disentangling effects from preferences, abilities and discrimination; therefore, labor market discrimination against ethnic minorities can only be identified with discrete choice models (occupational outcome models) by using well-defined assumptions about relationships among variables; moreover, labor market discrimination against ethnic minorities is difficult to identify with observation because with observation techniques I could only distinguish the ethnic minorities from their traditional clothing. A deeper analysis of labor market discrimination requires long-term fieldwork which was not possible in the framework of this project due to the limitations of available funding and time, but also to the high political sensitivity of the overall topic.

Based on the theoretical framework, I postulated that accessing employment in the non-agricultural sector is subject to three major constraints: 1) individuals' abilities are not adequate to perform non-agricultural work, 2) ethnic minority workers are discriminated against in the non-agricultural sector and 3) non-agricultural work is not available in the area. In line with these three constraints, I developed testable hypotheses, which inquired about the major human capital factors and about the geographic locations. The empirical analysis served to prove: 1) whether the Bouyei, the Miao and the Tujia face constraints in accessing non-agricultural employment, 2) whether fewer years of education hinder access to non-agricultural employment, 3) whether the Bouyei, the Miao and the Tujia have on average fewer years of education than do Han and, therefore, have a lower probability of working in non-agricultural employment, 4) whether men and women have differences in accessing non-agricultural employment, 5) whether age has an influence on accessing non-agricultural employment, 6) whether the counties show occupational differences and 7) whether smaller and larger villages have occupational differences.

In the discrete choice models (occupational outcome models) I used samples from the China Health and Nutrition Survey for the years 1997, 2000 and 2004 (chapter four). With linear probability and logit models, I analyzed the effects of the following independent variables: ethnic status (Bouyei, Miao, Tujia and Han), education, gender, age, counties and village size on the binary dependent variable agriculture versus non-agriculture and subgroups of both sectors. When the overall model fit improved, I included interaction terms between ethnic status and years of education. I also tested whether or not occupational outcomes can be combined and then estimated multinomial logit and ordered logit models on the categories considered. The results of the multinomial logit models are, however, questionable as the independence from irrelevant alternatives assumption was violated, and the results of the ordered logit estimation are questionable because the calculated probabilities were not correctly representing the data.

Regarding the dependent variable, I observed that in Guizhou agriculture is accomplished with traditional methods because the mountainous topography of the province makes it impossible to employ modern technologies to do the work more efficiently. SCHEIN (2000, pp. 161-162) observed the same during the rice harvest in Xijiang. Non-agricultural positions were mainly in the construction and service sectors and their corresponding subcategories.

There are almost no differences between Han, the Tujia and the Bouyei in occupational outcomes. Being Miao, however, negatively influences access to non-agricultural employment. This is in contrast with the inquiry by MAURER-FAZIO et al. (2004, 2005), which finds that in the 1982 census the Miao have a higher labor market participation rate than do Han. In 2004 better educated Miao, however, have a higher probability of working in non-agriculture in general and in blue-collar positions in particular than do Han.

A secondary occupation alongside agriculture is a very important income source for the Bouyei, Miao and Tujia. In 2004 better educated Bouyei, the Miao and the Tujia all have a higher probability of having a secondary occupation alongside agriculture than do Han. Ethnic minorities have a higher probability of having a secondary occupation alongside agriculture in Guizhou because the tourism industry particularly offers secondary occupations for ethnic minorities, as their traditional lifestyle is a major tourist attraction. This is in line with GUSTAFSSON and LI's findings (2003), which show that ethnic minorities in Guizhou and Yunnan particularly benefit from a growing tourism industry.

The results show that the Bouyei and the Tujia are not discriminated against. Not all Miao people suffer employer discrimination, only less educated Miao do. There is, therefore, not a general stereotype against Miao people but some taste-based discrimination against less educated Miao; less education implies an inadequate *wenhua* (cultural) level and weaker Mandarin language skills among other factors (see figure 2.5). Less educated Miao are probably unable to fulfill the *wenhua* level demanded in better paying non-agricultural positions; therefore, they may lose optimism about better job prospects and wages and may lose the incentive to work harder and/or to invest in human capital. This puts Miao with less education in a vicious cycle, and they continue working in agriculture. In this case the initial negative belief about the Miao's lower *wenhua* level is self-confirming (cf., COATE and LOURY, 1993); this implies that an inadequate *wenhua* level and inadequate language acquisition are both causes and symptoms of the Miao's different employment outcomes. The cultural and social variety that lurks behind the ethnic label "Miao", however, demands additional inquiries. In contrast Han see as unfair the preferential policies in education towards ethnic minorities because ethnic minorities are direct competitors in university entrance examinations.

Regarding the other constraints, I found that each additional year of education above the average positively influences access to non-agricultural employment,

that women and the elderly are underrepresented in non-agricultural employment and that smaller villages are linked to a higher probability of agricultural employment. From 1997 to 2004 education remains essential for finding full-time employment in the non-agricultural sector, yet in the same period the importance of more education declines for having a secondary occupation alongside agriculture. I found that from 2000 to 2004 gender inequalities increased. In 2004 it was easier for men rather than for women to access non-agricultural employment, while in previous years there had been no gender difference. The results concur with SCHEIN's, who finds that in Xijiang men work in non-agricultural jobs and women in agricultural jobs (SCHEIN, 2000, p. 174). I generally observed that jobs which demand a lot of strength were only done by men rather than by women. Older age negatively influences the access to non-agricultural employment in samples considered. Job inequalities between younger and older individuals even increased between 1997 and 2004. As there are no preferential policies in employment for the elderly (ROSS et al., 2007), age inequalities between agricultural and non-agricultural employment might persist.

In most areas the remoteness of villages and available natural resources determine employment possibilities. There are differences in occupational outcomes depending on the counties considered. I found that county effects even changed their magnitudes between 1997 and 2004, which can be explained by increasing infrastructure investments in some counties. The results, moreover, show that residence in larger villages positively influences access to non-agricultural employment. Generally better developed infrastructure and larger local markets are linked to larger village size and to higher demand for non-agricultural employment. The results concur with GUSTAFSSON and SAI's, who find that in China inequalities in income and in poverty are linked to geographic location (GUSTAFSSON and SAI, 2006, 2008); moreover, the results concur with SCHEIN's, who finds that in Guizhou locals rank villages based on their infrastructure connectivity (SCHEIN, 2000, p. 240).

5.3 Policy implications

In China there are several preferential policies to improve ethnic minorities' education and employment outcomes. Ethnic minorities enjoy benefits in accessing school and university, in hiring and promotion of cadres, in starting a business, in accessing positions in the people's congress and leading positions in autonomous areas (SAUTMAN, 1997, p. 3). The employment promotion law which was implemented in 2008 enforces non-discrimination measures for ethnic minorities in employment.

Although I have not directly analyzed the impact of these policies, there are some positive results: the Bouyei and the Tujia are not discriminated against; there is employer discrimination only against less educated Miao, not against all Miao people. To reach a sustainable integration of ethnic minorities in non-agricultural employment, ethnic minorities must have the required education; each additional

year of education generally improves chances of working in non-agricultural employment. Policy makers should, therefore, particularly encourage the Miao to increase their education because the Miao are more likely to work in agriculture than are other ethnic groups. To reach a sustainable integration, however, ethnic minorities should have the right to maintain their cultural diversity and mother tongues because if these rights are not guaranteed, ethnic minorities could suffer from culture shock, which could in turn trigger resentment against Han, as is the case in TAR and XUAR. To combine sustainable development and protection of cultural heritage requires that all ethnic groups understand, learn from and respect each other. Adapted school and university syllabi and information campaigns could help improve mutual awareness among ethnic groups. Additionally the lower representation of women and of the elderly in non-agricultural positions requires further governmental consideration.

Another problem in Guizhou is the fragile topography. Guizhou's land area has 62 % karst landforms and 19 % stony desert; as a result it is very fragile and hinders economic development (CHINA DAILY, 2012). These topographical conditions have a clear impact on employment options; in remote areas there are usually fewer employment possibilities than in better developed areas nearer to urban centers. Guizhou's topography makes development of rural areas through, for example, infrastructure investments, special economic zones and touristic zones, comparatively difficult. The government should, therefore, further improve the urbanization rate (currently around 51.3 % in China)[27] and encourage workers from impoverished rural areas to find employment in better developed areas. This requires an adjustment of the *hukou* household registration system, which currently impedes migrant workers from becoming permanent residents in cities and from accessing basic social services outside their home villages. Allowing villagers the full right to work and settle outside their home villages, furthermore, requires an adjustment of property ownership as currently all land officially belongs to the state.

5.4 Outlook on further research

China officially has 55 ethnic minority groups, but only some of them have been investigated with occupational outcome models (discrete choice models). Before my study only two studies had applied occupational outcome models: HANNUM and XIE (1998), who analyzed occupational outcomes of Han and ethnic minorities in XUAR, and ZANG (2008), who compared occupational outcomes of Han and Hui in Gansu. The China Health and Nutrition Survey comprises nine provinces. It would be fruitful to analyze employment discrimination against ethnic minorities in Guangxi Zhuang Autonomous Region, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning and Shandong, in the same manner I have examined

[27] CHINA.ORG.CN (2012).

the issue in Guizhou. Further research in this field could also be conducted by using additional explanatory variables or by changing occupational categories; moreover, regarding the dependent variable, it would also be important to analyze unemployment and migration. In figure 2-5 I pointed out many relationships among explanatory variables; in addition to the variables I considered, ethnicity, education, gender, age and geographic location, it could also be helpful to include other explanatory variables, such as psychological factors (e.g., personality traits and intelligence), socioeconomic factors (e.g., household characteristics, pre-labor market information, human capital investment and intergenerational transfers). Other quantitative methodologies (e.g., wage equation and occupational segregation index) and qualitative methodologies (e.g., audit studies) could, moreover, be applied to analyze occupational and wage differences among ethnic groups. Audit studies are a promising quasi-experimental approach to measure employer discrimination more accurately. Finally, further research is needed to understand why females and the elderly have a lower probability of non-agricultural employment.

The diamond of theories for measuring ethnic differences in occupational outcomes and wages gives a comprehensive theoretical framework and an optimal starting point for further theoretical extensions (figure 2-4). For example as yet no one has formally included employer discrimination as a constraint in farm household models and in non-farm rural employment theories. Because of employer discrimination, ethnic minorities could be in a distress-push situation and could be forced to either work in badly paid non-agricultural jobs or to stay in agriculture. Employer discrimination, therefore, could influence time allocation of ethnic minorities in farm household models.

To get more accurate results, it is crucial for empirical and for theoretical extensions of occupational outcome analysis to consider the decision-making processes which determine final occupational outcomes. BLAU et al. already pointed out this need in 1956. The empirical implementation to measure decision-making processes requires observing individuals' job decisions throughout their life. To apply this kind of research involves a sustainable partner (e.g., university or research institute) in the country of interest and interviewees who agree to be observed from preschool to the time their final job decision is made. This requires long-term funding and a persistent research team.

Other studies could be conducted regarding the classification of ethnic groups and of autonomous areas in China. Interested researchers may study ethnic identity, especially in those groups who have not officially received their own ethnic status, but who are classified as another ethnic group. In Guizhou this is the case for 15 groups (ZHOU, 2003, p. 14-15). The question remains to be answered as to whether these groups have kept their original ethnic identity or whether they have assimilated into the cultural mainstream. In future studies about cultural identity, the manners and dress of young ethnic minority women could be essential to finding out about preservation of traditional clothing. Another important topic to look at

could be the relationship between Han of Tunpu and people of the surrounding ethnic minority villages; it could be interesting to investigate whether ethnic minorities still regard the Tunpu as conquerors and occupants.

Future research on Guizhou may also concern many other open questions (chapter three). In rural Guizhou more researchers could be fruitfully investigate the following topics: the infrastructural development and safety regulations of roads and railways, the influence of western media and of western commodities on locals, people's awareness of environmental pollution, cow horns' symbolic meaning for the Miao, the importance of ethnic festivals for locals, etc.

EXECUTIVE SUMMARY

This monograph investigates the presence of labor market discrimination against ethnic minorities in rural Guizhou (province in southwestern China). The focus is on two key research questions: 1) are minorities, due to their ethnic affiliation, discriminated against in the rural labor market in Guizhou, and 2) to what extent is discrimination empirically measurable.

The theoretical approaches for analyzing the first question can be arranged in a diamond of theories, which postulates that there are four major reasons for ethnic differences in occupational outcomes: differences in employer discrimination, in institutional discrimination, in abilities and in preferences. The complexity of these four theoretical concepts requires assumptions for empirical application (see subchapter 2.5.1). I made assumptions regarding preferences and institutional discrimination: 1) in Guizhou agricultural work is seen as inferior work; therefore, all individuals regardless of their ethnic affiliation prefer working in non-agricultural than in agricultural employment, 2) China has a preferential policy framework to improve ethnic minorities' education and employment outcomes; therefore, there is no institutional discrimination (discrimination by law) against ethnic minorities in China.

The major theoretical approaches which serve to explain ethnic differences in occupational outcomes in rural Guizhou are, thus, approaches which analyze differences in employer discrimination and differences in abilities. Geographic location is another important determinant for accessing different employment possibilities. Altogether ethnic minorities, thus, face three possible constraints for accessing non-agricultural employment: 1) individual abilities are not adequate to perform non-agricultural work, 2) employers in the non-agricultural sector discriminate against ethnic minority workers, and 3) non-agricultural work is not available in the area. In line with these three constraints, I developed hypotheses to test whether human capital factors and geographic characteristics have an effect on the probability of accessing non-agricultural employment.

Labor market discrimination against ethnic minorities is difficult to analyze empirically because such discrimination is forbidden in China, and field investigations about ethnic minorities are restricted; therefore, empirical research methods must be adapted to these circumstances. In the empirical application of ethnic differences in occupational outcomes, I used methodological triangulation by combining discrete choice modeling in the quantitative analysis and participant observation in the qualitative analysis. In the quantitative analysis I used cross-sectional data

(1997, 2000 and 2004) from the China Health and Nutrition Survey. I used discrete choice models to estimate effects of the independent variables [ethnic status (Bouyei, Miao, Tujia and Han), education, gender, age, counties and village size] on the binary dependent variable agriculture versus non-agriculture and their sub-groups considering mixed-ethnic communities. I included interaction terms between ethnic status and years of education when the overall model fit improved. In the qualitative analysis I applied participant observation in April 2010 in selected autonomous areas of Bouyei and Miao groups. I acquired additional knowledge regarding my investigations through conversations with foreign and Chinese scholars of Guizhou University during winter term 2010.

Quantitative and qualitative methodologies both have shortcomings in the measurement of ethnic differences in occupational outcomes, particularly in disentangling effects from preferences, abilities and discrimination; therefore, labor market discrimination against ethnic minorities can only be identified with discrete choice models by using well-defined assumptions about relationships among variables; moreover, labor market discrimination against ethnic minorities is difficult to identify with observation because with observation techniques I could only distinguish the ethnic minorities from their traditional clothing. A deeper analysis of labor market discrimination requires long-term fieldwork which was not possible in the framework of this project due to the limitations of available funding and time, but also to the high political sensitivity of the overall topic.

In Guizhou agriculture is accomplished with traditional methods. The major reason is that the mountainous terrain of the province makes it impossible to employ modern technologies to do the work more efficiently. Non-agricultural employment is mainly in the construction and in the service sectors. Estimation results indicate that between Han, the Tujia and the Bouyei there are almost no occupational differences. The Miao have a higher probability of working in agriculture than do Han, yet in 2004 better educated Miao have a higher probability of working in non-agriculture in general and in blue-collar positions in particular than do Han. Secondary occupations alongside agriculture are important for all three ethnic minorities; Guizhou's tourism industry particularly offers secondary occupations for ethnic minorities because their traditional lifestyle is a major tourist attraction. Regarding the other factors, I found that more years of education positively influence access to non-agricultural employment, that women and the elderly are underrepresented in non-agricultural employment and that smaller villages are linked to a higher probability of agricultural employment. Better educated ethnic minorities have a similar or even a higher probability of accessing non-agricultural employment than do Han; therefore, employers' prejudices cannot be against the ethnic affiliation but must be against less education and against factors linked to less education, such as an inadequate *wenhua* (cultural) level and weaker Mandarin skills among other factors (see figure 2.5); this implies that an inadequate *wenhua* level and inadequate language acquisition are both causes and symptoms of the

Miao's different employment outcomes. The higher share of Miao in agricultural employment could be explained by their more traditional lifestyle and lower level of sinicization. To increase the number of Miao in non-agricultural employment requires improving their education and also their Mandarin skills; however, in this process the Miao may lose their own cultural values and language, which may in turn lead to resentments against Han. The cultural and social variety that lurks behind the ethnic label "Miao" demands additional inquiries.

A preferential policy framework to improve ethnic minorities' labor market access has already been implemented in China; it seeks to overcome ethnic differences in abilities and in employer discrimination. There is, however, evidence that Han see as unfair the preferential policies for ethnic minorities. Han might believe that ethnic minorities are assigned university places and jobs because of preferential policies and not because of ethnic minorities' qualifications. Preferential policies may, therefore, intensify Han's prejudices about ethnic minorities' abilities. Ethnic uprisings in Tibet Autonomous Area and in Xinjiang Uyghur Autonomous Area indicate that Tibetans and Uyghurs do not approve of current policies either. Tibetans and Uyhgurs see their own cultural identities as at risk and despite preferential policies feel that they are discriminated against in accessing better-paying jobs.

These facts imply that integrating ethnic minorities in labor markets requires mutual understanding and respect among the groups. It might be important to strengthen policies which focus on individuals' vulnerability and poverty levels and not on ethnic affiliation; this may reduce ethnic resentments. The government should improve the urbanization rate and should encourage workers from impoverished rural areas to find employment in better developed areas because in times of recession scarcer capital resources might be allocated to more densely populated areas.

To get a complete overview of the 55 ethnic minorities in China's labor market, future research should also focus on analyzing occupational outcomes of other ethnic minorities. It appears to be a promising research field to observe the decision-making process which determines occupational outcomes. To improve theoretical understanding of ethnic differences in occupational outcomes in rural areas, farm household models and non-farm rural employment theories should be combined with employer discrimination theories.

ZUSAMMENFASSUNG

Die vorliegende Arbeit untersucht das Vorhandensein von Arbeitsmarktdiskriminierung gegenüber ethnischen Minderheiten im ländlichen Guizhou (Provinz im Südwesten Chinas). Im Mittelpunkt stehen dabei zwei zentrale Forschungsfragen: Werden ethnische Minderheiten aufgrund ihrer ethnischen Zugehörigkeit im ländlichen Arbeitsmarkt in Guizhou diskriminiert? In welchem Umfang können wissenschaftliche Forschungsmethoden Arbeitsmarktdiskriminierung wirklich messen?

Theorien zur Messung von ethnischen Unterschieden in der Arbeitsplatzwahl und in der Lohnverteilung lassen sich in einem Diamanten darstellen, der vier Hauptgründe für ethnische Unterschiede bei der Arbeitsplatzwahl sowie der Lohnverteilung hervorhebt: Unterschiede durch Arbeitgeberdiskriminierung, institutionelle Diskriminierung (durch das Gesetz), Fähigkeiten und Präferenzen. Da die vier theoretischen Zweige eng miteinander verwachsen sind, müssen fundierte Annahmen zum besseren Verständnis der Kausalbeziehungen getroffen werden.

Im Rahmen dieser Arbeit werden zwei Grundannahmen unterstellt. Die erste Annahme ist, dass in Guizhou unabhängig von ethnischer Zugehörigkeit bevorzugt außerlandwirtschaftliche Tätigkeiten wahrgenommen werden, also eine Präferenz für außerlandwirtschaftliche Beschäftigung vorliegt, da Landwirtschaft als minderwertige Beschäftigung angesehen wird. Die zweite Annahme basiert auf der Tatsache, dass ethnische Minderheiten in China nicht durch das Gesetz diskriminiert werden, sondern durch Antidiskriminierungs-Gesetze und Arbeitsförderungsmaßnahmen geschützt werden. Demzufolge sind ethnische Unterschiede in der Arbeitsplatzwahl und Lohnverteilung mit Arbeitgeberdiskriminierung und unterschiedlichen Fähigkeiten begründbar.

Der Zugang zur außerlandwirtschaftlichen Beschäftigung ist für ethnische Minderheiten im ländlichen Guizhou somit beschränkt, wenn 1) individuelle Fähigkeiten nicht den Anforderungen der außerlandwirtschaftlichen Tätigkeit entsprechen, 2) Arbeitgeber des Sektors gegen ethnische Minderheiten diskriminieren und 3) außerlandwirtschaftliche Beschäftigung im Einzugsgebiet nicht vorhanden ist. Im Einklang mit diesen drei Bedingungen wurden Hypothesen über die wichtigsten Humankapital-Faktoren (ethnischer Status, Bildung, Geschlecht und Alter) und dem geographischen Standort entwickelt.

Weil Arbeitgeberdiskriminierung in China allerdings verboten ist und Forschung über ethnische Minderheiten somit nur eingeschränkt möglich ist, müssen empirische Forschungsmethoden entsprechend den Gegebenheiten angepasst werden.

Zur Untersuchung der Hypothesen wurden eine Kombination aus diskreten Wahlmodellen im quantitativen Teil und teilnehmender Beobachtung im qualitativen Teil herangezogen. Die quantitative Analyse basiert auf Querschnittsdaten (1997, 2000 und 2004) des China Health and Nutrition Survey. Mit diskreten Wahlmodellen wurden die Auswirkungen der unabhängigen Variablen [ethnischer Status (Bouyei, Miao, Tujia und Han), Bildung, Geschlecht, Alter, Landkreise und Dorfgröße] auf die binären abhängigen Variablen, landwirtschaftliche Beschäftigung gegenüber außer-landwirtschaftlicher Beschäftigung und deren Untergruppen in ethnisch gemischten Gemeinden untersucht. Wenn es zu einer Modellverbesserung führte, wurden Interaktionsterme zwischen ethnischem Status und Bildung der Regression hinzugefügt. Die qualitative Analyse basiert auf teilnehmenden Beobachtungen, die im April 2010 in ausgewählten Bouyei- und Miao-Gebieten durchgeführt wurden, und auf Befragungen von Wissenschaftlern der Universität Guizhou, die im Wintersemester 2010 durchgeführt wurden.

Die angewendeten quantitativen und qualitativen Forschungsmethoden eignen sich nur unter den gegebenen Annahmen, um Arbeitsmarktdiskriminierung von ethnischen Minderheiten zu messen. Es ist besonders schwierig, Diskriminierung, Präferenzen und Humankapital-Faktoren in der empirischen Analyse zu entkoppeln. Arbeitsmarktdiskriminierung von ethnischen Minderheiten ist mit diskreten Wahlmodellen nur mit Hilfe von Annahmen über die Kausalbeziehungen, die aus der Theorie hervorgehen, messbar. Arbeitsmarktdiskriminierung ist mit Beobachtung schwer identifizierbar. Die Methode eignet sich lediglich zur Feststellung des ethnischen Status' und dies auch nur, wenn ethnische Minderheiten traditionelle Kleidung tragen. Eine tiefere Untersuchung von Arbeitsmarktdiskriminierung erfordert langfristige Feldforschung, die im Rahmen dieses Projekts so nicht möglich war, da nicht genug Mittel und Zeit vorhanden waren. Arbeitsmarktdiskriminierung von ethnischen Minderheiten ist ein politisch hoch sensibles Thema in China; die Untersuchung war somit nur eingeschränkt möglich.

Feldbeobachtungen in Guizhou zeigten, dass landwirtschaftliche Tätigkeiten sehr traditionell mit Hilfe von einfachen Arbeitsgeräten ausgeübt werden. Dies kann damit begründet werden, dass es Guizhou's Gebirgsstruktur unmöglich macht, effizientere Technologien zur Erleichterung der Arbeit einzusetzen. Außerlandwirtschaftliche Beschäftigung ist vorwiegend im Bau- und Dienstleistungssektor angesiedelt. Die Schätzergebnisse deuten darauf hin, dass zwischen Han, Tujia und Bouyei keine wesentlichen Unterschiede in den Berufstätigkeiten vorliegen. Die Miao haben im Vergleich mit den Han eine höhere Wahrscheinlichkeit, in der Landwirtschaft tätig zu sein. Besser ausgebildete ethnische Minderheiten haben aber im Vergleich mit den Han oftmals sogar eine höhere Wahrscheinlichkeit, im außerlandwirtschaftlichen Sektor zu arbeiten. Alle drei ethnischen Minderheiten sind darüber hinaus auf Nebentätigkeiten außerhalb der Landwirtschaft angewiesen. Tourismus als ein Bereich des Dienstleistungssektors ist für ethnische Minderheiten in Guizhou ein wichtiger Beschäftigungsmotor, weil der traditionelle

Lebensstil der ethnischen Minderheiten im Zentrum touristischer Attraktionen steht. Die Untersuchung der anderen Faktoren ergab, dass Bildung den Zugang zu außerlandwirtschaftlicher Beschäftigung positiv beeinflusst, dass Frauen und ältere Arbeitnehmer in außerlandwirtschaftlicher Beschäftigung unterrepräsentiert sind und dass in kleineren Dörfern eher in der Landwirtschaft gearbeitet wird als in größeren und besser vernetzten Dörfern.

Die höhere Wahrscheinlichkeit einer außerlandwirtschaftlichen Beschäftigung von besser gebildeten ethnischen Minderheiten deutet darauf hin, dass Vorurteile von Arbeitgebern nicht direkt gegen den ethnischen Minderheitsstatus gerichtet sind, sondern gegen den geringen Bildungsstand, der mit Kulturstandards (*wenhua*) und Mandarin-Sprachkenntnissen in Zusammenhang steht. Unterschiedliche Kulturstandards und unzureichende Bildung stellen neben anderen Faktoren (siehe Abbildung 2.5) somit Ursachen und Symptome für die landwirtschaftlich geprägte Beschäftigungstruktur der Miao dar. Der größere Miao-Anteil in landwirtschaftlicher Beschäftigung lässt sich somit durch ihre vergleichsweise stärkere traditionelle Lebensweise und geringere Sinisierung erklären. Um die kulturelle Vielfalt, die sich hinter dem ethnischen Label "Miao" verbirgt, besser zu verstehen, ist zusätzliche Feldforschung erforderlich. Um gleiche Bedingungen im Zugang zur außerlandwirtschaftlichen Beschäftigung zu erhalten, sollten die Miao besser in das chinesische Schulsystem integriert werden. Mit diesem Schritt können allerdings die eigene Kultur und Sprache der Miao verloren gehen, was zu Ressentiments gegenüber den Han führen könnte.

Eine Verbesserung des Zugangs zu Bildung und Arbeit von ethnischen Minderheiten ist im chinesischen Gesetz festgelegt und wird durch Fördermaßnahmen unterstützt, um ethnische Unterschiede in Fähigkeiten und Arbeitgeberdiskriminierung zu verringern. Die Feldanalyse zeigte jedoch, dass sich Han aufgrund der Arbeitsförderung von ethnischen Minderheiten benachteiligt fühlen. Negative Einstellungen zum Bildungsniveau von ethnischen Minderheiten können somit durch Politikmaßnahmen sogar verstärkt werden, weil ethnische Minderheiten zum Beispiel mit einem niedrigeren Notendurchschnitt ein Studium aufnehmen können. Damit bleibt ein latenter Unterschied im Bildungsniveau zwischen ethnischen Minderheiten und Han bestehen. Die Fördermaßnahmen können daher Vorurteile der Han gegenüber den Fähigkeiten von ethnischen Minderheiten schüren. Ethnische Aufstände in den Autonomen Regionen Tibet und Xinjiang zeigen außerdem, dass Tibeter und Uiguren die gegenwärtigen Politikmaßnahmen gleichfalls nicht tolerieren. Tibeter und Uiguren beklagen den Verlust ihrer kulturellen Identität und eine Unterrepräsentation, trotz der Anti-Diskriminierungsgesetze und Fördermaßnahmen, in besser bezahlten Berufen.

Diese Tatsachen deuten darauf hin, dass Arbeitsmarktintegration von ethnischen Minderheiten nur durch gegenseitiges Verständnis und Respekt zwischen den Gruppen erreicht werden kann. Fördermaßnahmen sollten auf Schutzbedürftigkeit und Armut fokussieren und nicht auf ethnische Zugehörigkeit. Auf diese Weise

können ethnische Feindseligkeiten vielleicht reduziert werden. Die Regierung sollte die Urbanisierungsrate erhöhen und Arbeitnehmer aus verarmten ländlichen Gebieten motivieren, Arbeit in besser entwickelten Regionen aufzunehmen, da in Zeiten der Rezession knappe öffentliche Mittel dort ausgegeben werden, wo sie den meisten Menschen nutzen.

Um ein umfassendes Bild der Arbeitsmarktsituation von ethnischen Minderheiten in China zu erlangen, sind zusätzliche empirische Analysen der insgesamt 55 ethnischen Minderheiten erforderlich. Entscheidungstheoretische Ansätze zum besseren Verständnis von Kulturunterschieden in der Berufsfindung stellen ein weiteres vielversprechendes Forschungsfeld dar. Um ein besseres theoretisches Verständnis von ethnischen Berufsunterschieden in ländlichen Gebieten zu erlangen, sollten landwirtschaftliche Haushaltsmodelle und Theorien der Diversifikation im außerlandwirtschaftlichen Sektor mit Theorien der Arbeitsmarktdiskriminierung verbunden werden.

REFERENCES

AHITUV, A., KIMHI, A. (2002): Off-farm work and capital accumulation decisions of farmers over the life-cycle: the role of heterogeneity and state dependence, *Journal of Development Economics, Vol. 68,* pp. 329-353.

AI, C., NORTON, E. C. (2003): Interaction terms in logit and probit models, *Economics Letters, Vol. 80*, pp. 123-129.

AIGNER, D. J., CAIN, G. G. (1977): Statistical theories of discrimination in labor markets, *Industrial and Labor Relations Review, Vol. 30,* pp. 175-187.

AJZEN, I. (1985): From intentions to actions: A theory of planned behavior, in: KUHI, J., BECKMANN, J. (eds.): Action-control: From cognition to behavior, Heidelberg: Springer Verlag, pp. 11-39.

ALDRICH, J. H., NELSON, F. D. (1984): Linear probability, logit, and probit models, Newbury Park, CA: Sage.

ALTONJI, J. G., BLANK, R. M. (1999): Race and gender in the labor market, in: ASHENFELTER, O., CARD, D. (eds.): Handbook of labor economics, Amsterdam: Elsevier, Vol. 3c, pp. 3143-3259.

ARROW, K. (1971): The theory of discrimination, working paper 30a, Industrial Relations Section, Princeton University.

ARROW, K. (1973): The theory of discrimination, in: ASHENFELTER, O. A., REES, A. (eds.): Discrimination in labor markets, Princeton: Princeton University Press, pp. 3-33.

ASTEBRO, T., CHEN, J., THOMPSON, P. (2008): Stars and misfits: A theory of occupational choice, working paper, available at http://ssrn.com/abstract=1113048.

BARRETT, C., REARDON, T., WEBB, P. (2001): Non-farm income diversification and household livelihood strategies in rural Africa: Concepts, dynamics and policy implications, *Food Policy, Vol. 26(4)*, pp. 315-332.

BASU, K., TZANNATOS, Z. (2003): The global child labor problem: What do we know and what can we do?, *The World Bank Economic Review, Vol. 17(2),* pp. 147-173.

BAUM, C. F. (2010): Factor variables and marginal effects in Stata 11, http://fmwww.bc.edu /ec-c/s2011/327/327facmarg.beamer.slides.pdf, [accessed 15 January 2011].

BECKER, G. S. (1965): A Theory of the allocation of time, *Economic Journal, Vol. 75(299),* pp. 493-517.

BECKER, G. S. (1957, 1971): The economics of discrimination, 2nd edition. Chicago, IL: The University of Chicago Press.

BECQUELIN, N. (2000): Xinjiang in the Nineties, *The China Journal, Vol. 44*, pp. 65-90.

BELLÉR-HANN, I. (1997): The peasant condition in Xinjiang, *Journal of Peasant Studies, Vol. 25(1),* pp. 87-112.

BELLÉR-HANN, I. (1998a): Crafts, entrepreneurship and gendered economic relations in Southern Xinjiang in the era of "socialist commodity economy", *Central Asian Survey, Vol. 17(4)*, pp. 701-718.

BELLÉR-HANN, I. (1998): Work and gender among Uyghur villagers in Southern Xinjiang, *Cahiers d'Etudes sur la Méditerranée Orientale et le monde Turco-Iranien* [En ligne], 25 | 1998, mis en ligne le. URL: http://cemoti.revues.org/55.

BENJAMIN, C., CORSI, A., GUYOMARD, H. (1996): Modelling labour decisions of French agricultural households, *Applied Economics, Vol. 28(12)*, pp. 1577-1589.

BENJAMIN, C., KIMHI, A. (2006): Farm work, off-farm work, and hired farm labour: Estimating a discrete-choice model of French farm couples' labour decisions, *European Review of Agricultural Economics, Vol. 33(2)*, pp. 149-171.

BERGMANN, B. R. (1974): Occupational segretation, wages and profits when employers discriminate by race or sex, *Eastern Economic Journal, Vol. 1(1,2)*, pp. 103-110.

BHALLA, A. S., QUI, S. (2006): Poverty and inequality among Chinese minorities, Cambridge: University of Cambridge.

BJERK, D. (2007): The differing nature of black-white wage inequality across occupational sectors, *Journal of Human Resources, Vol. XLII(2)*, pp. 398-434.

BLACK, D. A. (1995): Discrimination in an equilibrium search model, *Journal of Labor Economics, Vol. 13(2)*, pp. 309-334.

BLAU, P. M., GUSTAD, J. W., JESSOR, R., PARNES, H. S., WILCOCK, R. C. (1956): Occupational choice: A conceptual framework, *Industrial and Labor Relations Review, Vol. 9(4)*, pp. 531-543.

BLINDER, A. S. (1973): Wage discrimination: reduced form and structural variables, *Journal of Human Resources, Vol. 8*, pp. 436-455.

BOSKIN, M. J. (1974): A conditional logit model of occupational choice, *Journal of Political Economy, Vol. 82*, pp. 389-398.

BRAUN, A. J. (2005): Between integration and marginalisation: Women's issues in the labour market policies of the PR China, in: LEUTNER, M. (ed.): Berliner China Hefte, Chinese History and Society, Berlin: Free University of Berlin, Vol. 28, pp. 110-122.

BROSIG, S., GLAUBEN, T., HERZFELD, T., ROZELLE, S., WANG, X. (2007): The dynamics of Chinese rural household's participation in labor markets, *Agricultural Economics, Vol. 37*, pp. 167-178.

BROWN, M. J. (2001): Ethnic classification and culture: The case of the Tujia in Hubei, China, *Asian Ethnicity, Vol. 2(1)*, pp. 55-71.

BROWN, S., FRY, T. R. L., HARRIS, M. N. (2008): Untangling supply and demand in occupational choice, *Economics Letters, Vol. 99*, pp. 414-417.

BUCHENRIEDER, G., ERJAVEC, E., JUVANCIC, L., KNÜPFER, J. (2001): Summary of factors influencing non-farm income diversification, Stuttgart: University of Hohenheim, Department of Agricultural Development Theory and Policy in the Tropics and Subtropics.

BUCHENRIEDER, G., MÖLLERS, J. (2006): A synthesis of theoretical concepts for analysing non-farm rural employment, poster paper prepared for presentation at the International Association of Agricultural Economists Conference, Gold Coast, Australia, August 12-18, 2006.

CAIN, G. G. (1986): The economic analysis of labor market discrimination: A survey, in: ASHENFELTER, O., LAYNARD, R. (eds.): Handbook of labor economics, Amsterdam: Elsevier, Vol. 1, pp. 693-785.

CAMERON, A. C., TRIVEDI, P. K. (2005): Microeconometrics, methods and applications, Cambridge: Cambridge University Press.

CAMERON, A. C., TRIVEDI, P. K. (2009): Microeconometrics using Stata, College Station, TX: Stata Press.

CAMPBELL, D. T., FISKE, D. W. (1959): Convergent and discriminant validation by the multitrait-multimethod matrix, *Psychological Bulletin, Vol. 56(2)*, pp. 81-105.

CHAKRAVARTY, S. R., SILBER, J. (2007): A generalized index of employment segregation, *Mathematical Social Sciences, Vol. 53*, pp. 185-195.

CHARLES, K. K., GURYAN, J. (2008): Prejudice and wages: an empirical assessment of Becker's the economics of discrimination, *Journal of Political Economy, Vol. 116(5)*, pp. 773-809.

CHAYANOV, V. A. (1986): The theory of peasant economy, Madison, WI: University of Wisconsin Press.

CHINA DAILY (2012): Big boost for poverty-stricken province, http://www.chinadaily.com.cn/m/guizhou/2012-03/13/content_14825569.htm, 13.03.2012, [accessed 13 March 2012].

CHINA LABOR BULLETIN (2011): Wages in China, http://www.clb.org.hk/en/node/100206, 28.12.2011, [accessed 24 May 2012].

CHINA.ORG.CN (2005): Regional autonomy for ethnic minorities in China, White Papers of the Chinese Government, http://www.china.org.cn/e-white/20050301/index.htm, [accessed 17 September 2009].

CHINA.ORG.CN (2012): China urbanization rate exceeds 50 percent, http://www.china.org.cn/china/2012-05/30/content_25516139.htm, 30.05.2012, [accessed 31 October 2012].

CITS (2010): Gateway to Guizhou, http://www.toguizhou.com/html/Minorities/140.html, [accessed 02 May 2010].

COATE, S., LOURY, G. (1993): Will affirmative-action policies eliminate negative stereotypes?, *American Economic Review, Vol. 83(5)*, pp. 1220-1240.

DARDEN, J. (2005): Black occupational achievement in the Toronto census metropolitan area: Does race matter?, *Review of Black Political Economy, Vol. 33(2)*, pp. 31-54.

DEATON, A. (1997): The analysis of household surveys. A microeconometric approach to development policy, Baltimore: The Johns Hopkins University Press (for the World Bank).

DE BRAUW, A., ROZELLE, S. (2007): Returns to education in rural China, New York: Routledge.

DE JANVRY, A., FAFCHAMPS, M., SADOULET, E. (1991): Peasant household behavior with missing markets: some paradoxes explained, *The Economic Journal, Vol. 101*, pp. 1400-1417.

DEUTSCHE BANK (2012): Emerging markets, China chartbook, Guizhou province, 03.2012, http://www.dbresearch.de/PROD/DBR_INTERNET_DE-PROD/PROD000 0000000 2475 22.pdf, [accessed 13 March 2011].

DIAMOND, N. (1995): Defining the Miao: Ming, Qing and contemporary views, in: HARRELL, S. (ed.): Cultural encounters on China's ethnic frontiers, Seattle: University of Washington Press, pp. 92-116.

DING, C. (2003): Land policy reform in China: assessment and prospects, *Land Use Policy, Vol. 20*, pp. 109-120.

DROST, A. (2002): The dynamics of occupational choice: Theory and evidence, *Labour Review of Labour Economics and Industrial Relations, Vol. 16*, pp. 201-233.

DUNCAN, O. D., DUNCAN, B. (1955): A methodological analysis of segregation indexes, *American Sociological Review, Vol. 20(2)*, pp. 210-217.

EFSTRATOGLOU-TODOULOU, S. (1990): Pluriactivity in different socio-economic contexts: A test of the push-pull hypothesis in Greek farming, *Journal of Rural Studies, Vol. 6(4)*, pp. 407-413.

ELLIS, F. (1993): Peasant Economics, 2nd ed., Cambridge: Cambridge University Press.

FAFCHAMPS, M. (1992): Cash crop production, food price volatility, and rural market integration in the third world, *American Journal of Agricultural Economics, Vol. 74*, pp. 90-99.

FAIRLIE, R. W. (1999): The absence of the African-American owned business: An analysis of the dynamics of self-employment, *Journal of Labor Economics, Vol. 17(1)*, pp. 80-108.

FAIRLIE, R. W. (2006): An extension of the Blinder-Oaxaca decomposition technique to Logit and Probit models, Institute for the Study of Labor IZA DP no. 1917.

FAZ (2008): Chinas Vorgehen in Tibet: SPD fordert Wirtschaftssanktionen, http://www.faz.net/-gq5-wgkn, 26.03.2008, [accessed 26 November 2011].

FEIGL, H. (1958): The mental and the physical, in: FEIGL, H., SCRIVEN, M., MAXWELL, G. (eds.): Concepts, theories and the mind-body problem, Minneapolis: University of Minnesota Press, Vol. 2.

FINDEIS, L. J., LASS, A. D. (1994): Labor decisions by agricultural households: Interrelationships between off-farm labor supply and hired labor demand, *PRI working paper 94-08*, University Park, Pennsylvania.

FINKELSHTAIN, I., CHALFANT, A. J. (1991): Marketed surplus under risk: Do peasants agree with Sandmo?, *American Journal of Agricultural Economics, Vol. 73*, pp. 557-567.

FLICK, U. (1995): Qualitative Forschung. Theorie, Methoden, Anwendung in Psychologie und Sozialwissenschaften, Reinbek: Rowohlt.

FLÜCKIGER, Y., SILBER, J. (1999): The measurement of segregation in the labor force, Heidelberg: Physica Verlag.

FN (2010): Fieldnotes conducted by Bente Castro Campos (March, 27 to April, 29 2010 in Guizhou province).

GERBER, C. (2008): Vielfalt auf Chinesisch, http://www.zeit.de/online/2008/14/china-minderheiten, 27.03.2008, [accessed 11 December 2011].

GIGERENZER, G., SELTEN, R. (2001): Rethinking rationality, in: GIGERENZER, G., SELTEN, R. (eds.): Bounded Rationality, the adaptive toolbox, Cambridge, MA: MIT Press, pp. 1-12.

GILLEY, B. (2001): Uighurs need not apply, *Far Eastern Economic Review*, 23.08.2001, pp. 26-27.

GLADNEY, D. C. (2004): Dislocating China: reflections on Muslims, minorities, and other subaltern subjects, London: Hurst.

GLEWWE, P., KREMER, M. (2006): Schools, teachers, and education outcomes in developing countries, in: HANUSHEK, E. A., WELCH, F. (eds.): Handbook of the Economics of Education, Amsterdam: Elsevier, Vol. 2, pp. 945-1017.

GOTTFREDSON, L. S. (1981): Circumscription and Compromise: A developmental theory of occupational aspirations, *Journal of Counseling Psychology, Vol. 28(6)*, pp. 545-579.

GOTTFREDSON, L. S. (1996): Gottfredson's theory of circumscription and compromise, in: BROWN, D., BROOKS, L. (eds.): Career choice and development, San Francisco: Jossey-Bass, Vol. 3, pp. 179-232.

GOULD, W. W. (2000): Interpreting logistic regression in all its forms, Stata Technical Bulletin 53, pp. 19-29, reprinted in Stata Technical Bulletin Reprints, College Station, TX: Stata Press, Vol. 9, pp. 257-270.

GREENE, W. H. (2008): Econometric Analysis, 6th ed, Upper Saddle River, NJ: Prentice Hall.

GTAI, GERMANY TRADE AND INVEST (2011): Das Perlflussdelta, Bestell-Nr.: 16101, www.gtai.de.

GUSTAFSSON, B., LI, S. (2003): The ethnic minority-majority income gap in rural China during transition, *Economic Development and Cultural Change, Vol. 51*, pp. 806-822.

GUSTAFSSON, B., SAI, D. (2006): Villages where China's ethnic minorities live, Institute for the Study of Labor IZA DP no. 2418.

GUSTAFSSON, B., SAI, D. (2008): Temporary and persistent poverty among ethnic minorities and the majority in rural China, Institute for the Study of Labor IZA DP no. 3791.

HAM, R., JUNANKAR, P. N. R., WELLS, R. (2009): Occupational choice: Personality matters, Institute for the Study of Labor IZA DP no. 4105.

HANNUM, E., XIE, Y. (1998): Ethnic stratification in Northwest China: Occupational differences between Han Chinese and National Minorities in Xinjiang, 1982-1990, *Demography, Vol. 35(3)*, pp. 323-333.

HARRELL, S. (1995): The history of the history of the Yi, in: HARRELL, S. (ed.): Cultural encounters on China’s ethnic frontiers, Seattle: University of Washington Press, pp. 63-91.

HARRELL, S. (1990): Ethnicity, local interests, and the State: Yi communities in Southwest China, *Comparative Studies in Society and History, Vol. 32(3)*, pp. 515-548.

HAUSMAN, J., MCFADDEN, D. (1984): Specification tests for the Multinomial Logit Model, *Econometrica, Vol. 52(5)*, pp. 1219-1240.

HECKMAN, J. J., SIEGELMAN, P. (1993): The Urban Institute audit studies: Their methods and findings, in: FIX, M., STRUYCK, R. J. (eds.): Clear and convincing evidence: Measurement of discrimination in America, Washington, DC: Urban Institute Press.

HENSHER, D. A., ROSE, J. M., GREENE, W. H. (2005): Applied choice analysis, a primer, Cambridge: Cambridge University Press.

HILLMAN, B. (2008): Money can´t buy Tibetans' love, *Far Eastern Economic Review*, pp. 8-12.

HOPPER, B., WEBBER, M. (2009): Migration, modernisation and ethnic estrangement: Uyghur migration to Urumqi, Xinjiang Uyghur Autonomous Region, PRC, *Inner Asia, Vol. 11*, pp. 173-203.

HOSMER, D. W., LEMESHOW, S. (2000): Applied logistic regression, 2nd ed., New York: Wiley.

HTTP://BAIKE.BAIDU.COM/VIEW/3537.HTM, HTTP://WWW.NCIKU.COM/SEARCH/ZH/DETAIL/文化/1315983: Definition of Wenhua, [accessed November 2011].

HTTPS://WWW.CPC.UNC.EDU/PROJECTS/CHINA: China Health and Nutrition Survey, [accessed October 2008].

HTTP://WWW.CPC.UNC.EDU/PROJECTS/CHINA/DATA/QUESTIONNAIRES: China Health and Nutrition Survey questionnaires, [accessed October 2008].

HTTP://WWW.CPC.UNC.EDU/PROJECTS/CHINA/PROJ_DESC/SURVEY: China Health and Nutrition Survey project description, [accessed February 2011].

HTTP://WWW.MERRIAM-WEBSTER.COM/DICTIONARY/SINICIZE: Definition of sinicize, [accessed August 2012].

HUFFMAN, E. W. (1980): Farm and off-farm work decision: the role of human capital, *The Review of Economics and Statistics, Vol. 62*, pp. 14-23.

HUFFMAN, E. W., LANGE, M. D. (1989): Off-farm work decisions of husbands and wives: Joint decision making, *The Review of Economics and Statistics, Vol. 71(3)*, pp. 471-480.

HUTCHENS, R. M. (1991): Segregation curves, Lorenz curves and inequality in the distribution of people across occupations, *Mathematical Social Sciences, Vol. 21*, pp. 31-51.

HUTCHENS, R. M. (2001): Numerical measures of segregation: Desirable properties and their implications, *Mathematical Social Sciences, Vol. 42*, pp. 13-29.

HUTCHENS, R. M. (2004): One measure of segregation, *International Economic Review, Vol. 45*, pp. 555-578.

JACOBS, G. (2007): An occupational choice model for developing countries, proceedings of the German Development Economics Conference, Göttingen 2007, 15, Verein für Sozialpolitik, Research Committee Development Economics.

JOHNSON, G. E., STAFFORD, F. P. (1998): Alternative approaches to occupational exclusion: in: PERSSON, I., JONUNG, C. (eds.), Women's work and wages, London: Routledge.

JONES, A. (2007): Applied econometrics for health economists: A practical guide, 2nd ed, Abingdon, UK: Radcliffe.

JONES, F. L. (1983): On decomposing the wage gap: A critical comment on Blinder's method, *Journal of Human Resources, Vol. 18(1)*, pp. 126-130.

JONES, F. L., MCMILLAN, J. (2001): Scoring occupational categories for social research: A review of current practice, with Australian examples, *Work Employment Society, Vol. 15(3)*, pp. 539-563.

JUNANKAR, P. N., MAHUTEAU, S. (2005): Do migrants get good jobs? New migrant settlement in Australia, *Economic Record, Vol. 81*, pp. 34-46.

KAI MING, C. (2003): Bildungsverwaltung, in: INSTITUT FÜR ASIEN STUDIEN HAMBURG (eds.): Das Grosse China Lexikon, Darmstadt: Wissenschaftliche Buchgesellschaft, pp. 103-104.

KARACA-MANDIC, P., NORTON, E. C., DOWD, B. (2012): Interaction terms in nonlinear models, *HSR: Health Services Research, Vol. 47(1)*, pp. 255-274.

KELLE, U. (1997): Empirisch begründete Theoriebildung: Zur Logik und Methodologie interpretativer Sozialforschung, Weinheim: Deutscher Studienverlag.

KEY, N., SADOULET, E., DE JANVRY, A. (2000): Transaction costs and agricultural household supply response, *American Journal of Agricultural Economics, Vol. 82*, pp. 245-259.

KIMURA, M., YASUI, D. (2006): Occupational choice, educational attainment, and fertility, *Economics Letters, Vol. 94*, pp. 228-234.

KLEIN, G. (2001): The fiction of optimization, in: GIGERENZER, G., SELTEN, R. (eds.): Bounded rationality: The adaptive toolbox, Cambridge, MA: MIT Press, pp. 103-122.

KLEINBAUM, D. G., KLEIN, M. (2002): Logistic regression: A self-learning text, 2nd ed, New York: Springer.

KOLONKO, P. (2005): Armut in China: Kein Wasser und nur einmal im Jahr Fleisch, http://www.faz.net/-gq5-qv0i, 03.07.2005, [accessed 21 December 2011].

KORF, B. (2004): Conflict, space and institutions property rights and the political economy of war in Sri Lanka, Aachen: Shaker Verlag.

KUPFER, K. (2011a): Konfliktportät: China – Tibet, Bundeszentrale für politische Bildung, http://www.bpb.de/themen/4GQA2G.html, 04.10.2011, [accessed 01 December 2011].

KUPFER, K. (2011b): Konfliktportät: China – Xinjiang, Bundeszentrale für politische Bildung, http://www.bpb.de/themen/324JYK.html, 17.10.2011, [accessed 01 December 2011].

LANG, K. (1986): A language theory of discrimination, *The Quarterly Journal of Economics, Vol. 101(2)*, pp. 363-382.

LARSON, L. M., ROTTINGHAUS, P. J., BORGEN, F. H. (2002): Meta-analyses of big six interests and big five personality factors, *Journal of Vocational Behavior, Vol. 61(2)*, pp. 217-239.

LASS, D. A., GEMPESAW II, C. M. G. (1992): The supply of off-farm labor: A random coefficients approach, *American Journal of Agricultural Economics, Vol. 74(2)*, pp. 400-411.

LE, A. T., MILLER, P. W. (2001): Occupational status: Why do some workers miss out?, *Australian Economic Papers, Vol. 40(3)*, pp. 352-372.

LIST, J. A., RASUL, I. (2011): Field experiments in labor economics, in: ASHENFELTER, O., CARD, D. (eds.): Handbook of labor economics, Amsterdam: Elsevier, Vol. 4a, pp. 103-228.

LONG, J. S. (1997): Regression models for categorical and limited dependent variables, Thousand Oaks, CA: Sage.

LONG, J. S., FREESE, J. (2003): Regression models for categorical dependent variables using Stata, 2nd ed, College Station, TX: Stata Press.

LOW, A. (1986): Agricultural development in Southern Africa: Farm household-economics and the food crisis, London: James Curtey.

LUNDBERG, S. J. (1991): The enforcement of equal opportunity laws under imperfect information: Affirmative action and alternatives, *Quarterly Journal of Economics, Vol. 106(1)*, pp. 309-326.

LUNDBERG, S. J., STARTZ, R. (1983): Private discrimination and social intervention in competitive labor markets, *American Economic Review, Vol. 73*, pp. 340-347.

LUO, L. (2008): Minderheiten in China: Wenn ein Han mit einer Miao, http://www.faz.net/-gdd-wy0c, 08.06.2008, [accessed 25 November 2011].

MACKERRAS, C. (2003): China's ethnic minorities and globalisation, Abingdon: Routledge Curzon.

MACKERRAS, C. (2004): Conclusion: Some major issues in ethnic classification, *China Information: A Journal on Contemporary China Studies, Vol. 18(2)*, pp. 303-313.

MAHUTEAU, S., JUNANKAR, P. N. (2008): Do migrants get good jobs in Australia? The role of ethnic networks in job search, *Economic Record, Vol. 84(1)*, pp. 115-130.

MAURER-FAZIO, M., HUGHES, J. W., ZHANG, D. (2004): The economic status of China's ethnic minorities, preliminary draft of paper being prepared for the International Research Conference: Poverty, inequality, labour market and welfare reform in China, Australia National University, August 25-27, 2004, http://econrsss.anu.edu.au/pdf/china-abstract-pdf/maurer-fazio.pdf.

MAURER-FAZIO, M., HUGHES, J. W., ZHANG, D. (2005): A comparison of reform-era labor force participation rates of China's ethnic minorities and Han majority, William Davidson Institute working paper no. 795.

MCKHANN, C. F. (1995): The Naxi and the nationalities question, in: Harrell, S. (ed.), Cultural encounters on China's ethnic frontiers, Seattle: University of Washington Press, pp. 39-62.

MENG, X., MILLER, P. (1995): Occupational segregation and its impact on gender wage discrimination in China's rural industrial sector, *Oxford Economic Papers, Vol. 47(1)*, pp. 136-155.

MICHAEL, R. T., BECKER, G. S. (1973): The new theory of consumer behaviour, *Swedish Journal of Economics, Vol. 75(4)*, pp. 378-396.

MÖLLERS, J. (2006): Außerlandwirtschaftliche Diversifikation im Transformationsprozess: Diversifikationsentscheidungen und -strategien ländlicher Haushalte in Slowenien und Mazedonien, Halle (Saale): Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO).

MOHAPATRA, S. (2004): Complementarities, constraints and contracts: Incentive design and occupational choice in China, Davis: University of California, Davis.

NATIONAL BUREAU OF STATISTICS OF CHINA (2009): 60 years of new China, Beijing: China Statistics Press.

OAXACA, R. L. (1973): Male-female wage differentials in urban labor markets, *International Economic Review, Vol. 14*, pp. 693-709.

OAXACA, R. L., RANSOM, M. L. (1999): Identification in detailed wage decompositions, *Review of Economic Statistics, Vol. 81(1)*, pp. 154-157.

OETTINGER, G. (1996): Statistical discrimination and the early career evolution of the black-white wage gap, *Journal of Labor Economics, Vol. 14(1)*, pp. 52-78.

ONLINE CAMBRIDGE DICTIONARY (2011): The meaning of choice, http://dictionary.cambridge.org/dictionary/british/choice_1?q=choice, [accessed 14 July 2011].

ONLINE CAMBRIDGE DICTIONARY (2012): the meaning of primary, secondary and tertiary industry, http://dictionary.cambridge.org/dictionary/business-english/primary-industry? q= primary+ industry, http://dictionary.cambridge.org/dictionary/business-english/secondary-industry?q=secondary+industry, http://dictionary.cambridge.org/dictionary/business-english/ tertiary industry ?q=tertiary+industry, [accessed August 2012].

PAMPEL, F. C. (2000): Logistic regression: a primer, Thousand Oaks, CA: Sage.

PEOPLE'S DAILY (2011): Guizhou province ready to rebuild its image, http://english.people daily.com.cn/90001/90776/90882/7278417.html, 01.02.2011, [accessed 12 March 2012].

PETRICK, M. (2004): Can econometric analysis make (agricultural) economics a hard science? Critical remarks and implications for economic methodology, *IAMO Discussion paper no. 62*, Halle (Saale): Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO).

PHELPS, E. S. (1972): The statistical theory of racism and sexism, *American Economic Review, Vol. 62*, pp. 659-661.

PILLSBURY, B. (1973): Cohesion and cleavage in a Chinese muslim minority, PhD dissertation, Columbia University.

PORTER, S., UMBACH, P. (2006): College major choice: An analysis of person-environment fit, *Research in Higher Education, Vol. 47(4)*, pp. 429-449.

PORTER, M. E. (1990): The competitive advantage of nations, *Harvard Business Review, March-April*, pp. 73-93.

POTTER, S. H. (1983): The position of peasants in modern China's social order, *Modern China, Vol. 9(4)*, pp. 465-499.

POWERS, D. A., XIE, Y. (2008): Statistical methods for categorical data analysis, 2nd ed., Bingley, UK: Emerald.

QUILLIAN, L. (2006): New approaches to understanding racial prejudice and discrimination, *Annual Reviews of Sociology, Vol. 32*, pp. 299-328.

REAP (2008): Rural Education Action Project, http://www.reapchina.org.

REINA, L. (2005): From subjective expected utility theory to bounded rationality: An experimental investigation on categorization processes in integrative negotiation in committees' decision making and in decisions under risk, Dresden: Technische Universität Dresden.

ROSENBLOOM, J. L., ASH, R. A., DUPONT, B., CODER, L. (2008): Why are there so few women in information technology? Assessing the role of personality in career choices, *Journal of Economic Psychology, Vol. 29(4)*, pp. 543-554.

ROSS, L., WOLL, R., ZHOU, K. (2007): Changes in China's labor law environment, http://www.wilmerhale.com/publications/whPubsDetail.aspx?publication=8011, [accessed 27 May 2012].

SADOULET, E., DE JANVRY, A. (1995): Quantitative development policy analysis, Baltimore: The Johns Hopkins University Press.

SADOULET, E., DE JANVRY, A., BENJAMIN, C. (1998): Household behavior with imperfect labor markets, *Industrial Relations, Vol. 37*, pp. 85-108.

SAUTMAN, B. (1997): Affirmative action, ethnic minorities and China's universities, working paper in the social sciences no. 13, Hongkong: Division of Social Sciences, The Hong Kong University of Sciences and Technology.

SCHAEFER, R. T. (2007): Sociology, 10th ed., Beijing, Guangzhou, Shanghai, Xian: McGraw-Hill.

SCHEIN, L. (2000): Minority rules: The Miao and the femine in China's cultural politics, Durham, London: Duke University Press.

SCHMIDT-LAUBER, B. (2007): Feldforschung: Kulturanalyse durch teilnehmende Beobachtung, in: GÖTTSCH, S., LEHMANN, A. (eds.): Methoden der Volkskunde, Berlin: Dietrich Reimer Verlag GmbH, Vol. 2, pp. 219-248.

SCHMIDT, P., STRAUSS, R. P. (1975): The prediction of occupation using multiple logit models, *International Economic Review, Vol. 16(2)*, pp. 471-486.

SCHNELL, R., HILL, P.B., ESSER, E. (1998): Methoden der empirischen Sozialforschung, 8th ed., Munich: Oldenbourg Verlag.

SCHOMANN, S. (2011): Innere Mongolei: Chinas Wilder Westen, http://www.zeit.de/reisen/2011-11/mongolei, 21.11.2011, [accessed 26 November 2011].

SEAWRIGHT, J., COLLIER, D. (2004): Glossary, in: BRADY, H. E., COLLIER, D. (eds.): Rethinking social inquiry: Diverse tools, shared standards, Lanham, MD: Rowman and Littlefied Publishers, Ine.

SIMON, H. A. (1956): Rational choice and the structure of the environment, *Psychological Review, Vol. 63(2)*, pp. 129-138.

SKOUFIAS, E. (1994): Using shadow wages to estimate labor supply of agricultural households, *American Journal of Agricultural Economics, Vol. 76*, pp. 215-227.

SPRADLEY, J. P. (1980): Participant observation, New York: Holt, Rinehart and Winston.

STALIN, J. V. (1913): The national question and social-democracy, *Prosveshchenie, Vol. 3*, p. 54.

STALIN, J. V. (1953): Marxism and the national question, *Works, Vol. 2*, p. 307.

START, D. (2001): The rise and fall of the rural non-farm economy: Poverty impacts and policy options, *Development Policy Review, Vol. 19(4)*, pp. 491-505.

STATISTICS SOLUTIONS (2012): Data levels of measurement, http://www.statisticssolutions.com/resources/dissertation-resources/descriptive-statistics/data-levels-of-measurement, [accessed 04 June 2012].

SUMNER, A. D. (1982): The off-farm labor supply of farmers, *American Journal of Agricultural Economics, Vol. 64*, pp. 499-509.

TACCHI, J., SLATER, D., HEARN, G. (2003): Ethnographic action research, New Delhi: UNESCO.

THE ECONOMIST (2000): Go west, young Han: Plans to develop China's western provinces are about more than economics, http://www.economist.com/node/457567/print, 21.12.2000, [accessed 21 December 2011].

THEESFELD, I. (2005): A common-pool resource in transition: Determinants of institutional change in Bulgaria's postsocialist irrigation sector, Aachen: Shaker Verlag.

TOKLE, J. G., HUFFMAN, W. E. (1991): Local economic conditions and wage labor decisions of farm and rural nonfarm couples, *American Journal of Agricultural Economics, Vol. 73(3)*, pp. 652-670.

TRAIN, K. E. (2009): Discrete choice methods with simulation, 2nd ed., Cambridge, UK: Cambridge University Press.

TREIMAN, D. J. (2009): Quantitative data analysis: Doing social research to test ideas, San Francisco, CA: Jossey-Bass.

WANG, X. (2007): Labor market behavior of Chinese rural households during transition, Halle (Saale): Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO).

WILLIAMS, R. (2012): Using the margins command to estimate and interpret adjusted predictions and marginal effects, *The Stata Journal, Vol. 12(2)*, pp. 308-331.

WOLFINGER, N. H. (2002): On writing fieldnotes: Collection strategies and background expectancies, *Qualitative Research, Vol. 2(1)*, pp. 85-93.

WORLD BANK (2008): Social assessment report IPP 301, Guizhou cultural and natural heritage protection and development project of the World Bank Loan.

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.

WORLD BANK (2012): GNI per capita, atlas method (current US$), http://data.worldbank.org/indicator/NY.GNP.PCAP.CD/countries/CN-4E-XT?display=default, [accessed 8 September 2012].

ZANG, X. (1998): Ethnic representation in the current Chinese leadership, *The China Quarterly, Vol. 153*, pp. 107-127.

ZANG, X. (2008): Market reforms and Han-muslim variation in employment in the Chinese state sector in a Chinese city, *World Development, Vol. 36(11)*, pp. 2341-2352.

ZHANG, D. D. (2003): Guizhou, in: INSTITUT FÜR ASIEN STUDIEN HAMBURG (eds.): Das Grosse China Lexikon, Darmstadt: Wissenschaftliche Buchgesellschaft, pp. 282-283.

ZHANG, L., HUANG, J., ROZELLE, S. (2002): Employment, emerging labor markets, and the role of education in rural China, *China Economic Review, Vol. 13*, pp. 313-328.

ZHOU, M. (2003): Multilingualism in China: The politics of writing reforms for minority languages 1949-2002, in: FISHMAN, J. A. (ed.): Contributions to the sociology of language, Berlin: Walter de Gruyter GmbH & Co. KG.

APPENDIX

Table A1: Major geographic distribution of ethnic minorities in PRC

Ethnic Minority Groups	Major Region by Geographic Distribution	Population Censuses (person)				
		2000	**1990**	**1982**	**1964**	**1953**
Mongolian	Inner Mongolia, Liaoning, Jilin, Hebei, Heilongjiang, Xinjiang	5,813,947	4,802,407	3,411,367	1,965,766	1,451,035
Hui	Ningxia, Gansu, Henan, Xinjiang, Qinghai, Yunnan, Hebei, Shandong, Anhui, Liaoning, Beijing, Inner Mongolia, Tianjin, Heilongjiang, Shanxi, Guizhou, Jilin, Jiangsu, Sichuan	9,816,805	8,612,932	7,228,398	4,473,147	3,530,498
Tibetan	Tibet, Sichuan, Qinghai, Gansu, Yunnan	5,416,021	4,593,541	3,847,875	2,501,174	2,753,081
Uygur	Xinjiang	8,399,393	7,207,024	5,963,491	3,996,311	3,610,462
Miao	Guizhou, Hunan, Yunnan, Guangxi, Chongqing, Hubei, Sichuan	8,940,116	7,383,622	5,021,175	2,782,088	2,490,874
Yi	Yunnan, Sichuan, Guizhou	7,762,272	6,578,524	5,453,564	3,380,960	3,227,750
Zhuang	Guangxi, Yunnan, Guangdong	16,178,811	15,555,820	13,383,086	8,386,140	6,864,585
Bouyei	Guizhou	2,971,460	2,548,294	2,119,345	1,348,055	1,237,714
Korean	Jilin, Heilongjiang, Liaoning	1,923,842	1,923,361	1,765,204	1,339,569	1,111,275
Manchu	Liaoning, Hebei, Heilongjiang, Jilin, Inner Mongolia, Beijing	10,682,262	9,846,776	4,304,981	2,695,675	2,399,228
Dong	Guizhou, Hunan, Guangxi	2,960,293	2,508,624	1,426,400	836,123	712,802
Yao	Guangxi, Hunan, Yunnan, Guangdong	2,637,421	2,137,033	1,411,967	857,265	665,933
Bai	Yunnan, Guizhou, Hunan	1,858,063	1,598,052	1,132,224	706,623	567,119
Tujia	Hunan, Hubei, Chongqing, Guizhou	8,028,133	5,725,049	2,836,814	524,755	–
Hani	Yunnan	1,439,673	1,254,800	1,058,806	628,727	481,220
Kazak	Xinjiang	1,250,458	1,110,758	907,546	491,637	509,375
Dai	Yunnan	1,158,989	1,025,402	839,496	535,389	478,966
Li	Hainan	1,247,814	1,112,498	887,107	438,813	360,950
Lisu	Yunnan, Sichuan	634,912	574,589	481,884	270,628	317,465
Va	Yunnan	396,610	351,980	298,611	200,272	286,158
She	Fujian, Zhejiang, Jiangxi, Guangdong	709,592	634,700	371,965	234,167	–
Gaoshan	Taiwan, Fujian	4,461	2,877	1,650	366	329
Lahu	Yunnan	453,705	411,545	304,256	191,241	139,060
Shui	Guizhou, Guangxi	406,902	347,116	286,908	156,099	133,566
Dongxiang	Gansu, Xinjiang	513,805	373,669	279,523	147,443	155,761
Naxi	Yunnan	308,839	277,750	251,592	156,796	143,453
Jingpo	Yunnan	132,143	119,276	92,976	57,762	101,852
Kirgiz	Xinjiang	160,823	143,537	113,386	70,151	70,944
Tu	Qinghai, Gansu	241,198	192,568	159,632	77,349	53,277
Daur	Inner Mongolia, Heilongjiang	132,394	121,463	94,126	63,394	–
Mulao[28]	Guangxi	207,352	160,648	90,357	42,819	–
Qiang	Sichuan	306,072	198,303	102,815	49,105	35,660
Blang	Yunnan	91,882	82,398	58,473	39,411	–

[28] ZHOU (2003, p. 13) uses the name Mulam instead of Mulao.

Salar	Qinghai	104,503	87,546	69,135	34,664	30,658
Maonan	Guangxi	107,166	72,370	38,159	22,382	–
Gelao	Guizhou	579,357	438,192	54,164	26,852	–
Xibe	Liaoning, Xinjiang	188,824	172,932	83,683	33,438	19,022
Achang	Yunnan	33,936	27,718	20,433	12,032	–
Pumi/Primi	Yunnan	33,600	29,721	24,238	14,298	–
Tajik	Xinjiang	41,028	33,223	26,600	16,236	14,462
Nu	Yunnan	28,759	27,190	22,896	15,047	–
Uzbek	Xinjiang	12,370	14,763	12,213	7,717	13,626
Russian	Xinjiang, Heilongjiang	15,609	13,500	2,917	1,326	22,656
Ewenki	Inner Mongolia	30,505	26,379	19,398	9,681	4,957
Deang	Yunnan	17,935	15,461	12,297	7,261	–
Baoan/Bonan	Gansu	16,505	11,683	9,017	5,125	4,957
Yugur	Gansu	13,719	12,293	10,568	5,717	3,861
Jing	Guangxi	22,517	18,749	13,108	4,293	–
Tatar	Xinjiang	4,890	5,064	4,122	2,294	6,929
Derung	Yunnan	7,426	5,825	4,633	3,090	–
Oroqen	Heilongjiang, Inner Mongolia	8,196	7,004	4,103	2,709	2,262
Hezhen	Heilongjiang	4,640	2,115	1,489	718	–
Monba	Tibet	8,923	7,498	1,140	3,809	–
Lhoba	Tibet, Yunnan	2,965	2,322	1,066	–	–
Jino	Yunnan	20,899	18,022	11,962	–	–
Total		**104,490,735**	**90,566,506**	**66,434,341**	**39,873,909**	**34,013,782**

Source: Ethnic Statistical Yearbook 2007, population figures are from the 2000 Population census, population figures of 1990, 1982, 1964, 1953 are taken from ZHOU (2003, p. 12-13)

Table A2: Ethnic Autonomous Areas

Name of Ethnic Autonomous Area	Time of Founding	Capital	Area (square km)	Population (thousand)	Proportion of Ethnic Minority Population (%)
Five Autonomous Regions					
Inner Mongolia Autonomous Region	May 1, 1947	Hohhot City	1,197,547	23,796.1	21.25
Guangxi Zhuang Autonomous Region	March 15, 1958	Nanning City	237,693	48,570.0	38.17
Tibet Autonomous Region	Sept. 1, 1965	Lhasa City	1,247,910	2,592.1	95.93
Ningxia Hui Autonomous Region	Oct. 25, 1958	Yinchuan City	62,818	5,801.9	35.52
Xinjiang Uygur Autonomous Region	Oct. 1, 1955	Urumqi City	1,655,826	19,339.5	60.13
30 Autonomous Prefectures					
Jilin Province					
Yanbian Korean Autonomous Prefecture	Sept. 3, 1952	Yanji City	42,700	2,185.7	40.89
Hubei Province					
Enshi Tujia-Miao Autonomous Prefecture	Dec. 1, 1983	Enshi City	23,942	3,817.9	52.80
Hunan Province					
Xiangxi Tujia-Miao Autonomous Prefecture	Sept. 20, 1957	Jishou City	15,461	2,655.5	74.59
Sichuan Province					
Aba Tibetan-Qiang Autonomous Prefecture	Jan. 1, 1953	Maerkang County	84,242	847.1	73.35
Liangshan Yi Autonomous Prefecture	Oct. 1, 1952	Xichang City	60,423	4,154.8	47.34
Garze Tibetan Autonomous Prefecture	Nov. 24, 1950	Kangding County	152,629	904.9	81.73
Guizhou Province					
Qiandongnan Miao-Dong Autonomous Prefecture	July 23, 1956	Kaili City	30,337	4,193.8	77.10
Qiannan Bouyei-Miao Autonomous Prefecture	Aug. 8, 1956	Duyun City	26,193	3,790.1	55.28
Qianxinan Bouyei-Miao Autonomous Prefecture	May 1, 1982	Xingyi City	16,804	3,016.2	42.94
Yunnan Province					
Xishuangbanna Dai Autonomous Prefecture	Jan. 24, 1953	Jinghong County	19,700	869.2	74.83
Wenshan Zhuang-Miao Autonomous Prefecture	April 1, 1958	Wenshan County	32,239	3,322.7	56.64
Honghe Hani-Yi Autonomous Prefecture	Nov. 18, 1957	Gejiu City	32,931	4,014.5	56.26
Dehong Dai-Jingpo Autonomous Prefecture	July 24, 1953	Luxi County	11,526	1,048.0	51.61
Nujiang Lisu Autonomous Prefecture	Aug. 23, 1954	Lushui County	14,703	471.5	92.17

Deqen Tibetan Autonomous Prefecture	Sept. 13, 1957	Zhongdian County	23,870	353.8	86.54
Dali Bai Autonomous Prefecture	Nov. 22, 1956	Dali City	29,459	3,358.3	49.49
Chuxiong Yi Autonomous Prefecture	April 15, 1958	Chuxiong City	29,258	2,550.3	31.70
Gansu Province					
Linxia Hui Autonomous Prefecture	Nov. 19, 1956	Linxia City	8,417	1,913.7	56.88
Gannan Tibetan Autonomous Prefecture	Oct. 1, 1953	Xiahe County	40,201	682.9	57.16
Qinghai Province					
Haibei Tibetan Autonomous Prefecture	Dec. 31, 1953	Haiyan County	39,354	267.5	61.34
Huangnan Tibetan Autonomous Prefecture	Dec. 22, 1953	Tongren County	17,921	212.6	93.40
Hainan Tibetan Autonomous Prefecture	Dec. 6, 1953	Gonghe County	45,895	394.1	68.33
Golog Tibetan Autonomous Prefecture	Jan. 1, 1954	Maqen County	76,312	138.6	92.65
Yushu Tibetan Autonomous Prefecture	Dec. 25, 1951	Yushu County	188,794	274.8	95.79
Haixi Mongolian-Tibetan Autonomous Prefecture	Jan. 25, 1954	Delingha City	325,785	340.2	26.31
Xinjiang Uygur Autonomous Region					
Changji Hui Autonomous Prefecture	July 15, 1954	Changji City	77,582	1,543.3	43.36
Bayingolin Mongolian Autonomous Prefecture	June 23, 1954	Korla City	471,526	1,126.5	42.24
Kizilsu Kirgiz Autonomous Prefecture	July 14, 1954	Artux City	69,815	458.4	16.87
Bortala Mongolian Autonomous Prefecture	July 13, 1954	Bole City	24,900	438.6	32.28
Ili Kazak Autonomous Prefecture	Nov. 27, 1954	Yining City	269,168	4,083.3	54.83
120 Autonomous Counties					
Hebei Province					
Dachang Hui Autonomous County	Dec. 7, 1955	Dachang Town	176	112.0	24.27
Mengcun Hui Autonomous County	Nov. 30, 1955	Mengcun Town	393	187.2	23.98
Qinglong Manchu Autonomous County	May 10, 1987	Qinglong Town	3,309	516.5	68.40
Fengning Manchu Autonomous County	May 15, 1987	Dage Town	8,747	380.5	68.05
Weichang Manchu-Mongolian Autonomous County	June 12, 1990	Weichang Town	9,058	515.7	57.84
Kuancheng Manchu Autonomous County	June 16, 1990	Kuancheng Town	1,933	233.9	63.50

Inner Mongolia Autonomous Region					
Oroqen Autonomous Banner	Oct. 1, 1951	Alihe Town	13,800	283.1	11.52
Morin Dawa Daur Autonomous Banner	Aug. 15, 1958	Ni´erji Town	2,351	314.6	19.70
Ewenki Autonomous Banner	Aug. 1, 1958	Bayantuohai Town	16,800	144.3	39.56
Liaoning Province					
Fuxin Mongolian Autonomous County	April 7, 1958	Fuxin Town	6,246	731.5	20.30
Mongolian Autonomous County of Harqin Left Wing	April 1, 1958	Dachengzi Town	2,240	423.6	19.71
Xiuyan Manchu Autonomous County	June 11, 1985	Xiuyan Town	4,502	503.1	79.95
Xinbin Manchu Autonomous County	June 7, 1985	Xinbin Town	4,287	306.6	73.50
Qingyuan Manchu Autonomous County	June 6, 1990	Qingyuan Town	3,921	341.6	61.00
Benxi Manchu Autonomous County	June 8, 1990	Xiaoshi Town	3,362	299.9	63.85
Juanren Manchu Autonomous County	June 10, 1990	Juanren Town	3,547	302.9	59.00
Kuandian Manchu Autonomous County	June 12, 1990	Kuandian Town	6,186	436.4	54.92
Jilin Province					
Changbai Korean Autonomous County	Sept. 15, 1958	Changbai Town	2,496	85.1	15.86
Mongolian Autonomous County of Qian Gorlos	Sept. 1, 1956	Qianguo Town	5,117	575.4	9.99
Yitong Manchu Autonomous County	Aug. 30, 1989	Yitong Town	2,523	466.4	39.71
Heilongjiang Province					
Mongolian Autonomous County of Dorbod	Dec. 5, 1956	Taikang Town	6,427	248.8	21.00
Zhejiang Province					
Jingning She Autonomous County	Dec. 24, 1984	Hexi Town	1,950	179.3	9.94
Hubei Province					
Changyang Tujia Autonomous County	Dec. 8, 1984	Longzhouping Town	3,430	409.7	50.65
Wufeng Tujia Autonomous County	Dec. 12, 1984	Wufeng Town	2,072	205.9	84.88
Hunan Province					
Chengbu Miao Autonomous County	Nov. 30, 1956	Rulin Town	2,620	257.2	57.59
Tongdao Dong Autonomous County	May 7, 1954	Shuangjiang Town	2,225	221.1	88.50
Jianghua Yao Autonomous County	Nov. 25, 1955	Tuojiang Town	3,216	458.3	63.97
Xinhuang Dong Autonomous County	Dec. 5, 1956	Xinhuang Town	1,511	250.5	87.56

Zhijiang Dong Autonomous County	Sept. 24, 1987	Zhijiang Town	2,096	356.9	61.25
Jingzhou Miao-Dong Autonomous County	Sept. 27, 1987	Quyang Town	2,211	260.1	73.00
Mayang Miao Autonomous County	April 1, 1990	Gaocun Town	1,561	359.8	77.97
Guangdong Province					
Liannan Yao Autonomous County	Jan. 25, 1953	Sanjiang Town	1,231	155.6	51.55
Lianshan Zhuang-Yao Autonomous County	Sept. 26, 1962	Jitian Town	1,264	114.7	62.88
Ruyuan Yao Autonomous County	Oct. 1, 1963	Rucheng Town	2,125	201.3	11.38
Guangxi Zhuang Autonomous Region					
Du'an Yao Autonomous County	Dec. 15, 1955	Anyang Town	4,092	611.2	97.45
Rongshui Miao Autonomous County	Nov. 26, 1952	Rongshui Town	4,665	468.1	71.83
Sanjiang Dong Autonomous County	Dec. 3, 1952	Guyi Town	2,455	347.1	86.40
Longsheng Multi-ethnic Autonomous County	Aug. 19, 1951	Longsheng Town	2,537	165.5	77.40
Jinxiu Yao Autonomous County	May 28, 1952	Jinxiu Town	2,517	148.1	78.40
Longlin Multi-ethnic Autonomous County	Jan. 1, 1953	Xinzhou Town	3,542	357.6	80.18
Bama Yao Autonomous County	Feb. 6, 1956	Bama Town	1,966	240.2	86.64
Luocheng Mulam Autonomous County	Jan. 10, 1984	Dongmen Town	2,639	360.5	73.07
Fuchuan Yao Autonomous County	Jan. 1, 1984	Fuyang Town	1,572	299.8	46.77
Dahua Yao Autonomous County	Dec. 23, 1987	Dahua Town	2,754	402.4	93.96
Huanjiang Maonan Autonomous County	Nov. 24, 1987	Si¡en Town	4,558	366.9	91.67
Gongcheng Yao Autonomous County	Oct. 15, 1990	Gongcheng Town	2,149	279.6	58.98
Hainan Province					
Baisha Li Autonomous County	Dec. 30, 1987	Yacha Town	2,117	181.6	61.36
Changjiang Li Autonomous County	Dec. 30, 1987	Shilu Town	1,569	232.1	36.77
Ledong Li Autonomous County	Dec. 28, 1987	Baoyou Town	2,747	469.4	38.01
Lingshui Li Autonomous County	Dec. 30, 1987	Yelin Town	1,128	330.1	55.98
Qiongzhong Li-Miao Autonomous County	Dec. 28, 1987	Yinggen Town	2,706	203.6	56.56
Baoting Li-Miao Autonomous County	Dec. 30, 1987	Baocheng Town	1,161	105.5	90.57

Chongqing Municipality					
Shizhu Tujia Autonomous County	Nov. 18, 1984	Nanbin Town	3,013	512.3	69.49
Xiushan Tujia-Miao Autonomous County	Nov. 7, 1983	Zhonghe Town	2,450	606.0	52.41
Youyang Tujia-Miao Autonomous County	Nov. 11, 1984	Zhongduo Town	5,173	745.0	83.60
Pengshui Miao-Tujia Autonomous County	Nov. 10, 1984	Hanjia Town	3,903	630.3	59.63
Sichuan Province					
Beichuan Qiang Autonomous County	Oct. 25, 2003	Qushan Town	2,86 5	161.2	58.89
Muli Tibetan Autonomous County	Feb. 19, 1953	Qiaowa Town	13,252	126.3	78.42
Mabian Yi Autonomous County	Oct. 9, 1984	Minjian Town	2,383	180.2	40.57
Ebian Yi Autonomous County	Oct. 5, 1984	Shaping Town	2,395	148.9	31.24
Guizhou Province					
Songtao Miao Autonomous County	Dec. 31, 1956	Liaogao Town	2,861	639.9	42.49
Zhenning Bouyei-Miao Autonomous County	Sept. 11, 1963	Chengguan Town	1,721	334.6	58.61
Ziyun Miao-Bouyei Autonomous County	Feb. 11, 1966	Songshan Town	2,284	322.4	68.44
Weining Yi-Hui-Miao Autonomous County	Nov. 11, 1954	Caohai Town	6,296	1,095.9	25.37
Guanling Bouyei-Miao Autonomous County	Dec. 31, 1981	Guansuo Town	1,468	320.0	58.99
Sandu Shui Autonomous County	Jan. 2, 1957	Sanhe Town	2,383	314.7	96.85
Yuping Dong Autonomous County	Nov. 7, 1984	Pingxi Town	516	136.8	82.70
Daozhen Gelao-Miao Autonomous County	Nov. 29, 1987	Yuxi Town	2,156	336.6	79.18
Wuchuan Gelao-Miao Autonomous County	Nov. 26, 1987	Duru Town	2,773	419.3	96.25
Yinjiang Tujia-Miao Autonomous County	Nov. 20, 1987	Yinjiang Town	1,961	399.4	71.36
Yanhe Tujia Autonomous County	Nov. 23, 1987	Heping Town	2,469	558.7	55.74
Yunnan Province					
Eshan Yi Autonomous County	May 12, 1951	Shuangjiang Town	1,972	149.0	65.46
Shilin Yi Autonomous County	Dec. 31, 1956	Lufu Town	1,777	229.3	34.29
Cangyuan Va Autonomous County	Feb. 28, 1964	Mengdong Town	2,539	166.9	90.90
Gengma Dai-Va Autonomous County	Oct. 16, 1955	Gengma Town	3,837	255.0	51.60
Yulong Naxi Autonomous County	Dec. 26, 2002	Huangshan Town	6,521	209.7	85.03

Ninglang Yi Autonomous County	Sept. 20, 1956	Daxing Town	6,206	235.4	79.39
Jiangcheng Hani-Yi Autonomous County	May 18, 1954	Menglie Town	3,476	109.4	81.20
Lancang Lahu Autonomous County	April 7, 1953	Menglang Town	8,807	470.6	77.12
Menglian Dai-Lahu-Va Autonomous County	June 16, 1954	Nayun Town	1,957	114.2	85.84
Ximeng Va Autonomous County	March 5, 1965	Mengsuo Town	1,391	82.9	94.06
Hekou Yao Autonomous County	July 11, 1963	Hekou Town	1,313	78.2	63.49
Pingbian Miao Autonomous County	July 1, 1963	Yuping Town	1,905	146.2	61.95
Gongshan Drung-Nu Autonomous County	Oct. 1, 1956	Cikai Town	4,506	30.42	96.24
Weishan Yi-Hui Autonomous County	Nov. 9, 1956	Wenhua Town	2,266	301.7	43.20
Nanjian Yi Autonomous County	Nov. 27, 1965	Nanjian Town	1,802	215.4	49.32
Xundian Hui-Yi Autonomous County	Dec. 20, 1979	Rende Town	3,966	503.9	21.82
Yuanjiang Hani-Yi-Dai Autonomous County	Nov. 22, 1980	Lijiang Town	2,858	196.5	79.24
Xinping Yi-Dai Autonomous County	Nov. 25, 1980	Guishan Town	4,223	269.7	69.76
Mojiang Hani Autonomous County	Nov. 28, 1979	Lianzhu Town	5,459	351.0	73.97
Shuangjiang Lahu-Va-Blang-Dai Autonomous County	Dec. 30, 1985	Mengmeng Town	2,292	163.8	44.36
Lanping Bai-Pumi Autonomous County	May 25, 1988	Jinding Town	4,455	190.5	93.47
Weixi Lisu Autonomous County	Oct. 13, 1985	Baohe Town	4,661	144.6	83.28
Jingdong Yi Autonomous County	Dec. 20, 1985	Jinping Town	4,532	353.0	46.03
Jinggu Dai-Yi Autonomous County	Dec. 25, 1985	Weiyuan Town	7,777	291.7	46.44
Pu'er Hani-Yi Autonomous County	Dec. 15, 1985	Ning¡¯er Town	3,670	185.1	49.54
Yangbi Yi Autonomous County	Nov. 1, 1985	Shangjie Town	1,957	100.1	63.28
Luquan Yi-Miao Autonomous County	Nov. 25, 1985	Pingshan Town	4,378	448.8	30.42
Jinping Miao-Yao-Dai Autonomous County	Dec. 7, 1985	Jinhe Town	3,677	316.3	85.51
Zhenyuan Yi-Hani-Lahu Autonomous County	May 15, 1990	Enle Town	4,223	204.3	51.80
Gansu Province					
Zhangjiachuan Hui Autonomous County	July 6, 1953	Zhangjiachuan Town	1,311	315.0	69.75
Tianzhu Tibetan Autonomous County	May 6, 1950	Huazangsi Town	6,865	214.1	38.00

Sunan Yugur Autonomous County	Feb. 20, 1954	Hongwansi Town	20,456	35.3	55.30
Subei Mongolian Autonomous County	July 29, 1950	Dangchengwan Town	55,000	11.1	41.08
Aksay Kazak Autonomous County	April 27, 1954	Hongliuwan Town	31,374	8.0	32.60
Dongxiang Autonomous County	Sept. 25, 1950	Suonan Town	1,467	267.3	88.11
Jishishan Bao'an-Dongxiang-Salar Autonomous County	Sept. 30, 1981	Chuimatan Town	910	219.8	54.45
Qinghai Province					
Huzhu Tu Autonomous County	Feb. 17, 1954	Weiyuan Town	3,321	370.8	25.13
Hualong Hui Autonomous County	March 1, 1954	Bayan Town	2,740	232.5	78.57
Xunhua Salar Autonomous County	March 1, 1954	Jishi Town	1,749	116.3	94.05
Henan Mongolian Autonomous County	Oct. 16, 1954	Youganning Town	6,250	31.9	97.16
Menyuan Hui Autonomous County	Dec. 19, 1953	Haomen Town	6,896	150.6	56.93
Datong Hui-Tu Autonomous County	July 10, 1986	Qiaotou Town	3,090	425.5	46.50
Minhe Hui-Tu Autonomous County	June 27, 1986	Chuankou Town	1,780	376.7	54.84
Xinjiang Uygur Autonomous Region					
Barkol Kazak Autonomous County	Sept. 30, 1954	Barkol Town	36,947	101.0	33.81
Taxkorgan Tajik Autonomous County	Sept. 17, 1954	Taxkorgan Town	52,300	33.2	52.70
Mulei Kazak Autonomous County	July 17, 1954	Mulei Town	13,510	85.5	31.44
Yanqi Hui Autonomous County	March 15, 1954	Yanqi Town	2,429	124.7	54.77
Qapqal Xibe Autonomous County	March 25, 1954	Qapqal Town	4,482	164.8	63.60
Hoboksar Mongolian Autonomous County	Sept. 10, 1954	Hoboksar Town	28,799	49.5	65.80

Source: CHINA.ORG.CN (2005).

Stata do-files

Merging of CHNS databases:

```
use m07jobs;
keep hhid line wave b4 b5 b6 b7 b9a b9 b8 b2a commid t1 t2 t3 t4 t5;
sort hhid line wave;
save m07jobs_merge, replace;

use panel_ethnicity, clear;
keep hhid line wave gender a11 a12 commid t1 t2 t3 t4 t5 west_dob nationality;
sort hhid line wave;
save panel_ethnicity_merge, replace;

use m07rst;
keep hhid line wave a8 a5 a5b a5d a8b1 aa11 aa13 a14 a15 a16 a19 a20 a26 commid t1 t2 t3
t4 t5;
sort hhid line wave;
save m07rst_merge, replace;

use m07wed;
keep hhid line wave s215 s216 s217 s218 commid t1 t2 t3 t4 t5;
sort hhid line wave;
save m07wed_merge, replace;

use m07timea;
keep hhid line wave k12 commid t1 t2 t3 t4 t5;
sort hhid line wave;
save m07timea_merge, replace;

use m07emw;
keep hhid line wave s41 s42 s44 s45 commid t1 t2 t3 t4 t5;
sort hhid line wave;
save m07emw_merge, replace;

use m07pe;
keep hhid line wave u48a u129_mn u179 u200 gender age commid t1 t2 t3 t4 t5;
```

```
sort hhid line wave;
save m07pe_merge, replace;

use m07wages;
keep if job==1;
keep hhid line wave c6 c8 commid t1 t2 t3 t4 t5;
sort hhid line wave;
save m07wages_merge, replace;

use m07wages;
keep if job==2;
rename c6 c6a;
rename c8 c8a;
keep hhid line wave c6a c8a commid t1 t2 t3 t4 t5;
sort hhid line wave;
save m07wages_soc_merge, replace;

use m07farmg;
format hhid %12.0g;
egen id=group(hhid line);
bysort id: gen newid = 1 if _n==1;
drop if newid ==.;
keep hhid line wave source e2a e4c;
sort hhid line wave;
save m07farmg_merge, replace;

use m07livei;
format hhid %12.0g;
egen id=group(hhid line);
bysort id: gen newid = 1 if _n==1;
drop if newid ==.;
keep hhid line wave f4c f5;
sort hhid line wave;
save m07livei_merge, replace;
```

```
use m07fishi;
format hhid %12.0g;
egen id=group(hhid line);
bysort id: gen newid = 1 if _n==1;
drop if newid ==.;
keep hhid line wave g4c g5;
sort hhid line wave;
save m07fishi_merge, replace;

use c07indinc;
format hhid %12.0g;
egen id=group(hhid line);
bysort wave id: egen count= count(indinc);
drop if count==2;
keep hhid line wave indinc commid t1 urban;
sort hhid line wave;
save c07indinc_merge, replace;

merge hhid line wave using m07jobs_merge panel_ethnicity_merge m07rst_merge
m07wed_merge m07timea_merge m07emw_merge m07pe_merge m07wages_merge
m07wages_soc_merge m07farmg_merge m07livei_merge m07fishi_merge c07indinc_merge;
save merge_final, replace;
```

Database used in the study

```
set memory 300m;
set more off;
use merge_final;
*YEAR***;
***for the 1997 database use wave==1997, for the 2000 database use wave==2000, for the
2004 database use wave==2004;
*keep if wave==1997;
*keep if wave==2000;
*keep if wave==2004;
*GUIZHOU PROVINCE*** t1==52 ***;
rename t1 prov;
keep if prov==52;
```

```
*RURAL AREAS t2==2***;
keep if t2==2;
*ETHNICITY*** 1)Han 6)Miao 9)Bouyei 15)Tujia;
***dropped 20)Other, 2)Mongolian, 7)Yi***;
***new categorization 0)Han 1)Miao 2)Bouyei 3)Tujia***;
drop if nationality==.;
tab nationality;
rename nationality eth;
drop if eth==2 | eth==7 | eth==20;
replace eth = 0 if eth==1;
replace eth = 1 if eth==6;
replace eth = 2 if eth==9;
replace eth = 3 if eth==15;
tab(eth), gen(eth);
rename eth1 han;
rename eth2 miao;
rename eth3 bouyei;
rename eth4 tujia;
*EMPLOYMENT SITUATION***;
***b4 primary occupation***;
****1)senior professional/technical worker(doctor,professor,lawyer,architect,engineer)***;
****2)junior professional/technical worker (midwife,nurse,teacher,editor,photographer)***;
****3)administrator/executive/manager (working proprietor,government official, section
chief, department or bureau director,administrative cadre,village leader);
****4)office staff (secretary,office helper);
****5)farmer,fisherman,hunter
****6)skilled worker(foreman,group leader,craftsman);
****7)unskilled worker(ordinary laborer,logger);
****8)army officer, police officer;
****9)ordinary soldier, policeman;
***10)driver;
***11)service worker(housekeeper,cook,waiter,doorman,hairdresser,
salesperson,launderer,child care worker);
***12)athlete,actor,musician (no observation);
***13)other;
***-9)unknown (dropped);
```

```
tab b4;
drop if b4==-9;
drop if b4==13;
drop if b9a==-9;
drop if b9==-9;
drop if b2a==-9;
***own categories;
****1)Agriculture/Non-Agriculture;
generate agr=b4;
replace agr=0 if agr==1 | agr==2 | agr==3 |agr==4 |agr==6 | agr==7 |agr==8 | agr==9 |
agr==10 | agr==11;
replace agr=1 if agr==5;
****group occupations 1) Agriculture 2) BlueCollar 3) WhiteCollar***;
***0) 5;
***1) 7,9,10,11;
***2) 1,2,3,4,6,8;
generate outcome=b4;
replace outcome=0 if b4==5;
replace outcome=100 if b4==7 | b4==9 | b4==10 | b4==11;
replace outcome=200 if b4==1 | b4==2 | b4==3 | b4==4 | b4==6 | b4==8;
replace outcome=1 if outcome==100;
replace outcome=2 if outcome==200;
tab outcome wave,row;
*drop if outcome==.;
***employment position in this occupation***b5;
***type of work unit b6a***;
***working - not working b2, b2a***;
***why are you not working***
***1)seeking work, 2)doing housework, 3)disabled, 4)student, 5)retired, 6)other, 9) un-
known***;
***secondary occupation b9a****
*OTHER INDEPENDENT VARIABLES***;
**EDUCATION YEARS***;
tab(a11), gen(educa);
```

```
gen yearseduc=0*educa1+1*educa2+2*educa3+3*educa4+4*educa5+5*educa6+6*educa7+
7*educa8+8*educa9+9*educa10+10*educa11+11*educa12+12*educa13+13*educa14+14*ed
uca15+15*educa16+16*educa17+17*educa18+18*educa19+19*educa20;

drop if yearseduc==.;

**PARENTAL EDUCATION/OCCUPATION***(given a5b==father´s line, a5d mother´s
line)***;

egen fatherid=group(hhid a5b);

egen motherid=group(hhid a5d);

gen father_edu = yearsedu if fatherid !=.;

gen mother_edu = yearsedu if motherid !=.;

gen father_occ = outcome if fatherid !=.;

gen mother_occ = outcome if motherid !=.;

**AGE***;

gen millier=west_dob/10000;

gen yearofbirthbis=floor(millier);

rename age agebis;

gen ageter=wave-yearofbirthbis;

replace ageter=agebis if ageter==.;

gen age=floor(ageter);

replace yearofbirthbis=wave-age if yearofbirthbis==.;

gen yearofbirth=floor(yearofbirthbis);

drop if age==.;

**GENDER*male=1 and female=2***;

rename gender male;

replace male = 0 if male==2;

**MARITAL STATUS*****1)never married, 2)married, 3)divorced, 4)widowed,
5)separated***;

rename a8 ms;

**DUMMY MARRIED/UNMARRIED***;

replace ms = 0 if ms==1 | ms==3 | ms==4 | ms==5;

replace ms = 1 if ms==2;

*drop if ms==.;

***observed county (1-4)***t3;

keep hhid line commid outcome agr eth yearseduc male age ms b2a b5 b6 b9a b9 u129_mn;

sort commid;

******for the 1997 database use save full 1997, for the 2000 database use save full 2000, for
the 2004 database use save full 2004;
```

```
save full1997, replace;
***MERGING WITH COMMUNITY DATA***;
***for the 1997 database use wave==1997, for the 2000 database use wave==2000, for the 2004 database use wave==2004;
use m05comfm.dta;
keep if wave==1997;
rename t1 prov;
keep if prov==52;
keep if t2==2;
keep commid wave prov t2 r6b r6i r6 r16 r6j_1 r6j_6 r6j_9 r6j_15;
sort commid;
save m05comfm_merge, replace;
use m05comin;
keep if wave==1997;
rename t1 prov;
keep if prov==52;
keep if t2==2;
keep commid wave prov t2 o1c o9k o9m_1 o9m_2 o9m_3 o9m_4 o9n o9o o9p o40 o79 o81 o83 o85 o23 o33 o34 o35 o36 t3 o0a o271;
sort commid;
save mo5comin_merge, replace;
******save and merge the files according to the years considered;
merge commid using m05comfm_merge mo5comin_merge full1997;
save analysis_full_1997, replace;
```

Data analysis

```
set memory 300m;
set more off;
******use the database according to the years considered;
use analysis_full_1997.dta;
drop b2a;
***GUIZHOU PROVINCE** t1==52 ***;
*RURAL AREAS t2==2***;
*ETHNICITY**;
drop if eth==.;
tab(eth), gen(eth);
```

```
rename eth1 han;
rename eth2 miao;
rename eth3 bouyei;
rename eth4 tujia;
*OCCUPATIONAL OUTCOME***;
drop if outcome==.;
*tab (outcome), gen (outcome);
*OCCUPATIONAL OUTCOME***;
***secondary occupation***b9;
gen soc = b9;
gen agr_soc = soc if outcome==0;
replace agr_soc = 100 if agr_soc==2 | agr_soc==3 | agr_soc==4 | agr_soc==6 | agr_soc==7 |
agr_soc==10 | agr_soc==11 | agr_soc==12 | agr_soc==13;
replace outcome = 11 if agr_soc==100;
replace outcome = 100 if outcome==1;
replace outcome = 200 if outcome==2;
replace outcome =1 if outcome==11;
replace outcome =2 if outcome==100;
replace outcome =3 if outcome==200;
***categories 0) agriculture 1)agriculture + secondary occupation 2) blue collar 3) white
collar***;
*replace outcome =2 if outcome==3;
*****Agriculture vs. Nonagriculture***;
gen nonagr = outcome;
replace nonagr = 0 if outcome==1;
replace nonagr = 1 if outcome==2 | outcome==3;
***categories 0) agriculture, agriculture + secondary occupation 1) non-agriculture (BC
+WC***);
***Only Agriculture***comparison (0)agriculture and (1)agriculture+soc;
gen agrsoc = outcome;
replace agrsoc = . if outcome==2 | outcome==3;
***Only Non-Agriculture***comparison (0) BC and (1) WC;
gen bcwc = outcome;
replace bcwc = . if outcome==0 | outcome==1;
replace bcwc =0 if outcome==2;
replace bcwc =1 if outcome==3;
```

```
***primary occupation employment position b5***;

***primary occupation work unit b6***;

*Is the family planning policy the same for minorities as it is for Han nationality r6b***;

gen familyplan = r6b;

***Are minority couples in this village/neighborhood allowed to have two children r6k***

***0 no, 1 yes, but only if both the husband and wife are minorities, 2 yes, as long as either the husband or wife is a minority***;

***Are all couples in this village/neighborhood allowed to have more than two children? r6i***

***all not allowed see tabulation results***;

***Are all couples in this village/neighborhood allowed to have two children?***;

**not allowed**;

*Do couples receive a subsidy if they have only one child r16***;

gen childsubsidy = r16;

**Is there an open trade area, an open city, or a special economic zone near this village/neighborhood (within two hours by bus) O40***;

gen economiczone = o40;

***primary school in village o79***;

gen primary = o79;

***lower middle school in village o81***;

gen lowmiddle = o81;

***upper middle school in village o83***;

gen uppmiddle = o83;

***Vocational upper middle school or vocational technical school o85***;

gen vocational = o85;

***What is the most common characteristic of the roads in or around this village/neighborhood O23***;

***1 dirt (no observation), 2 stone, gravel, or mixed material, 3 paved road***;

*** dummy for paved road or not, paved road is 1, no paved road is 0***;

gen pavedroad = o23;

replace pavedroad=1 if pavedroad==3;

replace pavedroad=0 if pavedroad==2;

***Is there a bus stop (or long distance bus stop) in this village/neighborhood o33***;

gen busstop = o33;

***Is this village/neighborhood near a train station o35***;

gen trainstation = o35;
```

```
***how far away is the nearest busstop? (km) o34***;
gen buskm =o34;
***how far away is the nearest trainstation (km)? o36***;
gen trainkm=o36;
****counties t3, 1 first county 2 second county, 3 third county, 4 fourth county***;
***in every province four counties are randomly selected***;
gen county = t3;
tab(county), gen(county);
***distance to school (travel to school in minutes)u129_mn***;
gen distanceschool = u129_mn;
drop if distanceschool == 1200;
****households in the village/neighborhood* o0a***;
gen households = o0a;
***Interaction terms***;
gen miaoedu=miao*yearseduc;
gen bouyeiedu=bouyei*yearseduc;
gen tujiaedu=tujia*yearseduc;
gen agesq = age*age;
gen lnedu = ln(yearseduc);
****only mixed communities***;
drop if commid==522101;
drop if commid==522401;
drop if commid==522402;
drop if commid==522403;
drop if commid==522404;
gen abcwc = outcome;
replace abcwc=. if outcome==1;
replace abcwc=1 if abcwc==3 | abcwc==2;
gen abc = outcome;
replace abc=. if outcome==1 | outcome==3;
replace abc=1 if abc==2;
sum bouyei miao tujia yearseduc male age households county;
end;
```

LR-test (LR-tests were accomplished by including/excluding variables into/from the regression until the best model fit was achieved);

```
mlogit outcome bouyei miao tujia yearseduc male age households i.county2 i.county3
estimates store full
mlogit outcome bouyei miao tujia yearseduc male age households i.county2
estimates store restricted1
lrtest full restricted1, stats
```

Estimation;

NONAGR VS AG;

```
reg nonagr i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)
estimates store reg
outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)
logit nonagr i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)
estimates store reg
outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)
margins, dydx(*)
```

Stata’s "margins" command is constantly updated; please refer to Karaca-Mandic et al. (2012) and Williams (2012) for current information.

****agricultural sector A+soc vs A***;

```
reg agrsoc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)
estimates store reg
outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)
logit agrsoc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)
estimates store reg
outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)
margins, dydx(*)
```

****A vs BC***;

```
reg abc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)
estimates store reg
```

```
outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997",
word replace bdec(3)

logit abc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce
(cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997",
word replace bdec(3)

margins, dydx(*)

****A vs BC+WC***;

reg abcwc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce
(cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997",
word replace bdec(3)

logit abcwc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce
(cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997",
word replace bdec(3)

margins, dydx(*)

****Full outcome***;

****Combination of outcomes, book Long and Freese p. 204***;

mlogit outcome bouyei miao tujia yearseduc male age households county2 county3

mlogtest, combine

****combine BC and WC into NA***;

mlogit outcome bouyei miao tujia yearseduc male age households county2 county3

mlogtest, combine

mlogtest, wald

****FULL OUTCOME A, A+SOC, NA***;

mlogit outcome i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3,
vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997",
word replace bdec(3)

ologit outcome i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3,
vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997",
word replace bdec(3)
```

```
reg outcome i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce
(cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997",
word replace bdec(3)

***MARGINAL EFFECTS***;

margins, dydx(*) predict(outcome(0))

margins, dydx(*) predict(outcome(1))

margins, dydx(*) predict(outcome(2))


*SUEST TEST***;

mlogit outcome bouyei miao tujia yearseduc male age households county2 county3

estimates store m0

quietly mlogit outcome bouyei miao tujia yearseduc male age households county2 county3 if
outcome !=1

estimates store m1

suest m0 m1, vce(cluster commid)

test [m0_2 = m1_2], cons

test [m0_3 = m1_3], cons


mlogit outcome bouyei miao tujia yearseduc male age households county2 county3

estimates store m0

quietly mlogit outcome bouyei miao tujia yearseduc male age households county2 county3 if
outcome !=2

estimates store m2

suest m0 m2, vce(cluster commid)

test [m0_1 = m2_1], cons

test [m0_3 = m2_3], cons


mlogit outcome bouyei miao tujia yearseduc male age households county2 county3

estimates store m0

quietly mlogit outcome bouyei miao tujia yearseduc male age households county2 county3 if
outcome !=3

estimates store m3

suest m0 m3, vce(cluster commid)

test [m0_1 = m3_1], cons

test [m0_2 = m3_2], cons
```

```
estimates table m0 m1 m2 m3, star stats(N ll)
***MNL vs Ologit**;
ologit outcome i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3
predict Aologit Asocologit BCologit WCologit
label var Aologit "ologit-A"
label var Asocologit "ologit-A+soc"
label var BCologit "ologit-BC"
label var WCologit "ologit-WC"
mlogit outcome i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3
predict Amlogit Asocmlogit BCmlogit WCmlogit
label var Amlogit "mlogit-A"
label var Asocmlogit "mlogit-A+soc"
label var BCmlogit "mlogit-BC"
label var WCmlogit "mlogit-WC"

dotplot Aologit Amlogit
dotplot Asocologit Asocmlogit
dotplot BCologit BCmlogit
dotplot WCologit WCmlogit

corr Aologit Amlogit
corr Asocologit Asocmlogit
corr BCologit BCmlogit
corr WCologit WCmlogit

***NONLINEARITIES***;
****LR-tests, BIC and AIC criteria were used to determine which interaction terms to include into the model***;
***Example of LR-test
*Full model***;
logit abcwc i.bouyei i.miao i.tujia tujiaedu yearseduc i.male age households i.county2 i.county3
estimates store full
*Restricted model***;
logit abcwc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3
estimates store restricted1
```

```
*LR-test;

lrtest full restricted1, stats

*Nonlinearties 1997***;

logit agrsoc i.bouyei i.miao i.tujia tujiaedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)


reg agrsoc i.bouyei i.miao i.tujia tujiaedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)


logit abc i.bouyei i.miao i.tujia bouyeiedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)


reg abc i.bouyei i.miao i.tujia bouyeiedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS1997", word replace bdec(3)


*Nonlinearties 2004***;

reg nonagr i.bouyei i.miao i.tujia miaoedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)


logit agrsoc i.bouyei i.miao i.tujia bouyeiedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)
```

```
logit agrsoc i.bouyei i.miao i.tujia bouyeiedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)


reg agrsoc i.bouyei i.miao i.tujia bouyeiedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)


logit abc i.bouyei i.miao i.tujia miaoedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)


reg abc i.bouyei i.miao i.tujia miaoedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)


logit abcwc i.bouyei i.miao i.tujia miaoedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)


reg abcwc i.bouyei i.miao i.tujia miaoedu yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)


****PREDICTIONS AFTER LOGIT/LOGISTIC***;
```

```
logit nonagr i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

outreg2 using "F:\RESEARCH\DISS\CHAPTER6\ANALYSIS_DISS\ANALYSIS2004", word replace bdec(3)

predict p

bysort yearseduc: egen Prob_NA_All=mean(p)

bysort yearseduc: egen Prob_NA_Han=mean(p) if han==1

bysort yearseduc: egen Prob_NA_Miao=mean(p) if miao==1

bysort yearseduc: egen Prob_NA_Bouyei=mean(p) if bouyei==1

bysort yearseduc: egen Prob_NA_Tujia=mean(p) if tujia==1

line Prob_NA_All Prob_NA_Han Prob_NA_Miao Prob_NA_Bouyei Prob_NA_Tujia yearseduc, subtitle(, pos(12) ring(6) size(medium) nobox) legend(on) scheme(s1color) clcolor(black gs1 gs5 gs10 gs8) clpattern(solid shortdash longdash_shortdash dash_dot longdash) lw(medium medium) xtitle("Education in years") plotr(lc(white)) ylabel(#4,angle(0) labsize(small)) ytick(#4,angle(0)) ytitle("Prob Non-Agriculture")


logit agrsoc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

predict p

bysort yearseduc: egen Prob_A_soc_All=mean(p)

bysort yearseduc: egen Prob_A_soc_Han=mean(p) if han==1

bysort yearseduc: egen Prob_A_soc_Miao=mean(p) if miao==1

bysort yearseduc: egen Prob_A_soc_Bouyei=mean(p) if bouyei==1

bysort yearseduc: egen Prob_A_soc_Tujia=mean(p) if tujia==1

line Prob_A_soc_All Prob_A_soc_Han Prob_A_soc_Miao Prob_A_soc_Bouyei Prob_A_soc_Tujia yearseduc, subtitle(, pos(12) ring(6) size(medium) nobox) legend(on) scheme(s1color) clcolor(black gs1 gs5 gs10 gs8) clpattern(solid shortdash longdash_shortdash dash_dot longdash) lw(medium medium) xtitle("Education in years") plotr(lc(white)) ylabel(#4,angle(0) labsize(small)) ytick(#4,angle(0)) ytitle("Prob Agriculture + Sec Occupation")


logit abc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

predict p

bysort yearseduc: egen Prob_BC_All=mean(p)

bysort yearseduc: egen Prob_BC_Han=mean(p) if han==1
```

```
bysort yearseduc: egen Prob_BC_Miao=mean(p) if miao==1

bysort yearseduc: egen Prob_BC_Bouyei=mean(p) if bouyei==1

bysort yearseduc: egen Prob_BC_Tujia=mean(p) if tujia==1

line Prob_BC_All Prob_BC_Han Prob_BC_Miao Prob_BC_Bouyei Prob_BC_Tujia yearseduc, subtitle(, pos(12) ring(6) size(medium) nobox) legend(on) scheme(s1color) clcolor(black gs1 gs5 gs10 gs8) clpattern(solid shortdash longdash_shortdash dash_dot longdash) lw(medium medium) xtitle("Education in years") plotr(lc(white)) ylabel(#4,angle(0) labsize(small)) ytick(#4,angle(0)) ytitle("Prob Blue Collar")


logit abcwc i.bouyei i.miao i.tujia yearseduc i.male age households i.county2 i.county3, vce (cluster commid)

estimates store reg

predict p

bysort yearseduc: egen Prob_NA_All=mean(p)

bysort yearseduc: egen Prob_NA_Han=mean(p) if han==1

bysort yearseduc: egen Prob_NA_Miao=mean(p) if miao==1

bysort yearseduc: egen Prob_NA_Bouyei=mean(p) if bouyei==1

bysort yearseduc: egen Prob_NA_Tujia=mean(p) if tujia==1

line Prob_NA_All Prob_NA_Han Prob_NA_Miao Prob_NA_Bouyei Prob_NA_Tujia yearseduc, subtitle(, pos(12) ring(6) size(medium) nobox) legend(on) scheme(s1color) clcolor(black gs1 gs5 gs10 gs8) clpattern(solid shortdash longdash_shortdash dash_dot longdash) lw(medium medium) xtitle("Education in years") plotr(lc(white)) ylabel(#4,angle(0) labsize(small)) ytick(#4,angle(0)) ytitle("Prob Non Agriculture")


***DESCRIPTIVE STATISTICS***;

tab outcome eth, chi2 col

tab outcome eth, chi2 exact

***COMMUNITIES***;

tab commid eth, row

***INDIVIDUAL CHARACTERISTICS***;

sum yearseduc age if eth==0

sum yearseduc age if eth==1

sum yearseduc age if eth==2

sum yearseduc age if eth==3

sum yearseduc age

tab eth male, row

**T-TESTS***Ho= diff=0, values larger than critical values mean that H0 is true, cannot be rejected;
```

```
***COMPARISON HAN MIAO***;
gen han_miao_tt = han
replace han_miao_tt =. if bouyei==1
replace han_miao_tt =. if tujia==1
ttest yearseduc, by(han_miao_tt)
ttest male, by(han_miao_tt)
ttest age, by(han_miao_tt)
***COMPARISON HAN BOUYEI***;
gen han_bouyei_tt = han
replace han_bouyei_tt =. if miao==1
replace han_bouyei_tt =. if tujia==1
ttest yearseduc, by(han_bouyei_tt)
ttest male, by(han_bouyei_tt)
ttest age, by(han_bouyei_tt)
***COMPARISON HAN TUJIA***;
gen han_tujia_tt = han
replace han_tujia_tt =. if miao==1
replace han_tujia_tt =. if bouyei==1
ttest yearseduc, by(han_tujia_tt)
ttest male, by(han_tujia_tt)
ttest age, by(han_tujia_tt)
****Geographic Location**;
tab county commid, row
tab county eth, row
tab eth county, row
tab county nonagr, row
tab county outcome, row
tab outcome eth, row
tab eth outcome, row
```

Studies on the Agricultural and Food Sector in Central and Eastern Europe
edited by Leibniz Institute of Agricultural Development in Central and Eastern Europe (IAMO)
ISSN 1436-221X

Vol. 1 **The importance of institutions for the transition in Central and Eastern Europe with emphasis on agricultural and food industry**
ed. by Klaus Frohberg and Witold-Roger Poganietz
1998, 137 pages, ISBN 3-8175-0258-3

Vol. 2 **The significance of politics and institutions for the design and formation of agricultural Policies**
ed. by Klaus Frohberg and Peter Weingarten
1999, 254 pages, ISBN 3-8175-0289-3

Vol. 3 **Food processing and distribution in transition countries. Problems and perspectives**
ed. by Monika Hartmann and Jürgen Wandel
1999, 349 pages, ISBN 3-8175-0293-1

Vol. 4 **Die private Nachfrage nach Nahrungsmitteln im Transformationsprozeß Tschechiens und Polens**
Stephan Brosig (PhD)
2000, 171 Seiten, ISBN 3-8175-0319-9

Vol. 5 **Integrating Estonia into the EU: Quantitative analysis of the agricultural and food sector**
Achim Fock (PhD)
2000, 286 pages, ISBN 3-8175-0320-2

Vol. 6 **Competitiveness of agricultural enterprises and farm activities in transition countries**
ed. by Peter Tillack and Frauke Pirscher
2000, 216 pages, ISBN 3-8175-0322-9

Vol. 7 **Конкурентоспособность сельскохозяйственных предприятий и фермерской деятельности в странах переходного периода**
под редакцией Петера Тиллака и Фрауке Пиршер
2000, 253 страницы, ISBN 3-8175-0324-5

Vol. 8 **Perspectives on agriculture in transition: Analytical issues, modelling approaches, and case study results**
ed. by Witold-Roger Poganietz, Alberto Zezza, Klaus Frohberg and Kostas G. Stamoulis
2000, 433 pages, ISBN 3-8175-0323-7

Vol. 9 **Land ownership, land markets and their influence on the efficiency of agricultural production in Central and Eastern Europe**
ed. by Peter Tillack and Eberhard Schulze
2000, 485 pages, ISBN 3-8175-0325-3

Vol. 10 **Landwirtschaft und Industrie in Russland – Der Transformationsprozeß in der Ernährungsindustrie**
Jürgen Wandel (PhD)
2000, 361 Seiten, ISBN 3-8175-0334-2

Vol. 11 **Food consumption in Russia. An econometric analysis based on household data**
Karin Elsner (PhD)
2001, 256 pages, ISBN 3-8175-0335-0

Vol. 12 **Alexander Wasiljewitsch Tschajanow – Die Tragödie eines großen Agrarökonomen**
hrsg. u. übers. von Eberhard Schulze
2001, 192 Seiten, ISBN 3-8175-0342-3

Vol. 13 **Analysis of food consumption in Central and Eastern Europe: Relevance and empirical methods**
ed. by Stephan Brosig and Monika Hartmann
2001, 253 pages, ISBN 3-8175-0349-0

Vol. 14 **Wettbewerbsprozesse und Firmenwachstum in der Transformation am Beispiel der polnischen Fleischindustrie**
Agata Pieniadz (PhD)
2002, 291 Seiten, ISBN 3-8175-0360-1

Vol. 15 **Agricultural enterprises in transition: Parallels and divergences in Eastern Germany, Poland and Hungary**
ed. by Ludger Hinners-Tobrägel and Jürgen Heinrich
2002, 455 pages, ISBN 3-8175-0366-0

Vol. 16 **Agricultural technology and economic development of Central and Eastern Europe. Results of the workshop in Halle, 2nd-3rd July 2001**
ed. by Peter Tillack and Ulrich Fiege
2002, 160 pages, ISBN 3-86037-199-1

Vol. 17 **Региональные аспекты аграрных преобразований: Политика, реструктуризация, рыночная адаптация**
под редакцией Петера Тиллака и Виталия Зиновчука
2003, 236 страницы, ISBN 3-928466-55-0

Vol. 18 **Alexander Vasilievich Chayanov – The tragedy of an outstanding agricultural economist**
ed. by Eberhard Schulze
2003, 188 pages, ISBN 3-86037-201-7

Vol. 19 **Development of agricultural market and trade policies in the CEE Candidate Countries**
by the Network of Independent Agricultural Experts in the CEE Candidate Countries
2003, 72 pages, ISBN 3-86037-212-2

Vol. 20 **Large farm management**
ed. by Alfons Balmann and Alexej Lissitsa
2003, 396 pages, ISBN 3-86037-213-0

Vol. 21 **Success and failures of transition – The Russian agriculture between fall and resurrection**
ed. by Eberhard Schulze, Elke Knappe, Eugenia Serova, Peter Wehrheim
2003, 521 pages, ISBN 3-9809270-1-6

Vol. 22 **Subsistence agriculture in Central and Eastern Europe: How to break the vicious circle?**
ed. by Steffen Abele and Klaus Frohberg
2003, 233 pages, ISBN 3-9809270-2-4

Vol. 23 **Pfadabhängigkeiten und Effizienz der Betriebsstrukturen in der ukrainischen Landwirtschaft – Eine theoretische und empirische Analyse**
Andriy Nedoborovskyy (PhD)
2004, 197 Seiten, ISBN 3-86037-216-5

Vol. 24 **Nichtmonetäre Transaktionen in der ukrainischen Landwirtschaft: Determinanten, Spezifika und Folgen**
Olena Dolud (PhD)
2004, 190 Seiten, ISBN 3-9809270-3-2

Vol. 25 **The role of agriculture in Central and Eastern European rural development: Engine of change or social buffer?**
ed. by Martin Petrick and Peter Weingarten
2004, 426 pages, ISBN 3-9809270-4-0

Vol. 26 **Credit rationing of Polish farm households – A theoretical and empirical analysis**
Martin Petrick (PhD)
2004, 254 pages, ISBN 3-9809270-6-7

Vol. 27 **Drei Jahrhunderte Agrarwissenschaft in Russland: Von 1700 bis zur Gegenwart**
Alexander Alexandrowitsch Nikonow und Eberhard Schulze
2004, 232 Seiten, ISBN 3-9809270-8-3

Vol. 28 **Russlands Weg vom Plan zum Markt: Sektorale Trends und regionale Spezifika**
Peter Voigt (PhD)
2004, 270 Seiten, ISBN 3-9809270-9-1

Vol. 29 **Auswirkungen des Transformationsprozesses auf die sozio-ökonomischen Funktionen ukrainischer Landwirtschaftsunternehmen**
Helga Biesold (PhD)
2004 182 Seiten, ISBN 3-938584-00-9

Vol. 30 **Agricultural policies and farm structures – Agent-based modelling and application to EU-policy reform**
Kathrin Happe (PhD)
2004, 291 pages, ISBN 3-938584-01-7

Vol. 31 **How effective is the invisible hand? Agricultural and food markets in Central and Eastern Europe**
ed. by Stephan Brosig and Heinrich Hockmann
2005, 361 pages, ISBN 3-938584-03-3

Vol. 32 **Erfolgsfaktoren von landwirtschaftlichen Unternehmen mit Marktfruchtanbau in Sachsen-Anhalt**
Kirsti Dautzenberg (PhD)
2005, 161 Seiten, ISBN 3-938584-06-8

Vol. 33 **Agriculture in the face of changing markets, institutions and policies: Challenges and strategies**
ed. by Jarmila Curtiss, Alfons Balmann, Kirsti Dautzenberg, Kathrin Happe
2006, 544 pages, ISBN 3-938584-10-6

Vol. 34 **Making rural households' livelihoods more resilient – The importance of social capital and the underlying social networks**
ed. by Gertrud Buchenrieder and Thomas Dufhues
2006, 106 pages, ISBN 3-938584-13-0

Vol. 35 **Außerlandwirtschaftliche Diversifikation im Transformationsprozess. Diversifikationsentscheidungen und -strategien ländlicher Haushalte in Slowenien und Mazedonien**
Judith Möllers (PhD)
2006, 323 Seiten, ISBN 3-938584-14-9

Vol. 36 **Accessing rural finance – The rural financial market in Northern Vietnam**
Thomas Dufhues (PhD)
2007, 166 Seiten, ISBN 3-938584-16-5

Vol. 37 **Страхование посевов в Казахстане: Анализ возможностей эффективного управления рисками**
Раушан Бокушева, Олаф Хайдельбах, Талгат Кусайынов
2007, 82 Seiten, ISBN 3-938584-17-3

Vol. 38 **Rethinking agricultural reform in Ukraine**
Zvi Lerman, David Sedik, Nikolai Pugachov, Aleksandr Goncharuk
2007, 167 Seiten, ISBN 3-938584-18-1

Vol. 39 **Sustainable rural development: What is the role of the agri-food sector?**
ed. by Martin Petrick, Gertrud Buchenrieder
2007, 293 pages, ISBN 3-938584-22-X

Vol. 40 **Efficiency of selected risk management instruments – An empirical analysis of risk reduction in Kazakhstani crop production**
Olaf Heidelbach (PhD)
2007, 223 Seiten, ISBN 3-938584-19-X

Vol. 41 **Marktstruktur und Preisbildung auf dem ukrainischen Markt für Rohmilch**
Oleksandr Perekhozhuk (PhD)
2007, 274 Seiten, ISBN 978-3-938584-24-8

Vol. 42 **Labor market behavior of Chinese rural households during transition**
Xiaobing Wang (PhD)
2007, 140 Seiten, ISBN 978-3-938584-25-5

Vol. 43 **Continuity and change: Land and water use reforms in rural Uzbekistan. Socio-economic and legal analyses for the region Khorezm**
ed. by Peter Wehrheim, Anja Schoeller-Schletter, Christopher Martius
2008, 211 Seiten, ISBN 978-3-938584-27-9

Vol. 44 **Agricultural economics and transition: What was expected, what we observed, the lessons learned (Vol I and II)**
ed. by Csaba Csáki, Csaba Forgács
2008, 634 Seiten, ISBN 978-3-938584-31-6

Vol. 45 **Theoretical and methodological topics in the institutional economics of European agriculture. With applications to farm organisation and rural credit arrangement**
Martin Petrick
2008, 223 Seiten, ISBN 978-3-938584-31-6

Vol. 46 **Agri-food business: Global challenges – Innovative solutions**
ed. by Thomas Glauben, Jon H. Hanf, Michael Kopsidis, Agata Pieniadz, Klaus Reinsberg
2008, 152 pages, ISBN 978-3-938584-33-0

Vol. 47 **Eine Analyse der Transformationsberatung für die "kollektive Landwirtschaft" während der ersten Transformationsphase (1989-1991) am Beispiel Ostdeutschlands: Lehren für Korea**
Jeong Nam Choi (PhD)
2009, 225 Seiten, ISBN 978-3-938584-36-1

Vol. 48 **Croatia's EU accession. Socio-economic assessment of farm households and policy recommendations**
Judith Möllers, Patrick Zier, Klaus Frohberg, Gertrud Buchenrieder and Štefan Bojnec
2009, 196 Seiten, ISBN 978-3-938584-35-4

Vol. 49 **Structural change in Europe's rural regions. Farm livelihoods between subsistence orientation, modernisation and non-farm diversification**
ed. by Gertrud Buchenrieder Judith Möllers
2009, 166 Seiten, ISBN 978-3-938584-39-2

Vol. 50 **Motive beim Weinkonsum – Unterschiede zwischen deutschen und ukrainischen Konsumenten**
Astrid Lucie Rewerts (PhD)
2009, 267 Seiten, ISBN 978-3-938584-40-8

Vol. 51 **Rural development as provision of local public goods: Theory and evidence from Poland**
Andreas Gramzow (PhD)
2009, 203 Seiten, ISBN 978-3-938584-41-5

Vol. 52 **Multi-level Processes of Integration and Disintegration. Proceedings of the Third Green Week Scientific Conference**
ed. by Franziska Schaft, Alfons Balmann
2009, 216 Seiten, ISBN 978-3-938584-42-2

Vol. 53 **Zur Bestimmung der Wettbewerbsfähigkeit des weißrussischen Milchsektors: Aussagefähigkeit von Wettbewerbsindikatoren und Entwicklung eines kohärenten Messungskonzepts**
Mikhail Ramanovich (PhD)
2010, 202 Seiten, ISBN 978-3-938584-44-6

Vol. 54 **Die Internationalisierung landwirtschaftlicher Unternehmen. Das Beispiel deutscher, dänischer und niederländischer Direktinvestitionen in den ukrainischen Agrarsektor**
Henriette Stange (PhD)
2010, 296 Seiten, ISBN 978-3-938584-45-3

Vol. 55 **Verhandlungsverhalten und Anspruchsanpassung im internationalen Verhandlungsprozess: Die WTO-Agrarverhandlungen zum Abbau exportwettbewerbsfördernder Maßnahmen**
Ildiko Lajtos (PhD)
2010, 195 Seiten, ISBN 978-3-938584-48-4

Vol. 56 **Challenges of education and innovation. Proceedings of the Fourth Green Week Scientific Conference**
ed. by Kelly Labar, Martin Petrick, Gertrud Buchenrieder
2010, 155 Seiten, ISBN 978-3-938584-49-1

Vol. 57 **Agriculture in the Western Balkan Countries**
ed. by Tina Volk
2010, 249 Seiten, ISBN 978-3-938584-51-4

Vol. 58 **Perspectives on Institutional Change – Water Management in Europe**
ed. by Insa Theesfeld, Frauke Pirscher
2011, 127 Seiten, ISBN 978-3-938584-52-1

Vol. 59 **Der ukrainische Außenhandel mit Produkten der Agrar- und Ernährungswirtschaft: Eine quantitative Analyse aus Sicht traditioneller und neuer Außenhandelstheorien**
Inna Levkovych (PhD)
2011, 232 Seiten, ISBN 978-3-938584-53-8

Vol. 60 **Regional structural change in European agriculture: Effects of decoupling and EU accession**
Christoph Sahrbacher (PhD)
2011, 244 Seiten, ISBN 978-3-938584-58-3

Vol. 61 **Structural Change in Agriculture and Rural Livelihoods: Policy Implications for the New Member States of the European Union**
ed. by Judith Möllers, Gertrud Buchenrieder, Csaba Csáki
2011, 247 Seiten, ISBN 978-3-938584-59-0

Vol. 62 **Improving the functioning of the rural financial markets of Armenia**
Milada Kasarjyan (PhD)
2011, 121 Seiten, ISBN 978-3-938584-60-6

Vol. 63 **Integrierte Strukturen im Agrar- und Ernährungssektor Russlands: Entstehungsgründe, Funktionsweise, Entwicklungsperspektiven und volkswirtschaftliche Auswirkungen**
Jürgen Wandel
2011, 758 Seiten, ISBN 978-3-938584-61-3

Vol. 64 **Goal Achievement in Supply Chain Networks – A Study of the Ukrainian Agri-Food Business**
Taras Gagalyuk (PhD)
2012, 204 Seiten, ISBN 978-3-938584-63-7

Vol. 65 **Impacts of CAP reforms on farm structures and performance disparities – An agent-based approach**
Amanda Sahrbacher (PhD)
2012, 284 Seiten, ISBN 978-3-938584-64-4

Vol. 66 **Land fragmentation and off-farm labor supply in China**
Lili Jia (PhD)
2012, 143 Seiten, ISBN 978-3-938584-65-1

Vol. 67 **Ausprägung interregionaler Disparitäten und Ansätze zur Entwicklung ländlicher Räume in Mittel- und Osteuropa**
Sabine Baum (PhD)
2012, 214 Seiten, ISBN 978-3-938584-68-2

Vol. 68 **Patterns Behind Rural Success Stories in the European Union: Major Lessons of Former Enlargements**
ed. by Axel Wolz, Carmen Hubbard, Judith Möllers, Matthew Gorton, Gertrud Buchenrieder
2012, 190 Seiten, ISBN 978-3-938584-69-9

Vol. 69 **Motives for remitting from Germany to Kosovo**
Wiebke Meyer (PhD)
2012, 142 Seiten, ISBN 978-3-938584-70-5

Vol. 70 **Effizienz russischer Geflügelfleischproduzenten: Entwicklung und Determinanten**
Elena Epelstejn (PhD)
2013, 272 Seiten, ISBN 978-3-938584-72-9

Vol. 71 **Econometric impact assessment of the Common Agricultural Policy in East German agriculture**
Patrick Zier (PhD)
2013, 172 Seiten, ISBN 978-3-938584-73-6

Vol. 72 **Determinants of non-farm entrepreneurial intentions in a transitional context: Evidence from rural Bulgaria**
Diana Traikova (PhD)
2013, 136 Seiten, ISBN 978-3-938584-75-0

Vol. 73 **Human capital differences or labor market discrimination? The occupational outcomes of ethnic minorities in rural Guizhou (China)**
Bente Castro Campos (PhD)
2013, 266 Seiten, ISBN 978-3-938584-76-7

In der Schriftenreihe *Studies on the Agricultural and Food Sector in Central and Eastern Europe* werden durch das IAMO Monographien und Tagungsberichte herausgegeben, die sich mit agrarökonomischen Fragestellungen zu Mittel- und Osteuropa beschäftigen. Wissenschaftlern, die in diesem Bereich forschen, steht die Schriftenreihe als Diskussionsforum offen.

In its series *Studies on the Agricultural and Food Sector in Central and Eastern Europe* IAMO publishes monographs and proceedings focusing on agricultural economic issues specific to Central and Eastern Europe. This series offers a forum to researchers studying this area.

ISSN 1436-221X
ISBN 978-3-938584-76-7