An investigation into the impact of health and health related behaviour on employment and wages in China.

Cui, Xuan

AN INVESTIGATION INTO THE IMPACT OF HEALTH AND HEALTH RELATED BEHAVIOUR ON EMPLOYMENT AND WAGES IN CHINA

XUAN CUI

Submitted to the University of Wales in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

School of Management

Swansea University

January 2014

Contents

List of Figures

List of Tables

List of Appendix Tables

Abbreviations

2SLS	Two Stage Least Squares
ADL	activities of daily living
ATET	Average Treatment Effect on the Treated
BHPS	British Household Panel Survey
BMI	Body Mass Index
CHFLS	Chinese Health and Family Life Survey
CHNS	China Health and Nutrition Survey
CMH	Commission on Macroeconomics and Health
CMS	Cooperative Medical Scheme
COEs	Collectively-Owned Enterprises
CPC	Communist Party of China
C-WGOC	Cooperative Meta-analysis Group of the Working Group on Obesity in China
DDP	Desirable Dietary Pattern
DWECS	Danish Work Environment Cohort Study
EBMI	Employee Basic Medical Insurance
ECHP	European Community Household Panel
ENDEF	Estudo Nacional da Despesa Familiar
HSE	Health Survey for England
FE	Fixed Effects
FIML	full information maximum likelihood
FIMSL	full information maximum simulated likelihood
GDP	Goss Domestic Product
GHQ	General Health Questionnaire
GIP	Government Insurance Program
GLS	General Least Squares
GSOEP	German Socio-Economic Panel
GSS	General Social Survey

HILDA	Household, Income and Labour Dynamics in Australia
HIS79	Health Interview Survey 1979
HRS	American Health and Retirement Survey
IV	Instrumental Variable
LATE	Local Average Treatment Effect
LIP	Labour Insurance Program
MLE	Maximum Likelihood Estimation
NCMS	New Cooperative Medical Scheme
NHS	Australian National Health Survey
NHSDA	National Household Surveys on Drug Abuse
NJCS	National Job Corps Study
NLS	National Longitudinal Survey
NLSY	National Longitudinal Survey of Youth
OLS	Ordinary Least Squares
PE	Physical Exam
PF	Physical function limitations
PSFD	Taiwan's Panel Survey of Family Dynamics
PSID	Michigan Panel Study of Income Dynamics
RE	Radom Effects
REC	Reemployment Centres
ROC	Receiver Operating Characteristic
SAMHSA	Substance Abuse and Mental Health Services Administration
SDW	Survey of Disability and Work
SEO	Survey of Economic Opportunity
SRHS	self-reported health status
SOEs	State-Owned Enterprises
UEBMI	Urban Resident Basic Medical Insurance

RLMS	Russia Longitudinal Monitoring Survey
RUBMI	Urban Resident Basic Medical Insurance Survey
WC	Waist Circumference
WHO	World Health Organization
WHR	Waist-to-Hip Ratio

Acknowledgements

I would like to express my sincere gratitude to may supervisor, Dr. Reza Arabsheibani and Dr. Melanie Jones, for their extraordinary support of my Ph.D study and research, for their patience, motivation, and immense knowledge. Without their illuminating instructions and intellectual comment, this thesis would not have been possible. I would also like to thank Professor John Sessions (external examiner from Bath University) and Professor David Blackaby (internal examiner) for your brilliant comments and suggestions.

I also appreciate the faculty members in the Economics department and my colleagues who contributed disputable viewpoints in research seminars and have given me many constructive suggestions.

Finally, and most importantly, I would like to take this opportunity to express my deepest thanks to my dear parents Tiejian Cui and Yinli Xie, for their endless love and encouragement through my entire life.

Summary

This thesis examines the impact of health and health-related behaviour on labour market outcomes in China. The data source used in this analysis is the second - seventh waves (1991, 1993, 1997, 2000, 2004 and 2006) of the China Health and Nutrition Survey (CHNS). The main analysis focuses on these related issues: the impact of health on employment; the impact of health on wages; and the impact of health-related behaviour on employment and wages.

This thesis provides new evidence to the limited existing evidence which examines the effect of health on labour market outcomes for the Chinese working-age population and therefore adds to the international literature in this area. It also provides evidence to policy makers, including the 17th and the 18th National Congress of China on improving the health of the whole nation. Chapter 1 outlines the motivation and objectives of the thesis. Chapter 2 briefly reviews the labour market and the health care system in China. Chapter 3 provides detailed description of the CHNS, and an overview of health and labour market outcomes in China.

Chapter 4 examines the effect of self-reported health status and obesity (as measured by Body Max Index (BMI)) on the probability of being employed. The empirical strategy involves a Random Effects (RE) probit model, a Fixed Effects (FE) logit model, and a recursive bivariate probit model which controls for the potential endogeneity of health. The results show that having better health status is found to be positively associated with employment for the Chinese working-age population. There is a statistically significant and negative relationship between having an unhealthy BMI and employment. These findings are consistent with the international literature. There is evidence of discrimination against people who have unhealthy body weight, when assuming SRHS completely capture the impact of health on productivity. The results are found to be vary within the population, the result finds that the positive relationship between better health and employment is larger for men and older working-age population than for women and the young working-age population, respectively. Furthermore, failure to account for the endogeneity of being overweight underestimates its impact on employment and this is consistent with international literature.

Chapter 5 estimates the impact of self-reported health status and obesity on hourly wages for non-agricultural employees. In addition to using the RE and the FE model, the analysis uses a Two Stage Least Squares (2SLS) and a Heckman (1979) selection correction method to separately control for the potential endogeneity of health and sample selection bias. The results show that having excellent health status is positively correlated with hourly wages for the Chinese workers. However, having an unhealthy BMI does not have a statistically significant effect on wages. However, for female workers being obese is predicted to reduce hourly wages by 22.40% compared to their normal weight counterparts, which is larger than that of the international literature. Based on the 2SLS method, the null hypothesis of exogeneity of health cannot be rejected for all sample groups within the Chinese population. However, the statistically significant and positive relationship between better health and wages largely disappears after controlling for sample selection bias.

Chapter 6 examines the simultaneous effect of drinking alcohol and smoking status on employment and wages. To allow for the potential endogeneity of drinking and smoking simultaneously, this chapter adopts Newey (1978)'s two-step and 2SLS methods to estimate the effect on employment and wages, respectively. For the overall population, both drinking and smoking status are found to be positively correlated with employment but not with hourly wages. The results confirm the literature that there is a positive and inverse U-shaped relationship between drinking and employment, which is also robust across different ways of measuring drinking behaviour. Smoking incidence is found to be positively correlated with employment for men but not for women. Furthermore, there is no evidence of endogeneity of drinking or smoking. Chapter 7 summarizes the key findings, explores the policy implications, and identifies limitations and issues for future research.

Chapter 1 Introduction

1.1 Context and Motivation of the Thesis

There are different understandings of views on the definition of health. The World Health Organization (WHO) defines health as '*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*' (WHO, 2006). Public policymakers usually emphasis the utilization and quality of medical care services when they talk about health, and mostly focus on education as human capital in the context of employment and productivity. In fact, health, along with education, job training and migration are four fundamental elements of human capital (Schultz, 1962). During the past three decades considerable attention has been paid to the role of education along with job training in labour market success. Human capital plays an import role in economic development (Simkovic, 2012). Importantly, research in labour economics has extended to investigate the relationship between health and labour market outcomes, which is also a pillar of the health economics (Fan, 2002).

Nowadays there is increasing public emphasis on the improvement of health. One reflection of this trend is the steady increase of inputs and expenditures in the health sector. It is a worldwide trend and China is no exception. However, economic resources are always scarce. A society has to optimize the allocation of limited resources to improve national health. For example, public policymakers must know the costs of poor health to society in order to conduct an efficient allocation of resources on preventing and treating diseases. As argued by Fan (2002), these costs take many forms. In addition to physical or mental pain, and the mental distress to family and friends, there are also economic costs. Direct economic costs include the medical expenses on the treatment, care and prevention of disease. Indirect costs include the economic losses that are caused by sickness. One important component of the latter is the lost output as a consequence of reduced labour market performers such as reduced working opportunities and salaries, including working hours and the chance of promotion (Contoyannis and Rice, 2001; Pelkowski and Berger *et al.*, 2004). It is therefore important to know health impact on labour market outcomes.

Most developed countries have fully recognized the importance of health care services. The financial expenditure of countries such as the UK, the US and Japan indicates that the share of expenditure on health care has become one of the important signs of prosperity and social advance of the country concerned (Wang, 2011). At present, China is going through economic transformation and has one third of working-age population in the word, which makes the gap between the health care system of China and that of developed countries particularly pronounced. The study of the relationship between health and labour market outcomes for the Chinese population is also new evidence in developing country where understanding the importance of the health is less advanced.

According to the Report of Commission on Macroeconomics and Health (CMH)[1] in China, 2005 (simply abbreviated as CMH, 2005 thereafter), investment in health has been put on China's development agenda. Furthermore, in the Eighteenth National Congress of the Communist Party of China (CPC) in 2012, general secretary of the CPC, Hu Jintao, also put promoting employment, increasing people's income and improving the level of people's health as three of the primary tasks on improving people's livelihood and innovation to strengthen the construction of social management. Therefore, a better understanding of the relationship between health, employment and wages which is the focus of this thesis is helpful in achieving these targets.

1.2 Aims and Objectives of the Thesis

The contribution of the human capital to economic development has been supported by many studies since the 1960s (Schultz, 1961; Becker, 1962, 1964). Although the relationship between health and economic development has come to the attention of economic researchers very early (Mushkin, 1962; Becker, 1964; Grossman, 1972a), most studies have focused on economic returns to education, for example, Card (1999), Blundell *et al.* (2000), Walker and Zhu (2003), and O'Leary and Sloane (2006). Compared to education, the studies on economic returns to health are relatively scarce (Zhang, 2003). Theoretically, health is often related to the labour

[1] The Report of the CMH, which is established by the WHO, aim at investigate the relationship between health, economic growth and poverty. So far this program has covered many countries in Africa, The Americas, Eastern Mediterranean, Europe, South-East Asia and Western Pacific.

market outcomes through productivity (Becker, 1964; Grossman, 1972b). Better health is likely to have a positive effect on the capacity or ability to perform a job and leads to higher productivity, hence individuals with better health are predicted to have a higher probability of being employed. On the other hand, low productivity is also associated with poor health, thereby decreasing individuals' wages.

Based on the China Health and Nutrition Survey (CHNS), the main area of analysis in this thesis lies with the impact of health on labour market outcomes in China. They correspond to the three main empirical chapters in this thesis, more specifically, Chapter 4 examines the impact of self-reported health status and Body Mass Index (BMI) on employment status; Chapter 5 investigates the impact of self-reported health status and Body Mass Index (BMI) on hourly wages; and Chapter 6 examines the impact of drinking and smoking on both employment status and wages. This thesis draws on the existing literature, and contributes new evidence in several ways.

First, it provides new evidence to the limited existing evidence on examining the effect of health on labour market outcome for the Chinese working-age population and therefore adds to the international literature in this area. A large body of international literature has shown that health has a substantial effect on employment (Luft, 1975; Stern, 1989; Bound, 1991; Baldwin and Johnson, 1994; Morris, 2004 and 2007; Greve, 2008; Cai and Kald, 2006; Zhang *et al.*, 2009; Cai, 2010; Garcia-Gomez *et al.*, 2010) and wages (Lee, 1982; Haveman *et al.*, 1994; Contoyannis and Rice, 2001; Cawley, 2004; Pelkowski and Berger, 2004; Gambin, 2005; Kline and Tobias, 2008; and Cai, 2009) in the US and in many other western countries such as the UK, Germany, Canada, and Australia. However, only a few studies, such as Benjamin *et al.*, (2003), Arcand and Labar (2005), Lin and Chen (2010) and Pan and Qin (2010), focus on specifically China. Compared to Arcand and Labar (2005), this thesis uses the data from the most recent waves of the CHNS. Further, unlike several of the studies above, this analysis is based on a representative sample of the Chinese population from CHNS. For example, Benjamin *et al.* (2003) and Pan and Qin (2008) rely on a sample of the elderly in rural areas and a sample of urban working-age adults, respectively, while Lin and Chen (2010) only focus on a special administrative region, Taiwan.

Second, this thesis examines multiple dimensions of health by using a range of measures of health. Currie and Madrian (1999) generalize the measures of health into eight types, which are: (1) the self-reported health status (SRHS); (2) nutritional status, for example, height, weight, BMI, or the level of energy intake; (3) the presence of health limitations; (4) the presence of functional limitations (e.g. activities of daily living); (5) the presence of chronic and acute diseases; (6) the presence of mental health problems or health related activities such as alcoholism and heavy smoking; (7) the utilization of medical care; and (8) mortality. The majority of studies focus on self-reported health status, nutritional status, and the presence of health limitations or chronic diseases, and the estimates of these effects on labour market outcomes are quite sensitive to the measure used (Currie and Madrian, 1999). Several studies that use multiple measures of health, such as Garcia-Gomez (2011) who analyse self-reported health status and the presence of chronic diseases, and Johansson *et al.* (2009) who use BMI, Waist Circumference (WC) and fat mass to measure obesity in estimating the effect of body weight on employment. Both Stern (1989) and Garcia-Gomez *et al.* (2010) use three types of health measures, namely self-reported health status, the presence of chronic diseases, and the presence of mental health problems to measure health, respectively. This thesis not only considers health as measured by self-reported health status and BMI, but also extends to two measures of health-related behaviour, including drinking alcohol and smoking cigarettes. While both these measures have previously been considered in the literature. Much of the existing evidence only focuses on the effect of drinking or smoking status. Auld (2005) argues that drinking and smoking are highly correlated, so that failing to control for one when estimating the effect of the other might lead to biased results. By considering health-related behaviour, this thesis also contributes to the literature relating to the simultaneous effect of drinking and smoking on labour market outcomes.

The data source used in this analysis is the second - seventh waves (1991, 1993, 1997, 2000, 2004 and 2006) of the CHNS. The data has several advantages including a stratified random sampling which increases precision and provides better coverage of the population, a large sample, and its panel nature, that is, the same individuals are interviewed over time. Unlike using the cross-sectional data, this chapter employs panel data analysis and therefore is able to control for individual unobserved factors which may bias the estimates (Arulampalam and Booth, 1998; Garcia-Gomez and Nicolas, 2006; Cai, 2010; and Garcia-Gomez *et al.*, 2010). In addition, the survey adopts almost the same index system and consistent definition of variables across time, such as consistent measure of weight and height (which are used to construct BMI) and same measure of monthly wages (which is used to construct hourly wages), thereby obtaining consistent estimates of the impact of health on labour market outcomes.

In addition to examining different measures of health, this thesis also considers other issues in estimating the relationship between health and labour market outcomes, particularly the potential endogeneity of health, and sample selection bias. The endogeneity of health might arise from either the reverse effect of labour market outcomes on health, or unobserved factors that affect health and/or labour market outcomes. Sample selection bias might arise when examining the impact of health on wages, since wages are only observed for employed workers, and the unobserved determinants of wages may be correlated to the employment status. A number of studies have employed an IV method to address the potential endogeneity of health in estimating its effect on employment (Stern, 1989; Bound *et al.*, 1996; Mullahy and Sindelar, 1996; Cawley, 2004; and Lin and Chen, 2010) and wages (Haveman *et al.*, 1994; Hamilton and Hamilton , 1997; Macdonald and Shields, 2001; Barrett, 2002; Lye and Hirschberg, 2004; Pelkowski and Berger, 2004; Auld, 2005; Jackle and Himmler, 2007; Cai, 2009; Lin and Chen, 2010; and Forbes *et al.*, 2010). This thesis uses IV methods and the Heckman (1979) selection model to control for the potential endogeneity of health and sample section bias, respectively.

The fifth advantage of this thesis is that the analysis is carried out for sub samples within the Chinese population, which facilitates on exploring the differences across genders, age groups and area of residence. Many international studies have estimated the effect of health on labour market outcomes between genders and relatively fewer studies capture the difference across age groups, although the evidence in examining these issues are somewhat mixed. Researchers argue that it is important to cover the entire working-age group rather than a narrowly defined sample of individuals such as studies focus on the retirement decisions to the older working-age population. In addition, the health status is likely to be generally qualitatively different across individuals at different age groups, so that empirical analyses should consider age groups separately (Mullahy and Sindelar, 1996). The analysis is also undertaken for urban and rural residents separately since composition of employment opportunities and health are expected to vary across these areas in China. Indeed employment reforms in the middle 1990s increased the employment pressure in urban areas, whereas in rural areas agriculture is still dominant and most of the workers are either self-employed or family-run enterprises. Furthermore, there are differences between urban and rural residents due to the substantial difference in the health care system between the urban and rural area in China (see Section 2.4 in Chapter 2).

1.3 Thesis Structure

The remainder of this thesis is structured as follows. Chapter 2 briefly reviews the labour market and the health care system in China. The Chinese labour market has experienced substantial changes in terms of job assignment and wage determination, along with the progress of the economic reforms over the past forty years. Over the same period the health of the population has improved as a consequence of changing health care system. Chapter 3 provides detailed description of the CHNS, and an overview of health and labour market outcomes in China.

The discussion of the theoretical and empirical issues regarding the main research themes, the corresponding reviews of literature and empirical models, together with results from the analyses are presented separately for each main research analyses. Chapter 4, 5 and 6 form the main part of the thesis. Chapter 4 examines the effect of self-reported health status and obesity (as measured by BMI) on the probability of being employed. In estimating the effect of having poor health status on employment, this thesis employs a recursive bivariate probit model, which takes into account the binary nature of both dependent and explanatory variables when controlling for the potential endogeneity of health. The instruments used in the above analyses involve information on whether individuals were born in the period of the 'Great Chinese Famine', 1958-1961, the prevalence of SRHS in the area in which the respondent lives, and information on individual diet.

Chapter 5 estimates the impact of self-reported health status and obesity on hourly wages. Based on an index measure of self-reported health and a continuous measure of BMI, the investigation of the relationship between health and wages involves a Two Stage Least Squares (2SLS) and a Heckman (1979) selection correction method to separately account for the potential endogeneity of health and sample selection bias. Following Bound (1991) and Cai (2009), the number of years having high blood pressure and the number of days being sick or injured during the last month are used as instruments for the self-reported health status. Following Haveman *et al.* (1994) and Jackle and Himmler (2007), the share of working-age adults who are obese in the area in which the respondent lives, along with the daily average energy intake of other working-age adults in respondent's household are used as instruments for BMI.

Chapter 6 examines the simultaneous effect of drinking and smoking status on employment and wages. To allow for the potential endogeneity of drinking and smoking simultaneously, this chapter adopts Newey (1978)'s two-step and 2SLS methods to estimate the effect on employment and wages, respectively. The instruments for drinking and smoking status, following Mullahy and Sindelar (1996), Macdonald and Shields (2001) and Barrett (2002), include the shares of the respondents' working-age household members who have the same type of drinking frequency, and the average amount of consumption for alcohol and for cigarettes of household members. The final chapter (Chapter 7) summarizes the key findings, explores the policy implications, limitations and identifies issues for future in this area.

Chapter 2 The Labour Market and the Health Care System in China

2.1 Background Information on the Chinese Economy

The Peoples' Republic of China was proclaimed in 1949 and the CPC came to power in the same year. Under the leadership of the Chairman Mao the First Five Year Plan was launched in 1953. This plan followed the Soviet economic model through large and socialized collective units in agriculture, state ownership in the modern sector, centralized planning and concentration of resources in heavy industry (Worden *et al.*, 1987). With the great help and support of the Soviet Union, the First Five Year Plan was implemented successfully[2]. However, this highly centralized, industry-biased plan was unable to be carried out continuously due to the increasing imbalance between agricultural and industrial growth, lack of flexibility in the decision-making process and dissatisfaction with inefficiency (Worden, *et al.*, 1987). Therefore, the Second Five Year Plan, which was also known as 'The Great Leap Forward' was launched in 1958 to mobilize the Chinese huge labour resources into an agricultural-based industrialization drive (Gonçalves, 2006). Unfortunately, the Second Five Year Plan failed badly because of serious disproportions of the economic structure, deterioration of China-Soviet relations, widespread famine and natural calamities which is also known as the 'The Great Chinese Famine' (Liu, 2010). Thereafter, China stepped into a readjustment and recovery period to repair the failed system. According to Worden *et al.* (1987), during the readjustment and recovery period, the economic model followed a balanced growth of the 'Agriculture First' policy. In other words, apart from laying particular stress on agricultural development, the model not only contained elements of the first industrially oriented and highly centralized model but also followed the decentralization of ownership and decision making that characterized the Second Five Year Plan, which can be treated as a combination of the previous two models (Worden *et al.*, 1987).

During the next period which was known as the Great Proletarian Cultural Revolution (also known as the Cultural Revolution) and lasted from 1966 until 1976, there was little change in the basic economic model or official economic policies

[2] Between 1952 and 1957, industrial production increased substantially, averaging rises of 19% a year, and the agricultural output increased at a rate of 4% a year. In addition, the national income increased at an average annual rate of 9% (Worden *et al.*, 1987).

(Worden *et al.*, 1987), and the economy ran smoothly[3]. The Cultural Revolution can be thought of as a political upheaval reflecting a power struggle inside the CPC (Joffe, 1966; Lee, 1978). However, the Cultural Revolution inflicted great damage upon the society. Further, the education of the Chinese youth was largely hit by this political movement (Deng and Treiman, 1997). As documented in Deng and Treiman (1997), Zhang *et al.* (2007), and Li *et al.* (2010), almost all primary schools, junior/senior high schools and universities were closed entirely during the climax of the Cultural Revolution (1966-1968). No teaching was carried out and no new students were admitted. Some primary schools and junior high schools reopened after 1968 and senior high schools reopened after 1971[4], therefore students of normal graduation age for primary school were able to go on to high school and children aged 7-9 began primary school. As for students who would have completed junior/senior high school between 1966 and 1968, although they did not receive a formal junior/senior high school education, still they were given diplomas. Universities were reopened in 1972, and students were given university degrees without formal education and allocated jobs. At the time when schools were closing, lots of students whose schooling was stopped or delayed actively participated in all kinds of political activities (these students were called the 'Red Guards'). As a result, there was a problem of having two extra cohorts of students at the same time after the high schools were reopened. At the end of 1968, Chairman Mao began 'Down to the Countryside Movement' in order to solve the problem. The main function of this rustication program was to '*simultaneously discharge the Red Guard, reduce unemployment in urban areas, and increase agricultural productivity*' (Deng and Treiman, 1997, pp.7). The send-down movement was ended in 1978, and most of the urban youth moved back to cities after the end of the movement.

Following the Cultural Revolution, as stated in Qiu (2007), the realization of the potential for rapid expansion of an economy with low technical and allocative efficiency stimulated the economic reforms launched in 1978. With Mao Zedong's death in 1976, Deng Xiaoping took power in 1978 and oversaw an unprecedented economic boom which continues today (Jakobson and Pursiainen, 2001). At the

[3] This did not include the worst period of the upheaval (1967-1968). See: http://www.edu.cn/20030821/3089681.shtml (Chinese version). [cited 30th Sept 2011].

[4] Rather than receiving traditional education, students spent much of their school time studying Marxism and Mao's thought, and leaning peasant and worker techniques.

Third Plenum of the National Party Congress's Eleventh Central Committee in December 1978, the Chinese authorities resolved to make economic development a top priority and decided to undertake a program of gradual but fundamental reform of the economic system (Worden *et al.*, 1987; Yueh, 2004). The economic reform was embodied in the 'Open Door' and the modernization of agriculture, industry, national defence, and science and technology (the so-called 'Four Modernizations'). Subsequently the economic reform began step by step, starting with the countryside. Qiu (2007) states that the reforms within the rural areas provided new opportunities for farm families to take control of their labour allocation decisions through the dissolution communes across the country and the implementation of the household responsibility system. According to Worden *et al.* (1987), the economy quickly recovered from the stagnation of the Cultural Revolution between 1976 and 1978. Agricultural production reached a record harvest in 1978 and the industrial output increased by 13% in the same year. Strictly speaking, 1979-1981 can be treated as a stage of readjustment. On the strength of successes in this readjustment stage, the reform program was expanded, and the leadership under Deng Xiaoping stressed time and again that China's basic policy was 'reform and opening', that is, reform of the economic system and opening to foreign trade (Worden *et al.*, 1987). Economic reforms within urban areas were intensely implemented since 1984 and the kernel of the reforms is the reforms of State-Owned Enterprises (SOEs) and wages. The market-oriented economic reform accelerated after 1992 especially in the coastal regions due to Deng's 'southern tour' (Bishop *et al.*, 2005; Park *et al.*, 2006). During the period when China was shifting from a planned economy to a market economy, market mechanisms increasingly playing a leading role in the processes of resource allocation. Li and Dong (2008) state that the gradual penetration of market-oriented reforms, together with the growing proportion of the private sector in the whole economy, implied that more enterprises would be face increasingly fierce market competition environment. However, it is argued that the rapid spread of the restructure and privatization within the SOEs caused massive layoffs, early retirements and drop out of the labour market. In general, economic reform gave managers greater power to decide wages as well as other operational autonomy. Meanwhile the government's intervention in the labour market become smaller, wages began to be determined by the market.

2.2 The Chinese Labour Market during the Pre-reform Period

China has carried out progressive reforms since late 1970s. China's labour market has also experienced a great change along with the progress of the reforms. Before the reforms, China implemented a centrally planned economy in which the central control power of labour allocation was in hands of the regional/local labour department assigned by the state authority. Qiu (2007) provides comprehensive descriptions of the Chinese labour market in the pre-reform period (during the late 1950s and the late 1970s). She states that the Chinese labour market is not a labour market in the conventional sense or as assumed in human capital theory[5]. Brooks and Tao (2003) state that the labour market in this period was characterized by direct allocation of jobs and administrative control of wages. The employment policy under this traditional system is to ensure the normal reproduction of labour force with low-wage levels. According to Yueh (2004), governmental planning, rather than the market, exercised a monopoly over the supply of and demand for labour. Hence major employers such as SOEs, Collectively-Owned Enterprises (COEs) were not allowed to recruit freely (Qiu, 2007). Regardless of the level of human capital, from university graduates and secondary technical school graduates, up to the demobilized soldiers and unemployed youths, all of them were subordinated themselves to unified state control. In other words, as stated in Qiu (2007), a high level of human capital stock was not linked to high productivity, which is chiefly embodied in three aspects: (i) enterprises were not able to hire or fire employees; (ii) employees were not able to choose jobs which would suit their abilities or interest; (iii) there was no mechanism to encourage individuals to work efficiently. Worker mobility was very limited (Arcand and Labar, 2005; Bishop *et al.*, 2005). The initial job assignment was of vital importance to employees as their first job was normally their only job in their whole life, the so-called 'iron rice bowl' (Yueh, 2004; Qiu, 2007). As defined by Ding and Warner (2001), the 'iron rice bowl' was characterized by a regime of unified job allocation, guaranteed life-long employment, and cradle-to-grave welfare. The authoritarian structure of job allocation was aimed at avoiding unemployment (Yueh, 2004). However, China's population remains predominantly rural[6]. Therofore,

[5] Human capital theory assumes that in a labour market with few imperfections individuals are able to decide how much education to obtain and to increase their marginal productivity and subsequently their future incomes.
[6] According to China Statistical Yearbook 2003, over 80% were classified as rural.

restriction on rural-urban labour migration was mainly achieved through the household registration system (the *hukou* system). Chan and Zhang (1999) state that the *hukou* system, which was formally set up in 1958, is a principal approach of social control adopted by the state and serves the function of migration control. Moreover, the *hukou* system effectively constrained the development of a national labour market (Brooks and Tao, 2003; Qiu, 2007). As described by Qiu (2007), people were required to register with their local authorities in order to obtain a permanent residency permit, and a place of work. According to, Brooks and Tao (2003), only people with an urban *hukou* were able to stay in cities and obtain preferential access to city services such as health, education and social security.

Prior to 1978, wage levels remained essentially unchanged and were determined by a major concern for equity. In other words, the Chinese labour-management system[7] was characterized by both guaranteed lifetime employment and centrally administered (fixed) wages (Kidd and Meng, 2001). Based on the Soviet model, a national unified wage-system which focuses on Chinese industries was established in 1950s (Ding and Warner, 2001; Qiu, 2007). The setting of the wage grade was based on individuals' personal endowments, especially the level of education and work experience (Kidd and Meng, 2001). Ding and Warner (2001) conclude that there were three primary features of this wage system: first, in order to increase capital accumulation for investment in heavy industries (such as energy and steel) China had maintained a 'low-wage policy' for a long time; second, the centrally fixed wage scales was a considerable 'flat' structure of reward in which wage differentials were purposely kept low; third, under this wage system, enterprises (or the employers) were deprived of autonomy to determine or adjust the salary classification, promotion method or the wage levels of their employees in the light of business needs. Although employees received low (but homogeneous) salaries, they still received much in-kind benefits such as medical care and subsidized housing (Arcand and Labar, 2005; Chen *et al.*, 2005). Zhang (2009) states that in the absence of the rights of the independent operation and management responsibility system, enterprises had no incentive to improve productivity. Moreover, the centrally

[7] Ding and Warner (2001) further describes three characteristics of the China's labour-management system in the pre-reform period, that is, life-time employment – 'iron rice bowl', centrally administered wages – 'iron wage', and the state-controlled appointment and promotion of managerial staff – 'iron chair'.

determined employment quota combined with the wage system indicated that enterprises faced an annual quota for their total wage bill (Kidd and Meng, 2001). Kidd and Meng (2001) highlight two major weaknesses of the Chinese labour market in the pre-reform era: (i) the lifetime secured employment system lead to over-staffing, idleness and low productivity; and (ii) the fixed wage system placed too much emphasis on individuals' personal endowments (qualification and work experience), thus differences in actual productivity of workers were not reflected in wages.

2.3 The Chinese Labour Market during the Reform Period

As discussed in Section 2.2, the link between labour productivity and human capital in the pre-reform Chinese labour market was broken since the labour arrangements were made without much reference to the needs of enterprises or the characteristics of workers (Yueh, 2004; Qiu, 2007). Therefore, understanding the inefficiency of low productivity associated with the pre-reform labour allocation had a significant stimulatory effect on the labour reforms (Qiu, 2007).

China's economic reform, which launched in 1978 and accelerated at the beginning of the 1990s, has led to substantial changes of the Chinese labour market. It implied a transition from a centrally planned to a more flexible and increasingly market-oriented labour market. Changes in both systems of job assignment and of wage determination were the most noticeable and were worth mentioning.

2.3.1 Labour Allocation and the Employment Reforms

From the end of 1970s to the early 1980s, a large increase in the working-age population[8] led to a relatively faster labour force growth in comparison to the population growth. According to statistics, the whole population and the working-age population increased 16% and 28%, from 1980 to 1990, respectively[9]. In addition, a large number of young people in the urban area became 'youth waiting for job

[8] According to Hardee-Cleaveland and Banister (1988), SEPI (2001), and Bergaglio (2008), 1960s-early 1970s is a 'baby boom generation' period.
[9] From 1990 to 2000, the whole population and the working-age population increased 11% and 16%, respectively.

assignments' due to historical reasons[10] (Ding and Warner, 2001). Therefore, in order to relieve the urban unemployment pressure, many enterprises and local labour departments were encouraged to provide training and job placement services for the jobless youth. In the rural area, communes were gradually replaced by the household responsibility system indicating that rural households were allowed to independently determine their own business. During the same period, the non-state sector, such as individual enterprises, domestic private enterprises, collective enterprises, domestic joint ventures and foreign-invested enterprises, emerged and grew (Chen *et al.,* 2005). Hence a high level of employment could be maintained under the central policy where surplus labour force was placed in both COEs and rural collectives (SEPI, 2002).

To replace the traditional life-guaranteed employment system, a labour contract system[11] was experimentally introduced in mid-1980s, and was formally established in 1994[12] (Ding and Warner, 2001). The presence of the labour contract system broke the 'iron rice bowl'. On the one hand, enterprises had more autonomy in terms of labour arrangement and recruitment method (Kidd and Meng, 2001); on the other hand, employees were provided opportunities to choose their jobs by removing restrictions on job mobility (Brooks and Tao, 2003). During the 1990s, there was a rapid growth in private, foreign-funded and other source-funded enterprises (Du, 2009).

The reforms were carried out mainly in urban China, particularly in SOEs and COEs, since 1984 (Qiu, 2007). The centrally maintained high level of employment vanished with the move toward a flexible labour market. Facing greater competition across sectors, employment in both SOEs and COEs decreased considerably during the mid 1990s and the early 2000s as enterprises laid off their workers in order to increase efficiency and productivity (Brooks and Tao, 2003; and Yueh, 2004). Such laid-off workers (also called 'xiagang' workers) were still officially employed by their enterprises without any working activities (SEPI, 2002). Instead of receiving regular

[10] As mentioned in Chapter 2, during the Cultural Revolution period, millions of urban youth were sent to the countryside to 'learn from the farmers', which is also called 'Down to the Countryside Movement'. After this movement, most of them returned to cities.

[11] As stated in Ding and Warner (2001), all urban workers were required to place on this new system regardless of the ownership type of their employers.

[12] Until the early 2000s, most urban employees were placed on this contract system.

earnings, they only received low income benefits[13]. In addition, they could join the Reemployment Centres (REC), and obtain skill retraining and job seeking assistance. However, these laid-off workers had to be transferred to the registered unemployed or leave the labour market if they failed to be reemployed by their old work unit or to find another job after more than three years (Brooks and Tao, 2003). Brooks and Tao (2003) also find that most of the laid-off workers possess two characteristics – a lack of skills and a low educational level. Table 2.1 shows that the total number of laid-off workers peaked in 2000 and reduced sharply after 2002. The number of laid-off workers declines over time because they have been reemployed by their original work unit, or have been transferred to either registered or unregistered unemployed, or have dropped out of the work force. Northeastern provinces have the highest absolute number of laid-off workers among the nine provinces within China and the Southwest has the lowest.

2.3.2 The Wage Reforms

Since 1978, wage reform has been undertaken in three main stages which focused on urban area to devise a more market-oriented wage system. In other words, the aim of these reforms is to promote efficiency and reward productivity.

The First Stage (Late 1970s - Early 1980s)

In the first step, the enterprise's centrally administered total wage quota was replaced by a more flexible total wage bill which gave a wider autonomy associated with the distribution of profits (Kidd and Meng, 2001). The new system allowed enterprises to use a share of their retained profits to provide employees with bonuses which were aimed at improving overall labour productivity (Bishop *et al.*, 2005; Chen *et al.*, 2005). However, Arcand and Labar (2005) argue that the structure of this bonus-scheme was still heavily equity-based.

The Second Stage (Early 1980s - Mid 1990s)

The reforms of the second period mainly focus on wages, that is, to link wages, rather than only bonuses, to the actual productivity and profitability of the enterprises (Ding and Warner, 2001; Yueh, 2004; Bishop *et al.*, 2005). During the mid-1980s,

[13] As argued by Yueh (2004), a number of laid-off workers normally cannot receive any benefits.

the nationally-unified system was converted into a wide range of wage systems (e.g. structural wages system and floating wage system) in which the wages of the workers were determined by various indicators such as education, experience, skill and physical effort (Ding and Warner, 2001; Kidd and Meng, 2001). According to Yueh (2004), enterprises were authorized to set their own wage structures based on their wage budgets since the beginning of the 1990s. In 1993, the state implemented a minimum wage system through the Labour Law. The minimum wage system requires that for employees who offer normal works within the legal working hours, the minimum payments must not be lower than the standard for the local minimum payment. The Labour Law stipulated that "*The determination and readjustment of the standards on minimum wages shall be made with reference to the following factors in a comprehensive manner: (1) the lowest living expenses of laborers themselves and the average family members they support; (2) the average wage level of the society as a whole; (3) labor productivity; (4) the situation of employment; and (5) the different levels of economic development between regions.*" (The Labour Law of the People's Republic of China, Chapter V, Article 49. Available: http://www.worldlawdirect.com/article/3213/labour-law-peoples-republic-china.html. Cited 12th Sep 2011).

The Third Stage (Mid 1990s - Early 2000s)

Since the recent wage reform period, state enterprises were given complete autonomy to set their own wage system (Ding and Warner, 2001). The monthly wages formerly consisted of six parts, including the basic wage, bonuses, benefits and subsidies, overtime wages, supplementary wages and an 'unknown' component, mainly based on hardship (Yueh, 2004). Instead of the previous wage categories, there have been two components of wages in enterprises: fixed wages (i.e. basic wage, seniority wage, insurance and a housing fund) and bonuses. Furthermore, the enterprises set their wage standards in accordance with job responsibilities, labour skills and labour intensity (Ding and Warner, 2001).

2.4 Background Information of the Chinese Health Care System

Over the years, the Chinese health care system has made considerable progress in improving the health outcomes of the Chinese population such as reducing mortality and increasing life expectancy (Chow, 2006); despite this, the rate of improvement has slowed in recent years (Hew, 2006). Figure 2.1 and Table 2.2 illustrate the expenditure on health for China and other five selected countries from different dimensions. Figure 2.1 (a) and (b) depicts the trend of total expenditure on health as a percentage of GDP and the trend of general government expenditure on health as a percentage of total government expenditure from 1995 to 2009, respectively. Both figures show that the health expenditure in China is comparably low among the six countries, which is consistent with Hew (2006). Table 2.2 summarises the distribution of health expenditure between government and private expenditure on health as a percentage of total expenditure on health. It shows that Japan and the UK have the largest share of government expenditure which both account for 81% on average over the fifteen years, whereas India has the smallest share (less than 35%). The distribution of health expenditure between government and private expenditure on health for Brazil and the US are relatively balanced. Interestingly, the government share on health expenditure in China decreases for the first seven years and then increases for the rest of the fifteen years (reaches rock-bottom in 2001), and a similar trend can be found in Figure 2.1 (b). These trends will be explained later in more detail. There are two different health care systems in China – one is to serve the urban residents, the other is to serve the rural residents (Wong *et al.*, 2007).

2.4.1 The Urban Health Care System

Prior to the mid 1990s, there were two primary elements of public health care system that offered the basic healthcare services in urban China – the Government Insurance Program (GIP) and the Labour Insurance Program (LIP). The GIP covered government employees and their families through which medical costs were financed partially or fully by government budget. The LIP covered employees in the SOEs and certain COEs, through which medical expenses were paid partially or fully by their enterprises (Liu and Zhao, 2006). Liu and Zhao (2006) state that under this health care system the health of the urban Chinese population improved considerably. However, the system has three major drawbacks: (i) cost inefficiency due to 'the moral hazard problem inherent in the old system' (Wong *et al.*, 2007); (ii) poor capacity of risk pooling across the work units (Liu, 2002; and Liu and Zhao, 2006); (iii) inadequate insurance coverage (Liu, 2002; Liu and Zhao, 2006; and Du, 2009).

Realization of the health care system deficiencies motivated China to reform its health insurance system.

In 1998, a new Employee Basic Medical Insurance (EBMI) consisting of individual medical saving account combined with socially risk pooled fund was established to replace the previous insurance system (Wong *et al.*, 2007). The new insurance system has two virtues: (i) it expanded insurance coverage to urban employees (Liu, 2002; and Chow, 2006); (ii) individuals began to share the healthcare responsibility with enterprises through a co-payment mechanism which was used to help in solving 'the moral hazard problem inherent in the old system' (Wong *et al.*, 2007).

Following the establishment of the EBMI, another health care program, Urban Resident Basic Medical Insurance (URBMI), was also implemented in 2007. This program has made great contributions to the Chinese health system in terms of covering the urban unemployed population and reducing financial barriers to care for population with poor health and socioeconomic status (Lin *et al.*, 2009).

2.4.2 The Rural Health Care System

Since the 1950s, a traditional Cooperative Medical Scheme (CMS) which was financed by the communes' welfare funds was adopted in rural China. Before economic reform the CMS was able to provide sufficient funds to the health care of the rural population (Chow, 2006). Since the collapse of the communes in the early 1980s, most counties suddenly lost their welfare funds to finance the CMS (Lei and Lin, 2009). Therefore, the major portion of the rural population became uninsured between 1985 and 2003, and was forced to pay all medical costs themselves, which in turn lead to rise in rural poverty (Liu, 2004).

To solve the low coverage of the health insurance and to improve the health care of the rural population, a New Cooperative Medical Scheme (NCMS) was introduced in 2003 (Lei and Lin, 2009). Using the CHNS, both Lei and Lin (2009) and Lv (2009) find that the implementation of the new health care system greatly expanded the coverage of the scheme. However, Lei and Lin (2009) argue that the new scheme did not improve health[14] of the rural population. As mentioned earlier, both in Figure 2.1

[14] In Lei and Lin (2009), health is measured by both self-reported health and incidence of sickness injury.

(b) and in Table 2.2 there are a downward trend for the first seven years and an upward trend for the following years. The establishment of the NCMS possibly explains these trends regarding the government expenditure on health in China.

In summary, UEBMI, URBMI, and NCMS jointly provide comprehensive healthcare benefits for the whole population in China. Meanwhile, private insurance has played an increasing role in field of the Chinese health care system due to both the increasing number of private insurance purchase rate and the increasing number of the private insurance types (Liu, 2002; and Du, 2009).

2.5 Health Outcomes in China

Over the last twenty years, China has made progress in improving the health outcomes among the population in terms of life expectancy and mortality trends. Figure 2.2 shows the trends of life expectancy at birth (in years) for China and the same five countries as used in Figure 2.1. It can be seen that the life expectancy for the Chinese population increased from 68 in 1980-1985 to 72 in 2005-2010. As a whole, life expectancies for the developed countries are longer than that for the developing countries. China had the longest life expectancy among the three developing countries. However, the growth rate of life expectancy for China was the slowest.

The trends of infant mortality rate (infant deaths per 1,000 live births) and the trends of under-five mortality rate (deaths under age five per 1,000 live births) are presented in Figure 2.3 (a) and Figure 2.3 (b), respectively. As shown in Figure 2.3 (a), infant mortality rate in China decreased by 42% (decreased from 38 per 1,000 people in 1980-1985 to 22 per 1,000 people in 2005-2010) and mortality rate under age five declined by over half (declined from 57 per 1,000 people in 1980-1985 to 26 per 1,000 people in 2005-2010). Overall, China had both the lowest infant mortality rate and the lowest under-five mortality rate among the three developing countries during the two decades. However, the decrease in infant mortality rate in China was far behind that of other two developing countries. Although the infant mortality rate and under-five mortality rate fell considerably in China, its absolute rate is still higher than that in the developed countries (the rate is less than 10 per 1,000 people for Japan, UK and US over the twenty years).

Figure 2.4 illustrates the changes in distribution on three causes of death across these six countries, that is, death due to communicable diseases (for example, infectious diseases, parasitic diseases and respiratory infections), non-communicable diseases (for example, diabetes mellitus, neoplasms, cardiovascular diseases, cerebrovascular disease, chronic obstructive pulmonary disease and heart disease) and injuries. As shown in this figure, the share of deaths due to non-communicable diseases and injuries in developed countries were greater than that in developing countries. In the three developed countries, deaths due to non-communicable diseases and injuries all accounted for more than 90% for both selected years. Among the three developing countries, China had the largest share of deaths due to non-communicable diseases and injuries (accounted for 67% in 2002 and 85% in 2008, respectively). Furthermore, non-communicable diseases such as cerebrovascular disease, chronic obstructive pulmonary disease and heart disease, are found to be the main causes of deaths in China (CMH, 2005).

2.6 One-Child Policy in China

China applied its unique one-child policy in 1979, which aimed at limiting the large population growth and promoting economic growth since the 1950s for both rural and urban areas. This policy is still in effect today. Under this policy, each couple was only allowed to have one child. In addition, couples who obeyed the policy were rewarded (such as receiving cash bonuses and preferential housing allocations), and those who violated the policy were penalized (such as paying large fines which vary in cities and counties). However, in most rural areas, families were allowed to have a second child if the first child was a girl (Qian, 1997; Li *et al.*, 2005; and Hu, 2009). Since rural families with low savings and without pensions need children to support them in old age, plus a fact that male labour is essential, the gender of the first child is more import to them, compared to the urban couples who normally received pensions after retirement (Worden *et al.*, 1987; Kane and Choi, 1999; and Rosenzweig and Zhang, 2009). In other words, the implementation of the one-child policy was more successful in urban than in rural areas. According to Li *et al.* (2005) and Li and Zhang (2007), the policy was initially introduced to the *Han* (the ethnic majority group) Chinese. Only two ethnic minority groups were subject to the one-child policy at the end of the 1980s: *Zhuang* and *Manchu*[15], which are first and second largest ethnic groups in China.

In terms of the population-related advantages, after the introduction of the policy the population growth slowed and the fertility rate reduced. For example, the birth rate in China decreased from 18.25 births per 1,000 in 1978 to 12.17 births per 1,000 in 2010, and the total fertility rate decreases from 2.91 births per women to 1.54 births per women[16]. The vice minister of the State Family Planning Commission, Zhao, said that the one-child policy has prevented 300 million births which further makes great effort on improving people's living standards and ensuring economic stability. [17] Moreover, this policy helped provide a better health care such as increasing access to vaccinations and improving disease treatment (Thompson, 2010). However, the one-child policy is not without problems. The potential impact of the fast shrink of the young population associated with the one-child policy on the China's ageing population has been argued by Worden *et al.* (1987). A journalist writing for the Times claimed that "*China may be forced to reconsider its one-child policy after census data revealed that rural-to-urban migration and rising life expectancies have led to a rapidly aging population.*" (The Times, April 29, 2011. Available: http://www.thetimes.co.uk/tto/public/profile/Leo-Lewis. Cited 5th Oct 2011). Economists predict that, by 2050, the number of elderly people in China will be over 300 million. The rapid reduction of the young labour force along with the increasing elderly population implies that in the future the burden of caring for elderly will increase. Moreover, Branigan (2011) states that one-child policy in conjunction with son preference has created gender imbalance problem.

2.7 Conclusion

To sum up, Chinese economy has carried out progressive reforms over the past forty years and the Chinese labour market has also experienced substantial changes along with the progress of the reforms. It implied a transition from a centrally planned to a more flexible and increasingly market-oriented labour market. Changes in both systems of job assignment and of wage determination were the most noticeable. In

[15] Most of *Zhuang* lived in Guangxi; *Manchu* can be found in Beijing, Inner Mongolia, Heilongjiang Hebei, Jilin and Liaoning.
[16] Statistical data is obtained from the CIA World Fact Book and the Trading Economics.
[17] http://www.china.org.cn/english/2002/Oct/46138.htm. [cited 1st Jan 2012].

other words, the reforms gave enterprises greater power to decide wages as well as other operational autonomy, although the rapid spread of the restructure and privatization within the SOEs caused some negative effects such as massive layoffs, early retirements and caused individuals to drop out of the labour market. Over the same period there were three main health insurance systems (which are UEBMI, URBMI, and NCMS) as well as the growing private insurance system to jointly provide comprehensive healthcare benefits for the Chinese population, which lead to considerable progress in improving the health outcomes. The mortality rate decreased and life expectancy increased, although the speed of these improvements were slow relative to other developing countries, and the actual measures of the health outcomes are still far behind that of the developed countries.

Figure 2.1 (a) Health Expenditure by Country

Total Expenditure on Health as a Percentage of Gross Domestic Product (GDP)

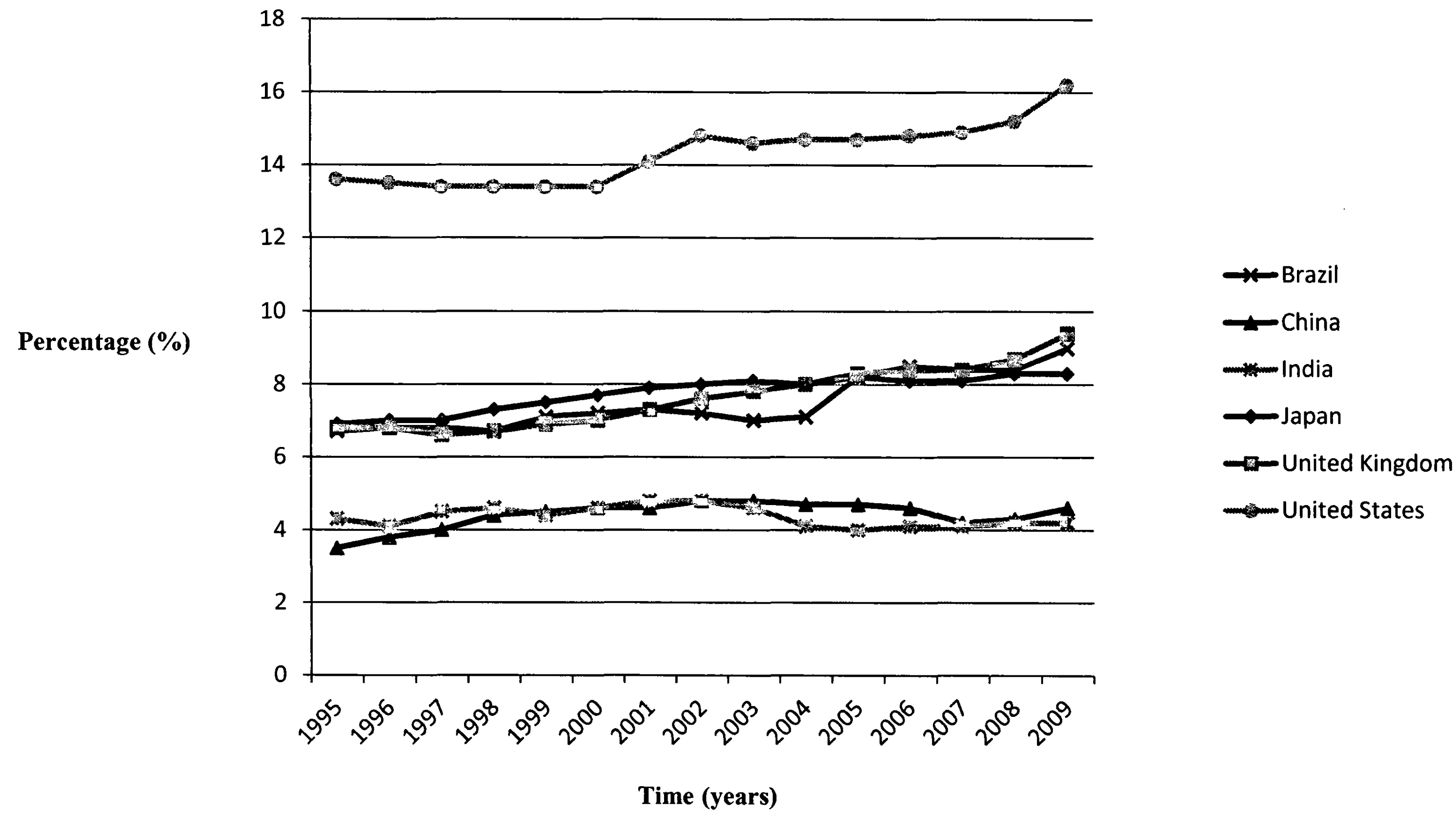
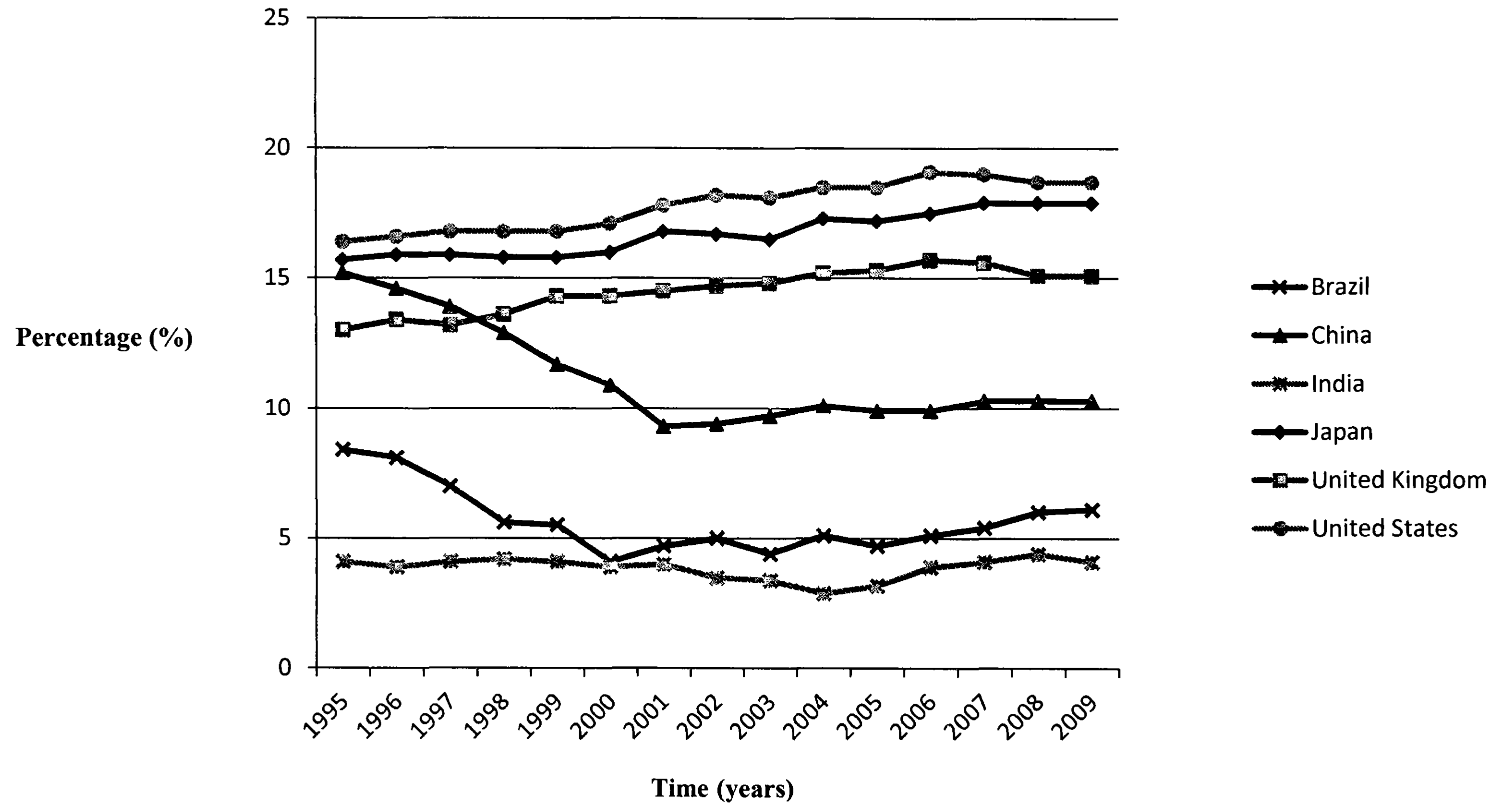


Figure 2.1 (b) Health Expenditure by Country

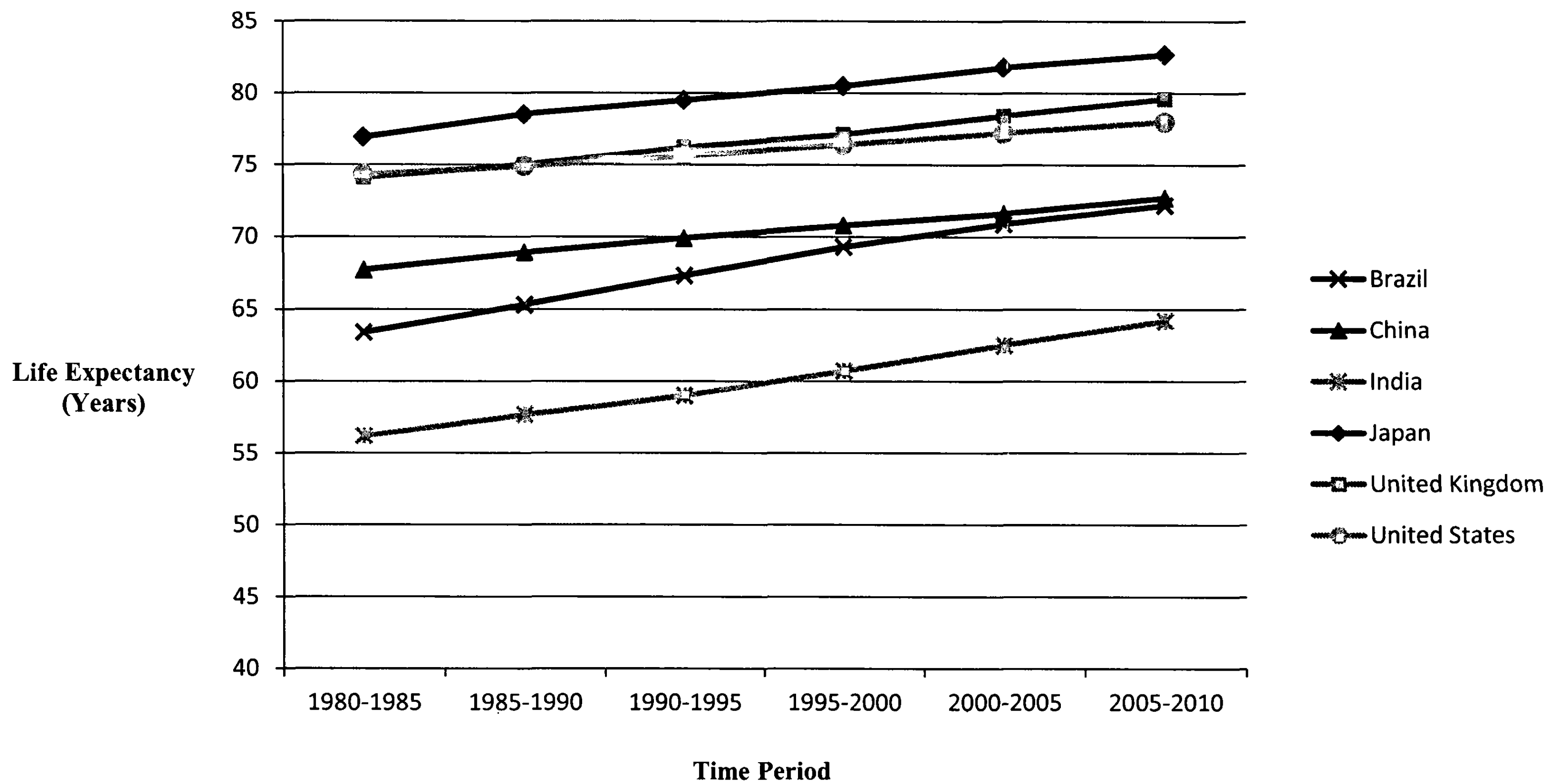
General Government Expenditure on Health as a Percentage of Total Government Expenditure[1]



Source: WHO World Health Statistics. Available: http://apps.who.int/ghodata/?vid=10012#.

[1] It includes the resources channelled through government budgets, and the expenditure on health by parastatals, extra budgetary entities and the compulsory health insurance.

Figure 2.2 Life Expectancy at Birth by Country



Sources: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, World Population Prospects: The 2010 Revision. Available: http://esa.un.org/unpd/wpp/index.htm.

Figure 2.3 (a) Mortality Rate by Country

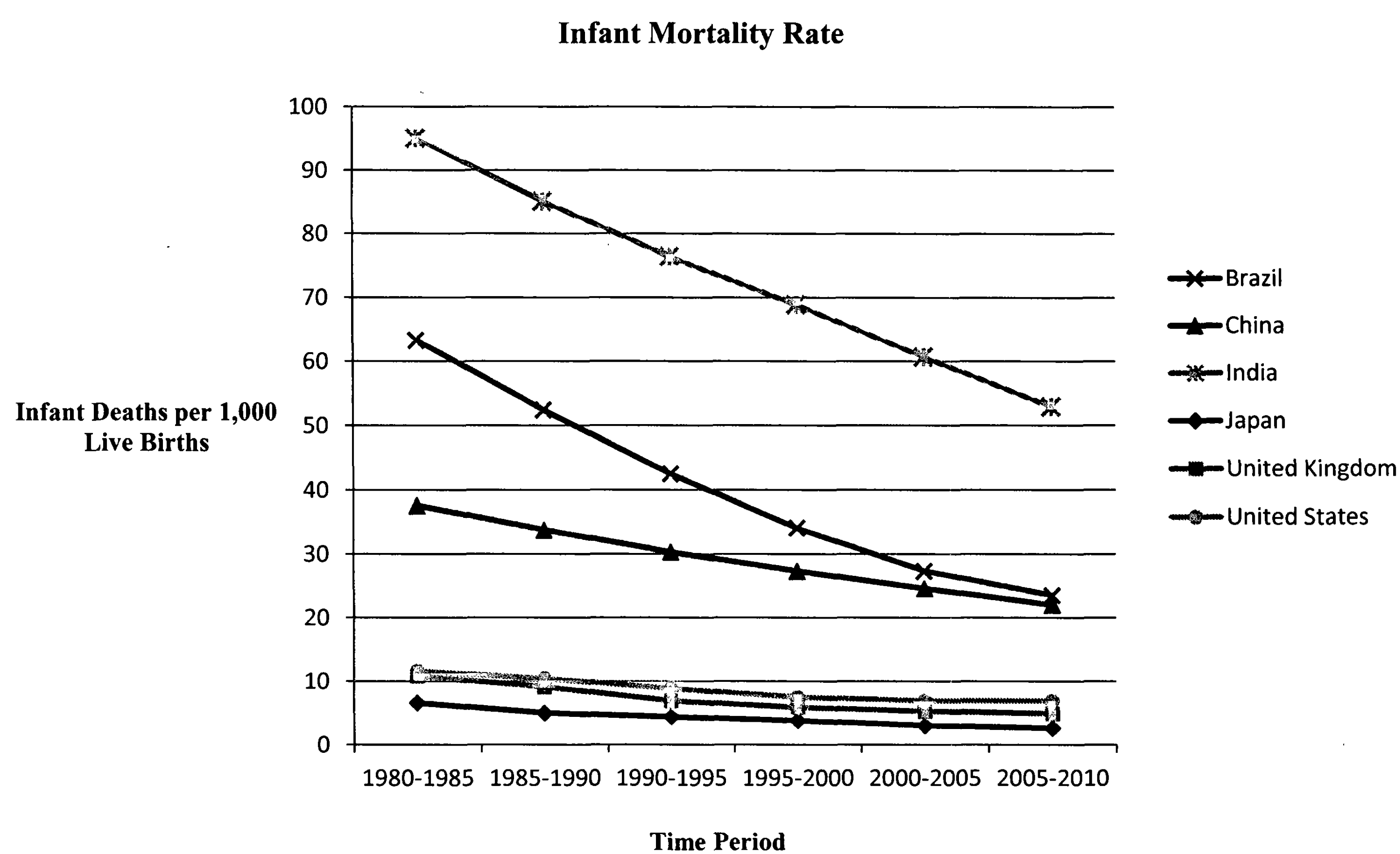
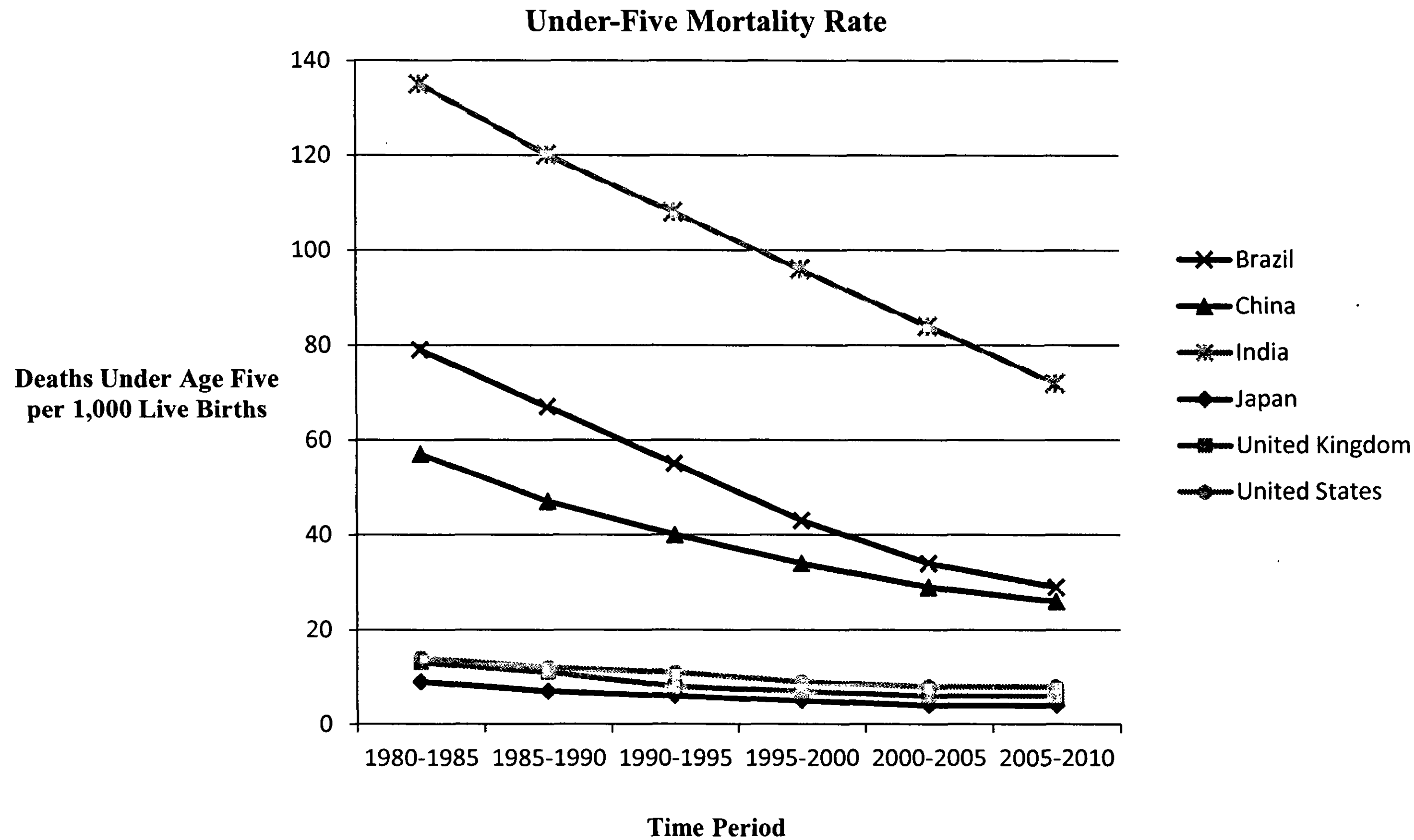
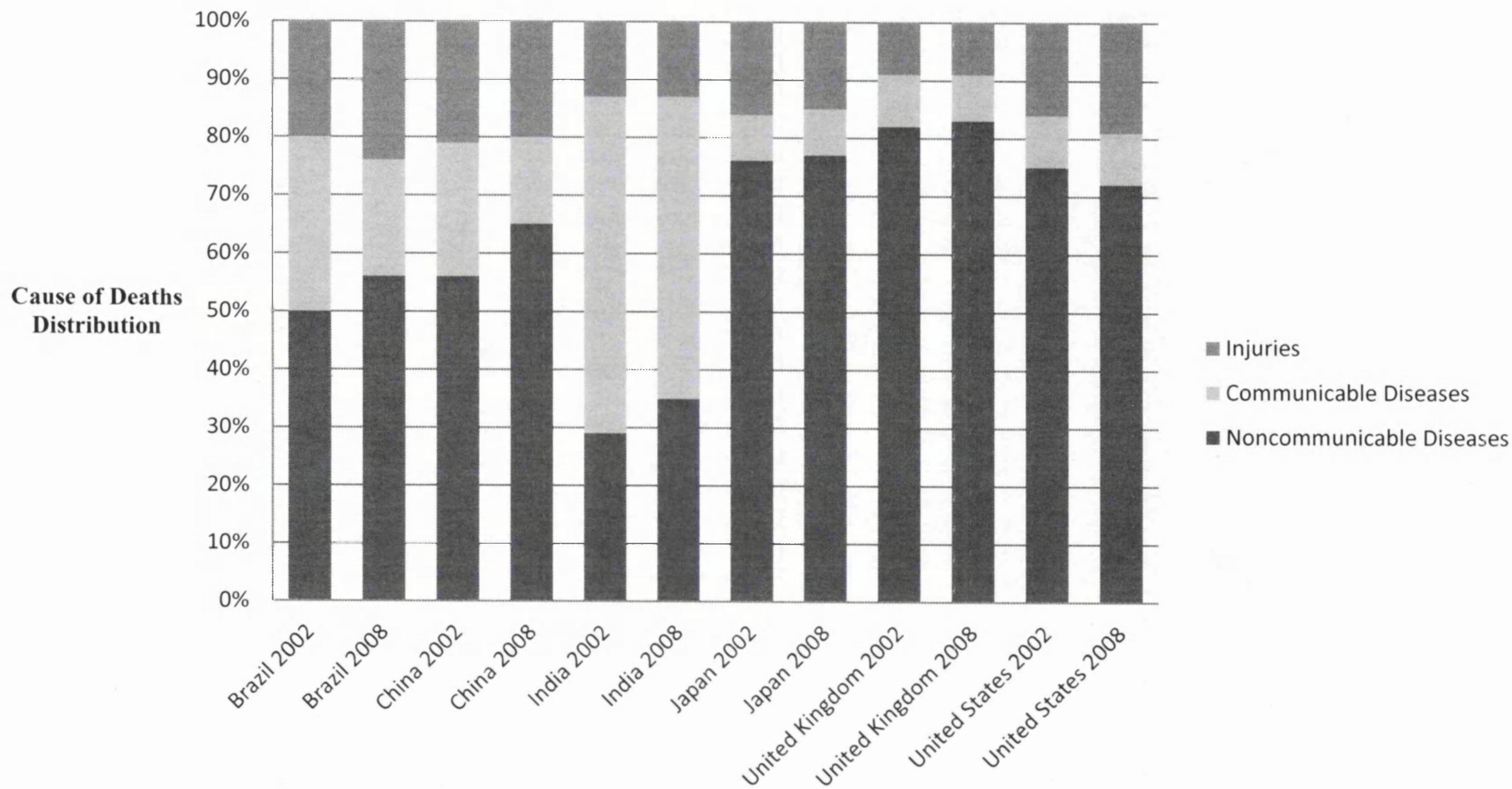


Figure 2.3 (b) Mortality Rate by Country



Sources: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat, World Population Prospects: The 2010 Revision. Available: http://esa.un.org/unpd/wpp/index.htm.

Figure 2.4 Cause of Deaths Distribution



Sources: WHO Core Health Indicators. Available: http://apps.who.int/whosis/database/core/core_select.cfm. WHO World Health Statistics. Available: http://apps.who.int/ghodata/?vid=10012#.

Table 2.1 Laid-off Workers by Year and Region

Region	**Laid-off Workers**[1] (in 10,000, at end of year)						
	1998	1999	2000	2001	2002	2003	2004
National	594.8	652.5	657.2	515.4	409.9	260.2	153.0
Northeast (selected Chinese provinces)							
Liaoning	58.9	70.8	67.2	37.0	7.4	-	-
Heilongjiang	52.8	74.1	69.9	49.1	44.2	34.1	11.5
East							
Jiangsu	24.5	17.6	14.4	8.4	1.9	0.5	0.3
Shandong	18.4	14.9	13.2	4.9	1.7	-	-
Central (selected Chinese provinces)							
Henan	25.5	28.8	30.1	25.9	25.0	17.5	8.4
Hubei	38.6	45.5	50.7	48.1	46.7	32.2	11.7
Hunan	48.1	52.5	50.0	49.1	42.1	19.5	19.5
Southwest (selected Chinese provinces)							
Guangxi	7.1	3.9	2.5	2.0	0.8	0.3	-
Guizhou	10.6	11.0	12.0	9.1	9.6	8.8	7.9

Sources: China Statistical Yearbook (2005)

[1] The number of laid-off workers in SOEs, joint state-state enterprises and companies with sole state ownership.

Table 2.2 Distribution of Health Expenditure by Country

Time Period	Brazil			China			India		
	Gov[1] (%)	Pri[2] (%)	Total (%)	Gov (%)	Pri (%)	Total (%)	Gov (%)	Pri (%)	Total (%)
1995	43.0	57.0	100	50.5	49.5	100	28.6	71.4	100
1996	40.5	59.5	100	46.6	53.4	100	27.9	72.1	100
1997	43.0	57.0	100	44.2	55.8	100	27.8	72.2	100
1998	42.6	57.4	100	41.8	58.2	100	28.4	71.6	100
1999	42.7	57.3	100	40.9	59.1	100	29.4	70.6	100
2000	40.3	59.7	100	38.3	61.7	100	27.5	72.5	100
2001	42.3	57.7	100	35.6	64.4	100	27.1	72.9	100
2002	44.6	55.4	100	35.8	64.2	100	25.0	75.0	100
2003	44.4	55.6	100	36.2	63.8	100	25.7	74.3	100
2004	47.0	53.0	100	38.0	62.0	100	22.5	77.5	100
2005	40.1	59.9	100	38.8	61.2	100	23.0	77.0	100
2006	41.7	58.3	100	40.7	59.3	100	27.5	72.5	100
2007	41.6	58.4	100	45.3	54.7	100	29.6	70.4	100
2008	44.0	56.0	100	47.3	52.7	100	32.4	67.6	100
2009	45.7	64.3	100	50.3	49.7	100	32.8	67.2	100
Time Period	**Japan**			**United Kingdom**			**United States**		
	Gov (%)	**Pri (%)**	**Total (%)**	**Gov (%)**	**Pri (%)**	**Total (%)**	**Gov (%)**	**Pri (%)**	**Total (%)**
1995	83.0	17.0	100	83.9	16.1	100	44.9	55.1	100
1996	82.8	17.2	100	82.9	17.1	100	45.0	55.0	100
1997	81.5	18.5	100	80.4	19.6	100	44.7	55.3	100
1998	80.8	19.2	100	80.4	19.6	100	43.5	56.5	100
1999	81.1	18.9	100	80.6	19.4	100	43.1	56.9	100
2000	81.3	18.7	100	79.3	20.7	100	43.2	56.8	100
2001	81.7	18.3	100	79.9	20.1	100	44.2	55.8	100
2002	81.5	18.5	100	79.9	20.1	100	44.1	55.9	100
2003	78.6	18.5	100	80.1	19.9	100	45.0	55.0	100

2004	79.4	18.3	100	81.4	18.6	100	45.4	54.6	100
2005	80.4	17.3	100	81.9	18.1	100	45.5	54.5	100
2006	79.4	18.7	100	81.9	18.1	100	46.4	53.6	100
2007	80.3	18.1	100	82.0	18.0	100	46.8	53.2	100
2008	80.5	18.0	100	82.6	17.4	100	47.8	52.2	100
2009	80.0	18.5	100	83.6	16.4	100	48.6	51.4	100

Source: WHO World Health Statistics. Available: http://apps.who.int/ghodata/?vid=10012#.

[1] 'Gov' refers to the general government expenditure on health as a percentage of total expenditure on health. It includes the resources channelled through government budgets to providers of health services, and the expenditure on health by parastatals, extra budgetary entities and the compulsory health insurance payments.

[2] 'Pri' refers to the private expenditure on health as a percentage of total expenditure on health. It includes expenditure from pooled resources with no government control, such as voluntary health insurance, and the direct payments for health by corporations (profit, non-for-profit) and households.

Chapter 3 The China Health and Nutrition Survey

3.1 Introduction

The data used in this thesis comes from a progressive International Cooperation Project carried out jointly by The University of North Carolina[18] and the Chinese Centre for Disease Control and Prevention[19] and is called the China Health and Nutrition Survey (CHNS). Popkin *et al* (2009) argue that China was selected for the project due to the opening up of its social and economic system. This project aims to investigate the labour market impact of health and nutrition status of the Chinese residents. The data include detailed information regarding demographic characteristics of residents, economic and social activities, and health-related factors. Thus, it has become an important data source on the study of the micro-behaviour of individuals in China.

The survey began in 1989. So far eight waves of data have been collected (1989, 1991, 1993, 1997, 2000, 2004, 2006 and 2009). In addition, the ninth rounds of the CHNS are proposed to be collected 2011. CHNS data has several advantages including a stratified random sampling, a large sample, and availability of panel data sets. The stratified random sampling increases precision and provides better coverage of the population and will be discussed later in this section. In addition, the survey adopts almost the same index system and consistent definition of variables across time.

The survey covered nine provinces[20] which represent various geographic and economic characteristics, public resources, health facilities and levels of health in China. These provinces include: Heilongjiang, Liaoning, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi and Guizhou. Figure 3.1 shows the map of the surveyed (dark green areas) regions. These nine provinces are distributed from the north to south of the counrty. However, Qiu (2007a) argues that the survey is not geographically representative since it fails to cover the north western provinces of China.

---

[18] Carolina Population Center, the University of North Carolina, Chapel Hill, NC, USA.

[19] National Institute of Nutrition and Food Safety, Chinese Center for Disease Control and Prevention, Beijing, China.

[20] In 1997 (the fourth wave), Heilongjiang province replaced Liaoning province. In 2000 (the fifth wave), Liaoning province returned to the survey.

As mentioned above the CHNS is an example of a multistage, random cluster sample[21]. Cities and counties in the nine provinces were divided by income level into low, middle, and high. In each province both urban and rural neighbourhoods are surveyed. In addition to that, two cities and four counties are randomly selected by employing a weighted sampling technique. Urban neighbourhoods and suburban villages (neighbourhoods) within the cities, and county towns and rural villages within counties are randomly selected (Popkin *et al.*, 2009). In principle, twenty households were picked randomly in each sampling unit (community) and all household members including children, adults and elders were surveyed. There were 216 sampling units from nine provinces involved [22], including 36 urban neighbourhoods, 36 suburban villages, 36 county towns and 108 rural villages. Currently around 4,400 households, which contain nearly 19,000 individuals, are investigated. The major areas which are covered by the survey are listed in Table 3.1, and the key features of the population surveyed in each year are summarized in Table 3.2.

As can be seen from Table 3.1, seven waves of cross-sectional data (1989-2006) are currently available.[23] For each wave of data there is a corresponding survey questionnaire of which most sections are arranged in tabular format.[24] Basically all questionnaires[25] can be divided into three major sub-surveys: Household and Individual Survey, Nutrition and Physical Exam Survey, and Community Survey.[26]

Details about demographic, financial status, labour force participation and health care facilities can be obtained from the Household and Individual Survey in all survey years, the only exception to this is the ever-married women survey which began at 1991. The first section of the Household and Individual Survey gathers the demographic information for all household members, including age/Date of Birth

[21] http://www.cpc.unc.edu/projects/china/design/survey.html [cited 6th May 2010].
[22] For the first four rounds of the CHNS (i.e. 1989-1997), 190 sampling units were involved: 32 urban neighbourhoods, 30 suburban villages, 32 county towns and 96 rural villages. Since 2000, the primary sampling units have grown to 216.
[23] For 2009, the cross-sectional dataset is not currently available but the panel dataset is.
[24] 'The questionnaire also functions as a codebook. The variable names in the data sets are usually listed to the right of or below their corresponding questions in the questionnaire.' See: http://www.cpc.unc.edu/projects/china/data/questionnaires.html [cited 6th May 2010].
[25] All the questionnaires which were originally written in Chinese version have been translated into English.
[26] Since 2004, Household and Individual Survey, and Nutrition and Physical Exam Survey started to use adult (aged 16 and above) and child (aged less than 16) questionnaires separately.

(DOB), gender, marital status, highest educational attainment and region of residence. The work activity section of the Household and Individual Survey collects adult (aged 16 and older) information about current employment status, and details regarding primary and secondary occupations, such as the type of the occupation, employment position, and size of the work unit. In the income section, income either from market activity (such as average monthly wages, annual bonuses, monthly subsidies, annual cash in-kind, annual pensions and retirement wages) or from non-market activities (such as annual income from household farming, raising poultry, fishing and home gardening) can be obtained. Moreover, the income section also includes information about working time, such as the number of months worked during the last year, the number of weekly working hours, the number of weekly working days and daily working hours.[27] The rest of the Household and Individual Survey collect information on the following, time allocation on home chores and care of children (aged 6 and below) for respondents aged 6 above, information about drinking water, environmental sanitation and household assets (including household equipment and other goods); accessibility and use of health care and medical services (for respondents aged 12 and above) and planned immunizations (for children aged 12 and below); marriage, pregnancy and birth history for ever-married women aged 52 below.

Details on household food consumption for three consecutive days are available in the nutrition section of the Nutrition and Physical Exam Survey, and dietary intakes for all household members for same three days were also surveyed. These nutrition data, as well as the physical examinations data, which include clinical measures of blood pressure (in millimetres of mercury), weight (in kilograms), height (in centimetres), head (head circumference was only surveyed for children in 1989) and upper arm circumference (both head and upper arm circumference are measured in centimetres), pulse (rate/minute), triceps skin fold (in millimetres), condition of eyes, arms and legs, are collected by a group of trained interviewers[28]. Furthermore, the

[27] In the income section, 'annual bonuses' refers to total value of bonuses received from the previous year of corresponding survey year, 'monthly subsidies' refers to the amount of subsidies received in the last month, etc.

[28] All field workers were professionally trained nutritionists. Some of the field workers have participated in other national surveys. Almost all interviews were graduates of post-secondary schools, and many of them were graduates of universities. Moreover, CHNS survey also provides 3 days of specific training in the collection of dietary data. Available: http://www.cpc.unc.edu/projects/china/design/datacoll.html [cited 3th May 2010].

survey asks details on current health status, smoking and alcohol drinking history, tea/coffee drinking history, injury and disease history for all household members. Current health status includes self-reported health status (SRHS), and physical activities (for children and adults separately). Details of SRHS, smoking and alcohol drinking history are discussed later in this section. The section of physical activities collects information on the type of and the time spent on participated physical activities and exercises. Since 1993, it also asks information about and life disruptions due to physical limitations or poor health for adults older than 50. The Community Survey contains details on: infrastructure such as transport, water and electricity; services such as health facilities and family planning; price of foods and living materials.

Table 3.2 shows the number of households and individuals included in the survey between 1989 and 2006. New members include the following:

i. Newly-formed households members – members who belonged to any of the households surveyed in last round, left the household after the last interview, then formed a new household or joined another household (still in the same neighbourhood/village)
ii. Members in replacement households (began on survey 1997) – there are two types of replacement households, added and substitute. If there are fewer than twenty households in a survey site, then a certain number of households are added to make up twenty. These households are called 'added households'. Another case is, if a few households cannot be contacted, then the same numbers of households at the same site are selected to replace those which have lost contact. Those households are called 'substitute households'.
iii. Members in new households and new sites. This is the case of 1997 when Liaoning was replaced by Heilongjiang (new province).

Old members include:

i. Members in previous households surveyed in previous years.
ii. New household members in old households– current family member who did not belong to the household in last wave. The most likely reason is being newborn, marriage and so on.

Clearly, it can be seen from Table 3.2 that the proportion of the new members increases sharply in 1997, accounting for 26.1% in this year. The main cause for this rise is because, in 1997, Liaoning Provinces was hit by floods; therefore it was replaced by Heilongjiang. In 2000 Liaoning rejoined the survey, hence the proportion of the new members falls back to 10.9%. It is not possible to follow up every individual in the survey due, for example, to death or inability to contact. The number of deaths is presented in Table 3.3. It grew steadily and more than doubled to 337 in 2006. A number of individuals were lost due to travel, relocation or immigration, or people who refuse to participate or were unable to answer certain questions. Table 3.2 also shows the sample of individuals in different age groups. Overall people who are of working age account for more than 60% of the individuals in each wave. The proportion of people under age 16 decreases over time from 26% in 1989 to 15.2% in 2006 and the proportion over the retirement age increases from 9.6% in 1989 to 13.4% in 2006.

As previously mentioned, the CHNS is a panel survey in that the same individuals are interviewed over time. In addition to cross-sectional versions of the annual data the data are also available in longitudinal versions which are drawn from the original cross-sectional datasets. So far 41 longitudinal datasets have been created which are listed in Table 3.4. Basically, all of these master longitudinal files are able to cover seven waves of the CHNS data except for the Infant Feeding File, Children Energy Expenses File and all Ever-Married Women Files. Jones (2007) states that panel data is an extension of the cross-sectional datasets which allows a further investigation of changes in individual behaviours over time. Thus, for instance, analysis of how the individuals' health conditions change over time can be performed by using the Master Physical Exam (PE) File which was created from the seven years of the Cross-Sectional Physical Exam Files. In addition, demographic information can be gathered from Roster File and PE File jointly. Information about individual occupation and income can be collected within Income Files.

3.2 Data and Variables

This thesis uses eight original longitudinal datasets, which are listed in Table 3.4, to create the merged CHNS longitudinal dataset. These original longitudinal datasets are Master Roster File, Individual Daily Nutrient Intakes Master File, Master

Physical Exam (PE) File, Individual Education Master File, Individual Occupation Master File, Individual Wage and Bonus Master File, Individual Income Master File and Individual Medical Services File. Table 3.5 provides descriptive statistics for demographics, current health status, anthropometric measurement information, nutrition and health-related behaviour (alcohol and tobacco consumption), namely the means, standard deviations, minimum and maximum values pooled across all waves, using the merged CHNS longitudinal dataset. Since the purpose of this research is examine the impact of health on employment, wages, and hours worked between men and women in China, this study based on adult samples, that is, individuals of working-age[29] (16-59 for male, 16-49 for female) who are non full-time students because students might not have established work patterns (Haveman *et al.*, 1994), and screens out the observations who have missing values for general demographics[30]. This study also excludes observations that have implausible or inconsistent information, thereby increasing the estimate precision (Tuan *et al.*, 2008). For instance, only individuals with plausible anthropometric information are kept for further analysis, that is, a waist circumference of 45-150 cm and a BMI of 15-40 kg/m^2 (Tuan *et al.*, 2008). Less than 0.6% of sample is excluded. According to Meng *et al.* (2009), daily calorie intake below 800 and over 10,000 should be treated to outliers, and these are also excluded (less than 0.3% of sample). Additionally this study drops outliers with nominal hourly wage income above 50 Chinese yuan which then exclude less than 0.1% of the working sample. If individuals respond that they are presently not working but have positive response on wage, working hours, occupation, employment position, type of work unit, or the size of the work unit, then these individuals (approximately 2.3% of sample), which are treated as inconsistent information, are also excluded from the sample.

3.2.1 Dependent Variables

---

[29] According to the 15th regulation in China Labour Law, the state prohibits hiring people under 16. In addition, basis "Provisional Measures on the workers retire or resign by state council", the legal retirement age for male is 60 and for female is 50 (for female who working at state organizations and institutions is 55). Figure 3.2 shows that the employment rate for males drops quickly after 59, and the employment rate for females drops quickly after 49.

[30] In this study, general demographic information is considered as information on gender, age, marital status, area of residence, region of residence, educational attainment and current working status (presently working or not working).

The dependent variables in the thesis include being in employment and logarithm of hourly income from wages[31]. The CHNS asks respondents '*Are you presently working?*'.[32] Accordingly, the employment variable is a dummy variable (*employed*) taking the value one if the individual is presently working[33] and zero if the individual is unemployed or not in the labour force.[34]. As there is no direct measure of hourly wage income, this study defines the hourly income as the ratio between the total annual wage income and total annual working hours. The total annual wage income is the sum of wage and bonuses received during the last year, that is, average monthly wage times the number of months worked, plus annual bonuses[35]. The annual working hours are calculated by multiply 4.3 weeks per month by weekly working hours, then times months worked. The weekly working hours can be obtained from the CHNS either directly (direct measure of hours worked in the last week) or indirectly (multiply average number of days worked in a week by average numbers of hours worked in a day). The average weekly working hours[36] is defined as the ratio between the total annual working hours and total number of weeks per year. As presented in Table 3.5, a large proportion (86%) of people in this sample report that they are presently working. The mean nominal hourly wage income and mean real hourly wage income for the Chinese adults (pooled all eight waves) are approximately 3.08 yuan and 3.73 yuan, respectively.[37]

Figures 3.3 and 3.4 present the distribution of the employment rate and real wage income by gender and wave, respectively. Figure 3.3 shows that the employment rate of Chinese adults decreases from 97% in 1989 to 76% in 2009, especially between 1997 and 2004 (that is, decreases from 88% to 74%), although it increases slightly

---

[31] Although the CHNS survey both primary and secondary occupations of the respondents, this study only focuses on wage income from respondent's primary occupation. 11% of the total working sample (total working sample, N=50,897) have secondary job, in addition, about 2% of the total working sample response their earnings from the secondary job.

[32] Presently indicates a short period of time before and after the interview.

[33] Employment, which is based on the International Labour Organization (ILO) definition of employment, includes workers who are self-employed.

[34] The definition of non-employed is similar to that in Jones (2011), for details, see footnote 5 in Jones (2011), p. 1004.

[35] In CHNS 1991, 1993 and 1997, information on annual bonuses was asked in the welfare subsidies section. After CHNS 1997, it was asked in the section of income from wages. According to the CHNS questionnaires, bonus refers to regular monthly bonus, holiday bonus, dividend bonus, bonus for increase in production, and any other bonuses from the work unit in the last year.

[36] In this study, less than 0.4% of the total working sample work less than twelve months.

[37] The official currency of China is Renminbi (RMB). Yuan is the Chinese currency unit. 1 yuan equals to 0.1 pound approximately.

after 2004. In Figure 3.4, there is a continued growth trend in average real wages from 1998 to 2009 except for 1991. The average real wages of the whole sample increase from 1.8 Chinese yuan in 1989 to 2.2 Chinese yuan in 1993, then increased to 3.5 Chinese yuan in 1997 and to 9.3 Chinese yuan in 2009. Note that the average weekly working hours for the first three waves (which are around 47-49 hours) are higher than those for the subsequent waves (which are around 40-45 hours)[38]. The Double Cease Day[39] has just been performed since 1995. Therefore, the number of days off for Chinese workers is one per week (i.e. 48 weekly standard working hours) for the first three waves, and increased to two days for the subsequent waves (i.e. 40 weekly standard working hours). As can be seen from Figures 3.3 and 3.4, the employment rate and real hourly wages for both genders follow the same trend as those for the total sample, and together are higher for men than for women.

It is important to note that 50,823 individuals are presently working, however, only 41% of them respond to questions on their income from wages (and 64% of them give information on working hours).[40] The possible reasons for such low response rate of wage income (and of working hours) are discussed in the next subsection.

3.2.2 Health and Health Related Variables

Self-Reported Health Status (SRHS)

The survey asks respondents '*Right now, how would you describe your health compare to that of other people of your age?*', and four possible responses are: *excellent*, *good*, *fair* and *poor*. The SRHS question is not available for waves 1989 and 2009. Based on the above response of the SRHS, four dummy variables, which take value of one if an individual has excellent health, has good health, has fair health or has poor health, are created. Table 3.5 shows that 60% of respondents report their current health status as good, 21% of and 16% of those are in fair and excellent health, respectively. Only the minority (3%) report their health status as poor. Figure 3.5 describes the distribution of the SRHS for the whole sample across

[38] The standard working hours for workers and staff in state organizations, social organization, enterprises and institutions and other organizations are eight hours a day.
[39] See: http://law.51labour.com/lawshow-50779.html. [cited 22nd Feb 2011, Chinese version].
[40] The response rate of hourly income is the ratio between the number of people gives information on income from wages and the number of people who are presently working, i.e. (20,744/50,823)*100%. The response rate of working hours is the ratio between the number of people gives information on weekly hours and the number of people who are presently working, i.e. (32,317/50,823)*100%.

six waves (1991-2006).[41] It shows that the majority of respondents have good health status for each survey year, and the percentage of the excellent health status and of the poor health status changes very little across six waves. However, there is a trend that the distribution of the SRHS becomes worse after wave 1997, which is found to be consistent with the trend found in Qiu (2007a). It is worthwhile to note that the position of SRHS question in the CHNS was changed after wave 1997. For wave 1991, 1993 and 1997, the SRHS question was listed in household questionnaire, whereas the following waves, it was listed in the physical examination survey questionnaire. The change of the position of SRHS question may makes individuals likely to evaluate their health from different perspectives. Therefore, the change of the SRHS after wave 2000 might not truly reflect a change of health.

Anthropometric Measurement

Height (barefoot; in meters) is measured by two professional by trained measurers who use height scale, and the average of two measurements is taken. The tallest individual in the whole sample is 1.90 meters and the shortest is 1.28 meters. The average height is 1.62 meters. When measuring weight, the respondent should wear only shorts and a shirt. Similar to the height measurement, two measurers who use a physician scale perform the weight measurement and take the average of two measurements. Weight is given in kilograms, and the mean value of the whole sample is approximately 59 kilograms.

BMI (kg/m^2) is a widely used measure of an individual's overweight/obesity and is defined as body weight in kilograms divided by the square of height in meters. According to the WHO on the International Classification of BMI, individuals with BMI between 18.5 and 25 are classified as healthy weight, while a BMI below this range is classified as underweight, those with BMI greater 25 is classified as overweight, and above 30 is classified as obese. Therefore, this study creates three dummy variables in accordance with the WHO BMI Classification, including *uweight_who* (BMI<18.5), *nweight_who* (BMI 18.5-25), and *oweight_who* (BMI>=25). Under the WHO BMI Classification, 75% of the whole sample is in the normal weight category, followed by overweight category (18%) and the

[41] The distribution of the SRHS for the working sample (i.e. dummy variable *employed* equals one) is similar to the distribution showed in Figure 3.5.

underweight category (7%). Figure 3.6 shows the distributions of WHO BMI Classification for the whole sample across eight waves. Between wave 1993 and 2009, the proportion of *oweight_who* increases significantly and is matched by a decrease on both the proportion of *nweight_who* and of *uweight_who*. A similar trend can be found in Zhai *et al.* (2009).

There are many studies which suggest optimal cutoffs of BMI and WC for overweight using Chinese population.[42] Yang *et al.* (2006), which use a cross-sectional sample of 1,109 males and 879 females aged 20-45 in central south China[43], examine the utilities of three anthropometric indices which involve BMI, Waist Circumference (WC, in centimeters) and Waist-to-Hip Ratio (WHR), by employing a Receiver Operating Characteristic (ROC) curves analysis.[44] They suggest that lower cutoffs of anthropometric indices for the diagnosis of health-related obesity should be applied to the Chinese population. Moreover, they propose that BMI and WC are better than WHR because of their high accuracy of predicting obesity. Similar conclusion can be found in Misra *et al.* (2005).

Bao *et al.* (2008) use BMI as a predictor of the metabolic syndrome (Mets) incidence, and study the optimal cutoffs of BMI for eastern Chinese population[45]. Similar to Yang *et al.* (2006), the authors also employ the ROC analysis to assess the usefulness of BMI associated with abdominal visceral obesity. The results show that the optimal cutoffs of BMI for male and female Chinese adults are 24.5 and 24.7, respectively. Based on optimal cutoffs of BMI for Chinese population proposed by Bao *et al.* (2008), a dummy variable, *oweight_bao*, is created. Table 3.5 shows that 21% of individuals are in the overweight category.

Tuan *et al.* (2008) define the optimal cutoffs of BMI for Chinese population by the increased risk of hypertension disease. The sample which includes 2,077 men and

---

[42] Qiu (2007a) adopts an adjusted measure of BMI in which gender and age effect are take into the consideration, and proposes a different healthy range (BMI 15-19) by using CHNS panel data. Qiu (2007a) states that comparing to the WHO healthy range (BMI 18.5-25) which is relevant to the obesity-related disease, the adjusted healthy BMI range is substantially different since it is relevant to SRHS.

[43] Yang *et al.* (2006) argue that their sample is not representative for the entire country since it only covers southern Chinese.

[44] WC and WHR are two most common approaches on measuring abdominal obesity (Misra *et al.*, 2005).

[45] Bao *et al.* (2008) also criticize that their sample might not represent the entire Chinese because the living standard and life styles are diverse in China.

2,415 non-pregnant or non-lactating women (aged 18-65) comes from a two-year CHNS panel data (2000 and 2004). The results based on the ROC analysis conclude that a value of 23 is an optimal BMI cutoff for Chinese adults. Specifically, a BMI of 22.5 and 23.5 are found to be the optimal BMI cutoffs for Chinese men and women, respectively. Therefore, a dummy variable *oweight_tuan* based on optimal cutoff of BMI proposed by Tuan *et al.* (2008) is created. Based on the optimal BMI cutoffs proposed by Tuan *et al.* (2008), 36% of individuals are in the overweight category. Figure 3.7 further describes the distribution of the three cutoffs of BMI on obesity (*oweight_who*, *oweight_bao* and *oweight_tuan*) for the whole sample across eight waves, respectively. As shown in Figure 3.7, both '*oweight_bao*' and '*oweight_tuan*' follow the same trend as '*oweight_who*'. Overall the proportion of '*oweight_bao*' is higher than that of '*oweight_who*', and the proportion of '*oweight_tuan*' is twice as much as the proportion of '*oweight_who*'.

Individual Daily Nutrient Intakes

The individual daily nutrient intakes include 3-day average intakes of energy (in kilocalories), and macronutrients which consist of protein (in grams), fat (in grams) and carbohydrate (in grams). This information is collected from the longitudinal Individual Daily Nutrient Intakes Master File. Four nutrition variables (continuous variables) are created in this study: *energy*, *protein*, *fat* and *carbohydrate*. Table 3.5 shows that available sample for each nutrition measurement contains nearly 51,700 individuals. In addition, Figure 3.8 describes the distribution of individual daily nutrient intakes (all in calories) for the whole sample across eight waves. It shows that the energy intake decrease from 2,970 kcal to 2,250 kcal, mainly due to the decrease of the carbohydrate intake. Furthermore, Figure 3.9 describes the distribution of the shares of calories from protein, fat and carbohydrate intakes across eight waves. *Zhai et al.* (2009) finds that, between wave 1989 and wave 2004, the share of calories from fat for the Chinese adults increases dramatically from 19% to 28%. However, the shares of calories from each macronutrient intakes only change slightly between wave 1989 and wave 2009 in the CHNS.

According to the human energy requirements proposed by FAO/WHO/UNU (2001), daily energy requirement is related to gender, age and body weight, which is listed in Table 3.6. For instance, the range of daily average energy requirement for a man aged 20 and weighing 60kg is between 2,754 kcal and 3,224 kcal, when he participates in moderate physical activity. Furthermore, according to the dietary intake recommendations proposed by WHO/FAO (2003), the healthy protein intake is between 10% and 15% of total energy, the healthy fat intake is between 15% and 30% of total energy, and the healthy carbohydrate intake is between 55% and 75% of total energy. The average daily energy intake and the share of calories from the three macronutrients (by gender and age) for the Chinese adults (the whole sample and the working sample) are presented in Table 3.7. The average daily energy intake for men aged less than 30 and aged 30-59 are 2,768 kcal and 2,642 kcal, respectively. The average daily energy intake for women aged less than 30 and aged 30-49 are 2,351 kcal and 2,323 kcal, respectively, which are lower than those for men in each corresponding age group. Comparing with the whole sample, the working sample has slightly higher energy intake in each corresponding gender-age group. Both gender-age groups in the whole sample and in the working sample have the same shares of calories from protein intake (12%). The shares of calories from fat intake for each gender-age groups in the whole sample (24%-27%) are similar to those in the working sample (24%-26%). The shares of calories from carbohydrate for each gender-age groups in the working sample (60%-63% across all gender-age groups) are slightly lower than those in the whole sample (61%-64% across all gender-age groups).

On the basis of the energy requirement range proposed by FAO/WHO/UNU (2001), this study creates three dummy variables: *lenergy* which takes value of one if individual daily average energy intake is below the FAO/WHO/UNU criteria; *nerengy* which takes value of one if individual daily average energy intake is within the FAO/WHO/UNU criteria; and *henergy* which takes value of one if individual daily average energy intake is above the FAO/WHO/UNU criteria. Similarly, according to the WHO/FAO criteria, nine dummy variables which reflect the levels of the three macronutrient intakes are created (*lprotein*, *nprotein*, *hprotein*, *lfat*, *nfat*, *hfat*, *lcarbohydrate*, *ncarbohydrate*, and *hcarbohydrate*). Table 3.5 shows that the number of individuals whose energy intake is below the FAO/WHO/UNU criteria accounts for more than half of the individuals whose weight information and energy intake information are both available. In addition, the number of individuals whose share of calories from protein, fat and carbohydrate does not meet the WHO/FAO criteria accounts for 22%, 17% and 29%, respectively.

Alcohol Consumption

The CHNS asks respondents whether or not they drank beer or any other alcoholic beverage during the past year. Those who answer in the affirmative are then asked information about their drinking frequency during the same period. Five possible responses are: '*almost every day*', '*3-4 times a week*', '*once or twice a week*', '*once or twice a month*' and '*no more than once a month*'.[46] According to these two self-reported drinking questions in the CHNS, two discrete indicators, which are similar to the alcohol consumption measures in Tekin (2002), are constructed. One of the indicators (*drink*) is constructed using the first survey question of alcohol consumption. It is a dummy variable which takes on the value of one if an individual consuming any alcoholic beverage during the previous year, and zero otherwise. The other indicator are constructed by combining the responses of the two questions above, that is, six dummy variables which are named '*drinkdaily*', '*drink3to4w*', '*drink1to2w*', '*drink1to2m*', '*drink0to1m*' and '*nodrink*', respectively. As shown in Table 3.5, about 60% of respondents report that they did not consume any alcohol during the last year. 10% of respondents were daily drinkers, and the similar number of individuals drank once or twice a week. A slightly lower proportion, 8% of respondents, drank once or twice a month, and 6% drank three to four times a week. Only about 4% of respondents who drink alcohol, drink less than once a month.

Smoking

The CHNS asks respondents whether or not they have ever smoked cigarettes, including rolled or manufactured. Those who answered in the affirmative are then asked whether or not they are smoking currently and their daily cigarette consumption.[47] Combining the answers to these three smoking questions, three smoking variables are created. The first variable, *eversmoked*, a dummy variable which takes the value of one if the respondent has ever smoked cigarettes, and zero

[46] For wave 1991, 1993 and 2000, the alcohol consumption questions were listed in the physical examination survey questionnaire. For other three waves, they were listed either in the household survey questionnaire or in the adult survey questionnaire.
[47] For wave 1991, 1993 and 2000, the cigarettes consumption questions were listed in the physical examination survey questionnaire. For other three waves, they were listed either in the household survey questionnaire or in the adult survey questionnaire.

otherwise. The second variable, *smokenow*, a dummy which takes value of one if the respondent is a current smoker, and zero otherwise. The last variable, *dailysmokeamt*, a continuous variable indicates the number of cigarettes being smoked per day. Table 3.5 shows that 35% of individuals have smoked, and 96% of them are still consuming cigarettes. On average, an individual consumes approximately 5 cigarettes with a minimum zero value indicates non-smokers and a maximum 60 cigarettes. For individuals who are current smokers, around 15 cigarettes are consumed daily on average, and this number is consistent with the descriptive statistics in Pan and Hu (2008). Pan and Hu (2008) use a five-year CHNS panel data (1991-2004) to examine the changes in both the prevalence and frequency of cigarette smoking in China. In their study, the prevalence of smoking is measured by the probability of smoking, and the frequency of smoking is measured by the number of cigarettes smoked daily. They find that there is a slight decreasing trend in the prevalence of smoking, and a slight increasing trend in the frequency of smoking in China. They also find that: i) individuals who are older, male, residents in urban area, currently working, having a second job, having a higher number of smokers in the household, and living in a smaller community size are found to be associated with a higher probability of smoking, whereas living in a larger household size seems to contribute to the decrease in the probability of smoking; ii) being older, male, currently working, having a second job, and a higher number of smokers in the household have positive impacts on the frequency of smoking, conversely, urban residence, a larger household size, a larger community size and a higher average community daily wages, are negative predictors of an increase in smoking frequency.

3.2.3 Other Explanatory Variables

The demographic variables utilized in this study include gender, age, area of residence (urban/rural), region of residence (province), marital status, number of children in each household, educational attainment and occupation.

As can be seen from Table 3.5, 55% of sample is male and 45% of is female. It includes linear term of age (*age*), and also includes four age group (*age16_25*, *age26_35*, *age36_45* and *age46_59*). The average age of the sample is approximately 36 years old and the distribution among four age categories is relatively uniform. Since the sample of CHNS data is collected in four types of residential areas which including urban neighbourhoods and suburban villages within the cities, and county towns and rural villages within the counties, an area of residence variable[48] (*urban*) can be created, that is, a dummy variable which takes the value of one if the respondent is a resident in urban neighbourhood or county town areas, and zero otherwise. 70% of respondents live in rural area which is consistent with the overall rural-urban population distribution in China (Liu *et al.*, 2008).

The nine dummy variables which capture region of residence are *liaoning*, *heilongjiang*, *jiangsu*, *shandong*, *henan*, *hubei*, *hunan*, *guangxi* and *guizhou*, respectively. Liaoning and Heilongjiang are two north eastern provinces of China[49], with an economy based mainly on heavy industry; Jiangsu and Shandong are two eastern coastal provinces of China in which economic reform commenced early; Henan, Hubei and Hunan are located in central region of China; Guangxi and Guizhou are located in south western part of China in which economic progress is comparatively slow (Qiu, 2007b). Table 3.5 shows that respondents living in each of nine provinces are well distributed. Table 3.8 lists the Gross Domestic Product (GDP) and *per capita* GDP by region of residence from wave 1997 to wave 2009. As developed coastal province in China, Jiangsu has the highest GDP and *per capita* GDP in the nine provinces for all five waves. In contrast, the GDP and *per capita* GDP in Guizhou is the lowest. Shandong has the second highest GDP for all five waves and the second highest *per capita* GDP for wave 2004, 2006 and 2009. Among the rest of the provinces, Hunan and Liaoning have relatively high GDP and *per capita* GDP, respectively, for all five waves.

This study further investigates the distribution of the SRHS across gender, area of residence and region of residence which is shown in Table 3.9. The second row of Table 3.9 shows that Jiangsu and Shandong have relatively small proportions of the poor health status (2% and 1%, respectively). On the other hand, Liaoning and Hubei have relatively large proportions of the poor health status (both accounts for 4%). Heilongjiang and Shandong have the highest percentages of the excellent health (29% and 27%, respectively) and lowest percentages of the fair health (18% and

[48] The area of residence variable is created by using the original sampling-unit variables of the CHNS data which are also used to construct the household ID variables.
[49] The Three Northeast Provinces of China which refer to Northeast China consist of Liaoning, Jilin and Heilongjiang. Since the Northeast is one of the earliest regions to develop heavy industry in China, therefore the urbanization level of this region are relatively high comparing with most parts of China.

15%, respectively) among nine provinces. Although Guangxi and Guizhou have the smallest proportions of the excellent health status (5% and 10%, respectively), the proportions of the good health status for these two provinces are relatively small among nine provinces. According to the third and the fourth rows of Table 3.9, there is only a small difference between the percentage of excellent health for men and for women in Heilongjiang and Hunan, a small difference between the percentage of good health for men and for women in Hubei and Guangxi, and a small difference between the percentage of fair health for men and for women in Liaoning and Guangxi. Both sexes show the same average percentage of good health (60%) and of poor health (3%). In addition, the difference between the average percentage of excellent health for men and for women, and the difference between the average percentage of fair health for men and for women are both very small. The fifth and the sixth rows present the distribution of SRHS by area of residence and region of residence. In Shandong, there are large urban and rural differences in the proportion of good health and of excellent health. There are also obvious urban and rural differences in the proportion of excellent health in Liaoning and of good health in Hubei. Both areas show the same average proportion of fair health (21%) and of poor health (3%). The average proportion of excellent health in urban areas is higher than that in the rural areas, whereas the average percentage of good health in urban areas is lower than that in the rural areas.

In terms of other control variables, marital status is captured by a dummy variable, *married*, which takes the value of one if respondents are married, and zero otherwise (refers to people whose current marital status are single, divorced, widowed or separated). Table 3.5 shows that three quarters of respondents are married. Since there is no direct measure on how many children an adult has in the CHNS, instead, this is calculated using the total number of children per household (*nchildrenhh*) by summing all respondents who are under 16 years old in each household. It also shows that each adult has less than one child on average[50] with a maximum of six children. However, note that '*nchildrenhh*' does not necessarily mean the child dependent on the adult individual. In other words, a value of one on *nchildrenhh* indicates that there is one child within that household, while it does not necessarily mean that every adult household member have an obligation to take care of the child.

[50] The One-Child Policy was implemented in 1979.

This paper further creates three binary indicators on the presence of children in the household: a binary indicator *child0* which takes value of one if there are no children in the household, and zero otherwise; a binary indicator *child1* which takes value of one if there is one child in the household, and zero otherwise; a binary indicator *child2m* which takes value of one if there are two or more children in the household, and zero otherwise. As shown in Table 3.5, 42% of observations have no children in their household, 36% of observations have one child in their household, and 22% of observations have two or more children in their household[51]. Li and Wu (2011) use CHNS panel data to examine the intrahousehold resource allocation and bargaining power in China. Based on a culture of son preference in China, they find that the nutrition intake (including protein and calorie intake) for a woman can be improved by having a first-born son. In addition, the probability of being underweight (a BMI less than 20) for a woman also can be reduced by having a first-born son.

Six binary indicators controls for the highest educational level achieved include *noedu*, *primary*, *lowermid*, *uppermid*, *techvoc* and *university*. According to Table 3.5, 15% of respondents do not attain any academic qualification. The proportion of respondents who obtain the highest qualification on lower middle school is relatively larger than others, which accounts for 39%. Only 4% of people have bachelor degree or above. Using a similar working age sample for the 1991, 1993 and 1997 panels of CHNS, Liu *et al.* (2008) find an even smaller proportion of individuals in '*university*' (3%).

For those in work, nine categories are used to describe the occupation of respondents. The first two categories are senior professional/technical personnel (such as doctor, lawyer, professor, engineer and architect) and junior professional/technical personnel (such as photographer, nurse, teacher, midwife and editor), respectively. The third category includes police officer, army officer, factory manager, administrator/executive/manager, section chief, government official, director and administrative staff. The fourth category includes office staff such as secretary and office helper. The fifth category includes farmer, fisherman, logger and hunter. The sixth and seventh categories are technical/skilled worker (such as craftsman and foreman) and non-technical/non-skilled worker (such as labourers). The eighth

[51] Among the observations who have two or more children in the household, 80% of them have two children in the household.

category includes driver, and service worker such as counter sales, waiter, cook, housekeeper, doorkeeper, barber/beautician, childcare and launderer. And the final category includes athlete, handicraft, musician, actor, policeman ordinary soldier, engagement in small commercial household business and other. Accordingly, nine binary variables are constructed using the above categories (*occup_1*, *occup_2*, *occup_3*, *occup_4*, *occup_5*, *occup_6*, *occup_7*, *occup_8* and *occup_9*). As shown in Table 3.5, agricultural occupation category (*occup_5*) accounts for half of the total occupation categories, technical, non-technical and service occupation category account for 8%, 12% and 11%, respectively, of the total. For the rest of the occupation categories, each represent 5% or less of the total.

This study utilizes three dummy variables to describe the employment status among those in work, which are *selfemp*, *employee*, and *otherposi*, respectively. The first dummy variable (*selfemp*) equals to one if individuals report their employment position as self-employed (such as independent operator with no employees and owner-manager with employees). The second variable (*employee*) takes on the value of one if individuals are paid family works, or work for another person or enterprise. The last dummy variable (*otherposi*) takes value of one if individuals are unpaid family workers, temporary workers or other. As shown in Table 3.5, almost 56% of respondents are self-employed and 36% are '*employee*', only minority report as '*otherposi*' which account for 8%.

The type of work unit is described using three binary indicators: *privatesec*, *government*, and *typeother*. In Table 3.5 there is a large proportion (63%) of respondents work in an individual or private enterprise, three source invested enterprise, household business, or family contract. Working in state enterprise or government units accounts for 21% of those in work. As mentioned in the previous subsection, 50,823 individuals are presently working, while only 40% and less than 63% of respondents give information on income from wages and on working hours, respectively. Similarly, there is a low response rate of the size of the work unit (59%)[52] comparing with that of other explanatory variables. Therefore, this study defines four dummy variables, i.e. *firm020*, *firm20100*, *firmmt100*, and *firmunk*, to

[52] The response rate of the size of the work unit is the ratio between the number of people who give information on the size of the work unit and the number of people who are presently working, i.e. (28,405/50,897)*100%.

measure the total number of employees within the work unit. The dummy variable, *frimunk*, is used to control for individuals who have missing value on the size of their work unit. In addition, a dummy variable, *secjob*, takes on the value one if an individual has a second job. As can be seen from Table 3.5, only a small number (10% of the sample) of people have second job.

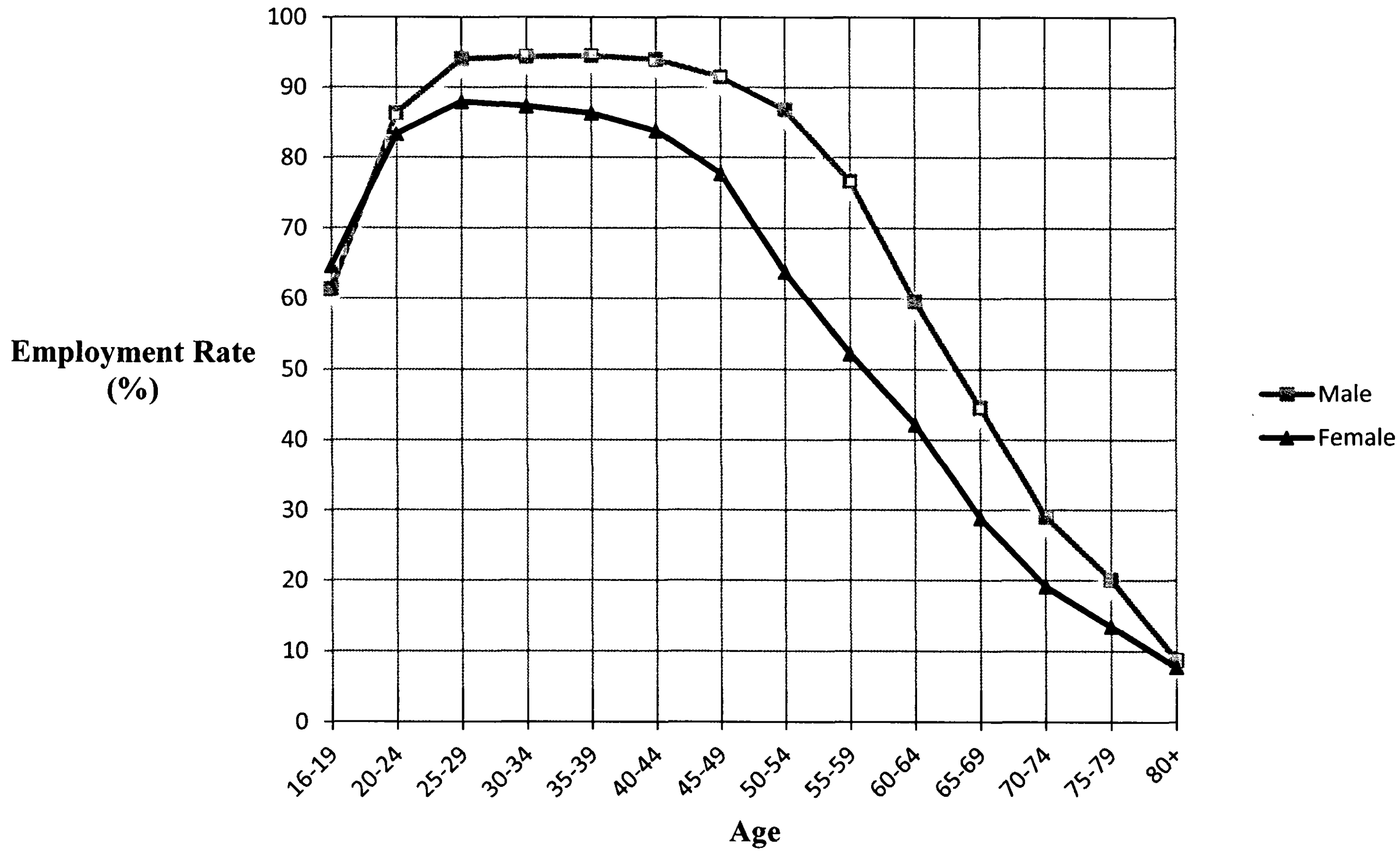
Table 3.10 presents the response rates for earnings, working hours and firm size across the nine occupation categories. The column (a) indicates the proportion employed in each occupation category; column (b) indicates the response rate of the average monthly earnings in each category; column (c) indicates the response rate of the weekly hours worked in each category; column (d) indicates the response rate of the hourly income in each category; and column (e) indicates the response rate of the size of the work unit in each category. Similarly, Table 3.11 shows these response rates across the three employment position categories. From Table 3.10 it is evident that half of the sample engages in what broadly can be described as agricultural type work, however, only 3% of them give their information on hourly income due to the low response rate on both monthly earnings (3%) and weekly working hours. Moreover, just 30% report their size of the work unit. A possible reason for the low response rate on wage income, working hours and the size of work unit is that people who engage in agricultural work are self-employed (94% of agricultural workers are self-employed in this study), therefore they might either be unwilling to respond or have difficulty in calculating their monthly earnings, and working hours. In studying the impact of health on wages in Brazil, Thomas and Strauss (1997) explain the reason for choosing the urban sample only, is because, there are many households operate farms in the rural area and most of them are family enterprises which contain more than one household member, implying the difficulty in gathering individual income information. Also note that only 35% of respondents whose occupation are engagement in small commercial household business, handicraft, athletes or other, give their information on hourly wage income due to the low response rate on monthly earnings (36%). Since nearly half of individuals in the above occupation category are self-employed, they might have difficulty on calculating their monthly earnings, working hours or the size of the work unit.

Interestingly, for individuals who are self-employed, similar results can be found in Table 3.11. Among the respondents who report their employment position, more than half are self-employed. Nevertheless, merely 7% of the self-employed people give their information on hourly income because of the low response rate on both monthly earnings (7%) and weekly working hours. As shown in Table 3.12, a large proportion (94%) of agricultural workers are self-employed, whereas for individuals who do not engage in agricultural work, only 19% are self-employed. Further, Table 3.13 examines the distribution of the occupation/employment position categories between the urban and rural areas. It shows that 65% of workers in the rural area work in agriculture whereas 7% are agricultural workers in urban area. In the rural area 69% of respondents are self-employed whereas 20% are self-employed in urban area.

Figure 3.1 Map of Surveyed Regions

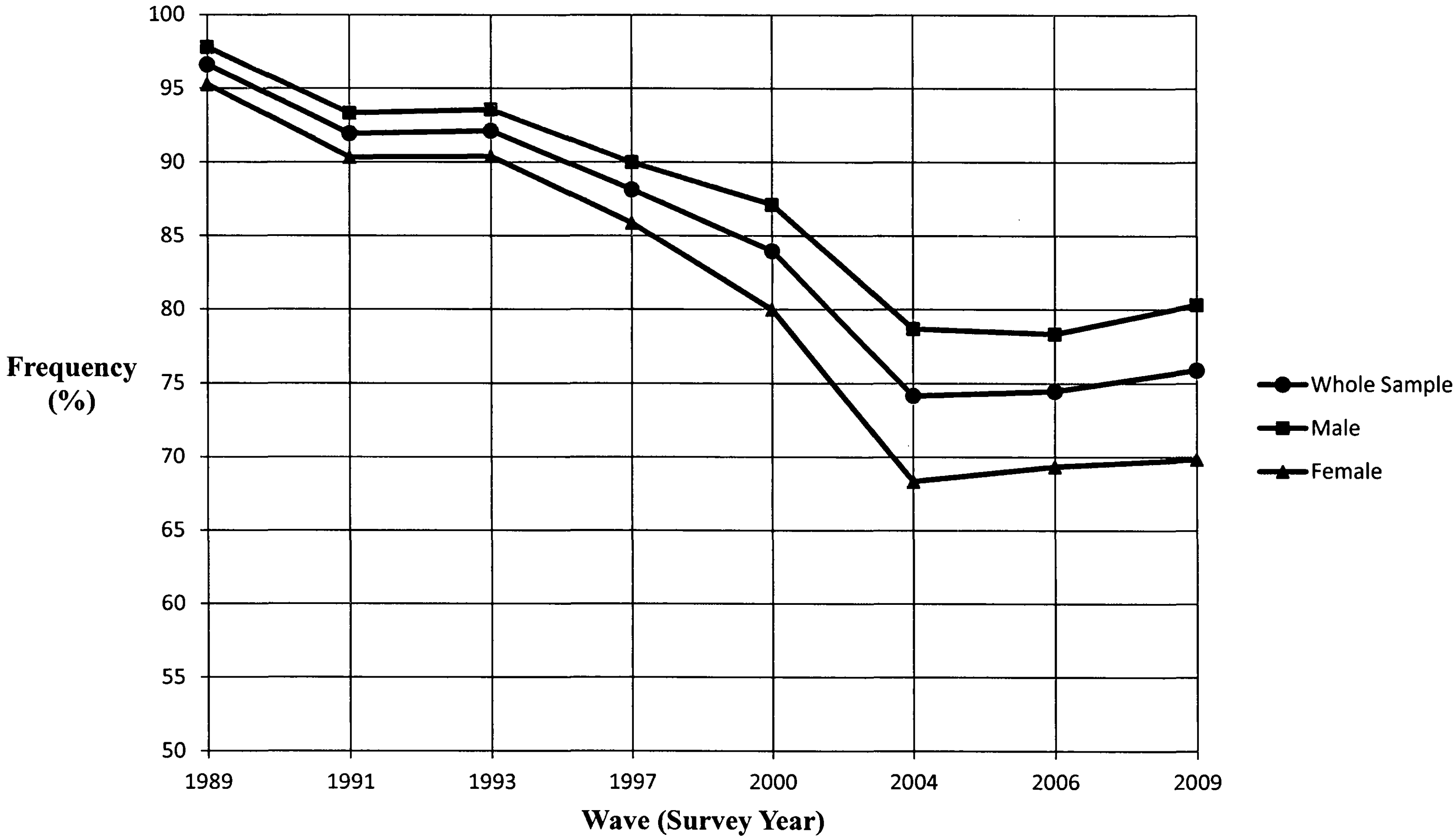
Source: http://www.cpc.unc.edu/projects/china/proj_desc/chinamap.html.

Figure 3.2 Employment Rate by Gender and Age (Longitudinal Dataset, pooled all waves, 1989-2009)



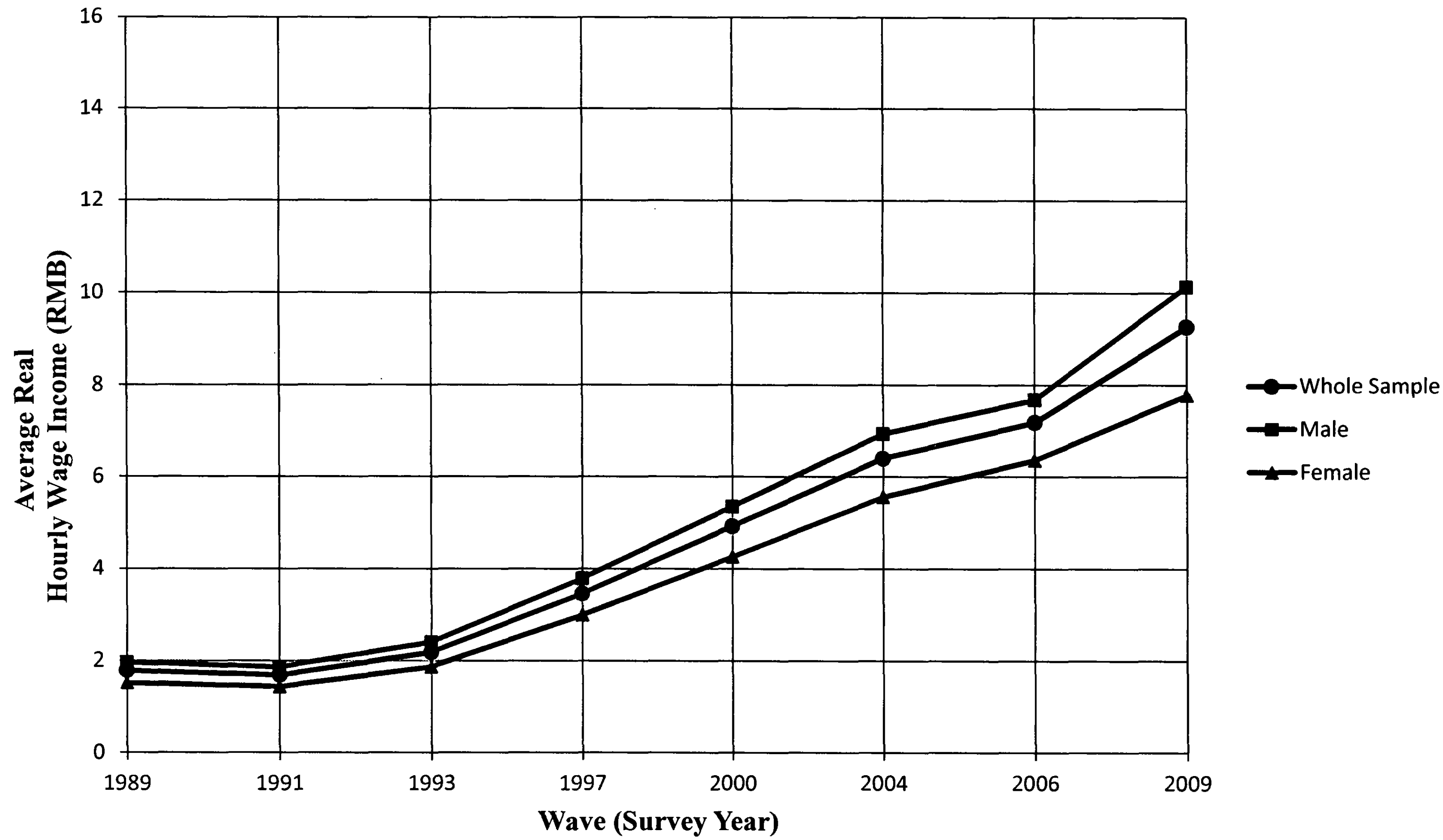
Source: Author's calculations using CHNS data.

Figure 3.3 Employment Rates by Gender and Wave (Longitudinal Dataset)



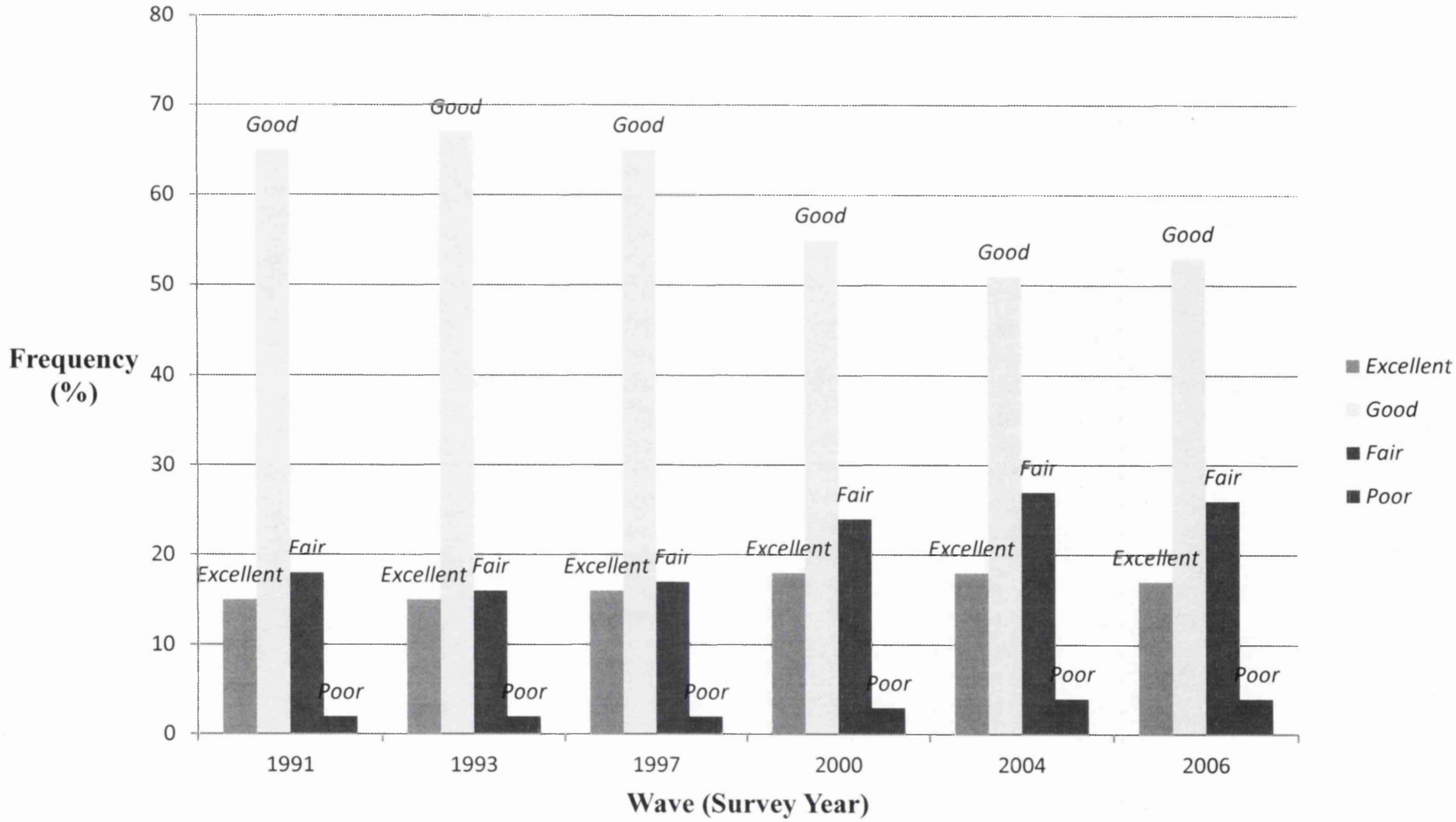
Source: Author's calculations using CHNS data.

Figure 3.4 Average Real Hourly Wage Income by Gender and Wave (Longitudinal Dataset)



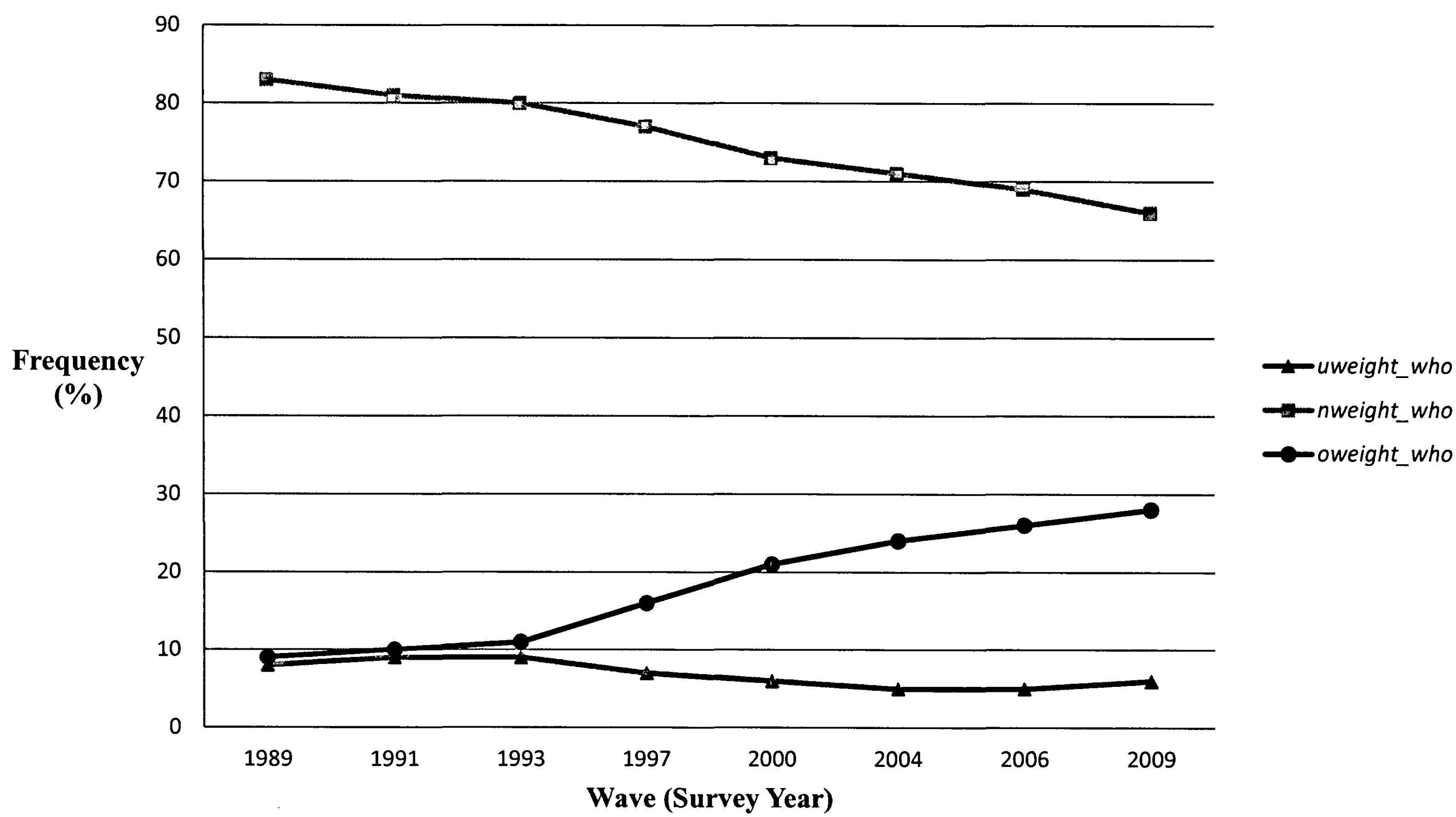
Source: Author's calculations using CHNS data.

Figure 3.5 SRHS by Wave (Longitudinal Dataset)



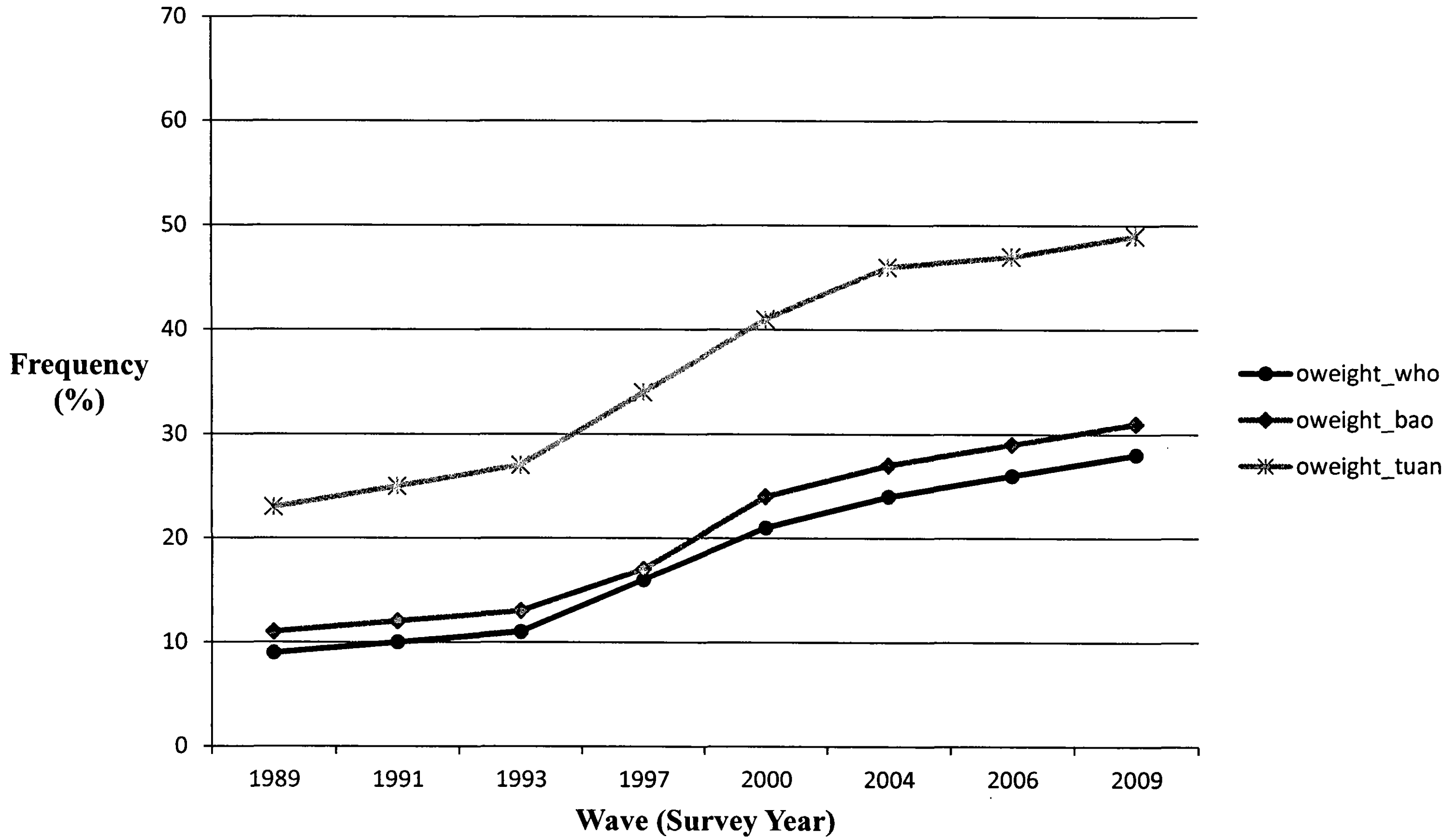
Source: Author's calculations using CHNS data.

Figure 3.6 WHO BMI Classifications by Wave (Longitudinal Dataset)



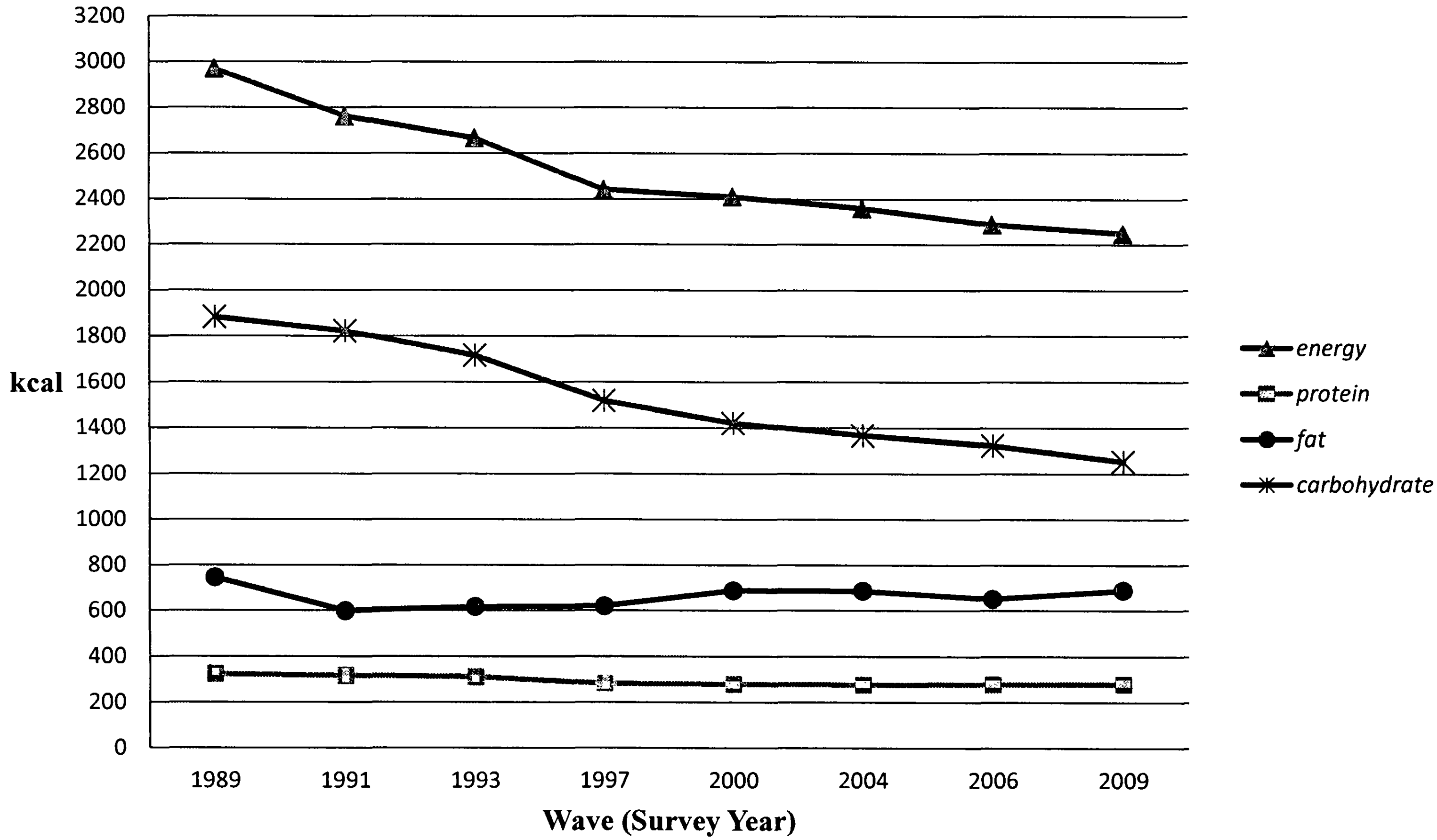
Source: Author's calculations using CHNS data.

Figure 3.7 Different Cutoffs of BMI on Obesity by Wave (Longitudinal Dataset)



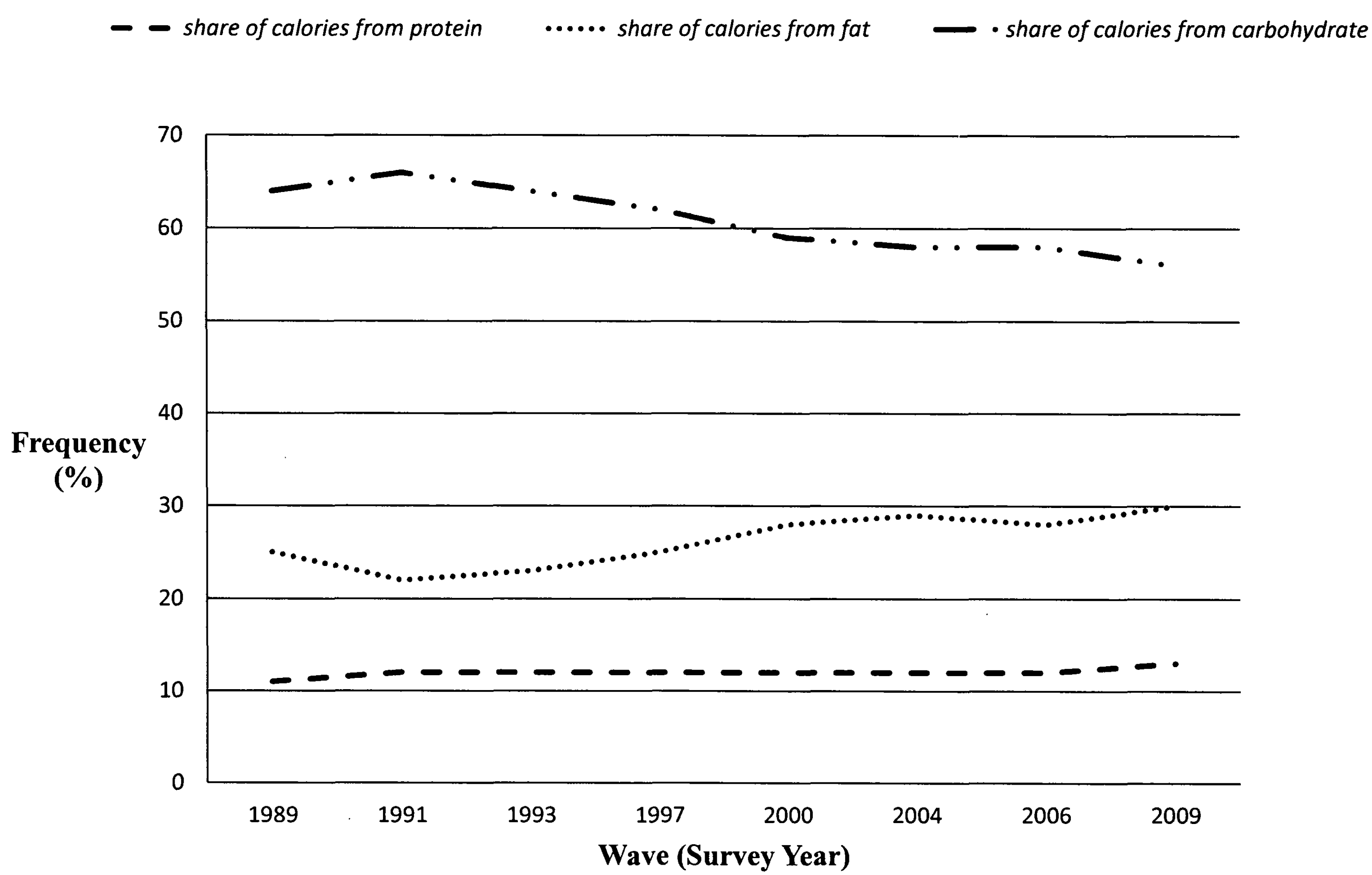
Source: Author's calculations using CHNS data.

Figure 3.8 Individual Daily Nutrient Intakes by Wave (Longitudinal Dataset)



Source: Author's calculations using CHNS data.

Figure 3.9 Shares of Calories from Macronutrient Intakes by Wave (Longitudinal Dataset)



Source: Author's calculations using CHNS data.

Table 3.1 Key Domains of Data (Cross-sectional Datasets) - CHNS

Data Domains	CHNS
Samples	Nine provinces
Number of surveys[1]	Seven (1989–2006), ongoing 2009 and 2011
Households' individuals (100%)	4600/19 000
Household and Individual Survey:	
Demographics (Age/DOB, Sex, Marital status, Education, Living Region and family size)	All surveys
Work activities and income	All surveys
Other income and expenses	All surveys
Time allocations for home activities	All surveys
Drinking water, sanitation, and household assets	All surveys
Accessibility of health care and medical services	All surveys
Marriages, birth preference and pregnancy history of ever-married women	Since 1991
Nutrition and Physical Exam Survey:	
3-Day record of household food consumption	All surveys
3-Day record of household meals per person per day	All surveys
3-Day record of individual daily food	All surveys
Anthropometric measurement and physical examination	All surveys
Hypertension, smoking, alcohol-drinking and injury history	Since 1991
Current health status	Not surveyed for 1989 or 2009
Tea/Coffee – drinking history	Since 1993
Energy record	Since 1997
Community Survey:	
Infrastructure, services and organization	All surveys
Prices of food and specific living material	All surveys

Source: China Health and Nutrition Survey Questionnaires. Available: http://www.cpc.unc.edu/projects/china/data/questionnaires.html.

[1] For 2009, only panel dataset is currently available.

Table 3.2 Sample Statistics (Cross-sectional Datasets) - CHNS

Survey	1989	1991	1993	1997	2000	2004	2006	Total
Sample households	3,794	3,616	3,435	3,834	4,346	4,386	4,468	27,897
Sample individuals (100%):	15,924	14,778	13,893	14,399	15,648	16,269	18,851	109,762
New members	-	600	504	3,764	1,699	2,506	1,381	10,454
(%)		(4%)	(3.6%)	(26.1%)	(10.9%)	(15.4%)	(7.3%)	(9.5%)
Old members	15,924	14,178	13,389	10,635	13,949	13,763	17,470	99,308
(%)	(100%)	(96%)	(96.4%)	(73.9%)	(89.1%)	(84.6%)	(92.7%)	(90.5)
Sample individuals[1] (100%):	15,910	14,625	13,749	14,289	15,436	16,265	18,850	109,124
Children (< 16)	4,618	3,718	3,322	3,189	3,044	2,418	2,857	23,166
(%)	(29%)	(25.4%)	(24.2%)	(22.3%)	(19.7%)	(14.9%)	(15.2%)	(21.2%)
Adults in retirement age (men > 59; women > 49)	1,525	1,581	1,615	1,769	2,051	2,599	2,537	13,677
(%)	(9.6%)	(10.8%)	(11.7%)	(12.4%)	(13.3%)	(16%)	(13.4%)	(12.6%)
Adults in working age (16-59)	9,767	9,326	8,812	9,331	10,341	11,248	13,456	72,281
(%)	(61.4%)	(63.8%)	(64.1%)	(65.3%)	(67%)	(69.1%)	(71.4%)	(66.2%)

Source: Author's calculations using CHNS data. Available: http://www.cpc.unc.edu/projects/china/data/datasets/data_downloads.

[1] Sample individuals refer to individuals that have available age or other equivalent variables.

Notes: China "Labour Law", the 15th regulation, has provided for the minimum working age that the state prohibits hiring people under the age of 16. "Provisional Measures on the workers retire or resign by state council" has provided for the legal retirement age: 60 for male and 50 for female (55 for female who working at state organizations and institutions).

Table 3.3 Sample of Death (Cross-sectional Datasets) - CHNS

Survey	No. of participants who participated in previous survey but died after the previous survey
1989	-
1991	149
1993	166
1997	259
2000	224
2004	281
2006	337

Source: Author's calculations using CHNS data.

Table 3.4 Longitudinal Datasets - CHNS

Main Type of the File	Description
Master ID File	1 file that contains part of demographic information of individuals (such as gender, age, nationality) and their current household (HHID) and individual (LINE) ID; also includes information on death.
Master Roster File	1 file that contains part of demographic information of individuals (such as marital status, living region and family size)
Master Physical Exam (PE) File	1 file that contains information regarding physical activity and all the anthropometric data, also including gender and age.
Master Nutrition Files: Household Daily Food Master Files Individual Daily Meal Master File Individual Daily Food Master File Individual Daily Nutrient Intakes Master File	5 files that provide information on diet, food consumption at both household level and individual level
Master Medical Services Files: Individual Medical Services File Household Medical Service File Master Medical Insurance File	3 files gives information on health insurance, health care availability and utility
Master Income Files: Household Income Master File Individual Income Master File Individual Occupation Master File Individual Wage and Bonus Master File Business Master Files (Household Business File and Individual Business File) Subsidies Master Files (Food Subsidies File, Other Household Subsidies File and Individual Subsidies File) Other Household Income Master File	10 files that cover wages, occupations, household businesses, subsidies and other income.
Individual Education Master File	1 individuals educational file
Master Agriculture Files: Master Farming Files (Household Farming File and Individual Gardening and Farming File)	9 agricultural files involving farming, gardening and fishing at both household level and individual level

Master Household Gardening File Master Crops Type File Master Livestock Files (Household Livestock File, Individual Livestock File and Livestock Type File) Master Fishing Files (Household Fishing File and Individual Fishing File)	
Household Assets File	1 file provides information on household facilities and assets
Individual Time Allocation Master File	1 time use file that gives the amount of time individuals spent taking care of the home, parents and children
Child Care Master Files: Individual Child Care (not surveyed for 1989) Child Care Available to Household	2 files that provide information on child care activities and availabilities.
Master Child/Parent Relationship File	1 file regarding children's household and line ID and their parents they have resided with
Ever-Married Women Files: EMW Master File (not surveyed for 1989) EMW Mass Media Master File (since 2000) Master Marriage History File (not surveyed for 1989) Master Pregnancy History File (not surveyed for 1989) Master Birth History File (since 2004)	5 files include child bearing, mass media, marriage, pregnancy, and birth history.
Infant Feeding Files: Infant Feeding Master File (only surveyed for 1989, 1991 and 1933) 1989 Infant Feeding File	2 files contain information on infant feeding

Source: Summary of the CHNS Longitudinal Datasets. Available: http://www.cpc.unc.edu/projects/china/longitudinal/datasets.

Notes: Family Planning Master Files are also available to download at website: https://www.cpc.unc.edu/projects/china/data/datasets/data_downloads.

Table 3.5 Descriptive Statistics
(Longitudinal Dataset, pooled all waves, 1989-2009: N = 59,356; Sample age: 16-59 for male; 16-49 for female)

Variables	Definition	Obs	Mean	Std. Dev.	Min	Max
Dependent Variables						
employed (working status: respondents who are presently working)	A dummy variable: presently working = 1, presently not working = 0	59,356	0.86	0.35	0	1
hrinc_nominal (calculated average hourly income = total annual wage income / total annual working hours = (monthly wage * working months + annual bonuses) / (4.3 * weekly working hours * working months))	A continuous value	20,744	3.08	4.11	0.02	58.14
hrinc (calculated average hourly income inflated to 2009 = hourlyinc_nominal / inflation index to 2009)	A continuous value	20,744	3.73	4.92	0.03	49.61
lnhrinc (logarithm of real hourly income: log (hourlyinc_real))	A continuous value	20,744	0.91	0.88	-3.53	3.90
Health Variables						
Self-reported Health Status (not surveyed for 1989 and 2009)						
excellent	A dummy variable: excellent = 1, otherwise = 0	42,378	0.16	0.37	0	1
good	A dummy variable: good = 1, otherwise = 0	42,378	0.60	0.49	0	1
fair	A dummy variable: fair = 1, otherwise = 0	42,378	0.21	0.41	0	1

poor	A dummy variable: poor = 1, otherwise = 0	42,378	0.03	0.17	0	1
Anthropometric Measurements						
height (in meters)	A continuous value	47,782	1.62	0.08	1.28	1.90
weight (in kilograms)	A continuous value	47,782	58.77	10.17	30	118.6
waist (in centimeters) (not available for 1989 and 1991)	A continuous value	35,514	78.61	9.75	49.5	128
bmi (calculated by weight/height^2)	A continuous value	47,782	22.32	3.01	15.02	39.63
WHO Criteria						
uweight_who (BMI<18.5)	A dummy variable: BMI<18.5 = 1, otherwise = 0	47,782	0.07	0.26	0	1
nweight_who (BMI 18.5-25)	A dummy variable: BMI 18.5-25 = 1, otherwise = 0	47,782	0.75	0.43	0	1
oweight_who (BMI >=25)	A dummy variable: BMI >=25 = 1, otherwise = 0	47,782	0.18	0.38	0	1
owaist_who	A dummy variable: male WC>102cm, and female WC>88cm = 1, otherwise = 0	35,514	0.06	0.24	0	1
The Optimal cutoffs of BMI for Overweight for the Chinese Population						
oweight_bao	A dummy variable: male BMI >=24.5, and female BMI >=24.7 = 1, otherwise = 0	47,782	0.21	0.41	0	1
oweight_tuan	A dummy variable: male BMI >=22.5, and female BMI >=23.5 = 1, otherwise = 0	47,782	0.36	0.48	0	1

Individual Daily Nutrient Intakes						
energy (3-day average intake in kilocalories)	A continuous value	51,683	2520.19	785.55	802.21	8081.16
protein (3-day average intake in grams)	A continuous value	51,683	73.78	25.53	18.07	243.95
fat (3-day average intake in grams)	A continuous value	51,683	73.22	42.08	5.01	397.90
carbohydrate (3-day average intake in grams)	A continuous value	51,683	385.87	146.40	57.99	1198.71
FAO/WHO/UNU Criteria (Daily Average Energy Requirement by Gender-Weight)						
lenergy	A dummy variable: energy intake does not meet the FAO/WHO/UNU Criteria = 1, otherwise = 0	46,192	56.38	49.59	0	1
nerengy	A dummy variable: energy intake meets the FAO/WHO/UNU Criteria = 1, otherwise = 0	46,192	22.07	41.47	0	1
henergy	A dummy variable: energy intake exceeds the FAO/WHO/UNU Criteria = 1, otherwise = 0	46,192	21.55	41.42	0	1
WHO/FAO Criteria (dietary intake recommendations)						
lprotein	A dummy variable: shares of calories from protein < 10% = 1, otherwise = 0	51,683	22.49	41.75	0	1
nprotein	A dummy variable: shares of calories from protein	51,683	67.3	46.91	0	1

	10-15% = 1, otherwise = 0					
hprotein	A dummy variable: shares of calories from protein > 15% = 1, otherwise = 0	51,683	10.21	30.28	0	1
lfat	A dummy variable: shares of calories from fat < 15% = 1, otherwise = 0	51,683	17.3	37.83	0	1
nfat	A dummy variable: shares of calories from fat 15-30% = 1, otherwise = 0	51,683	47.72	49.95	0	1
hfat	A dummy variable: shares of calories from fat > 30% = 1, otherwise = 0	51,683	34.98	47.69	0	1
lcarbo	A dummy variable: shares of calories from carbohydrate < 55% = 1, otherwise = 0	51,683	29.40	45.56	0	1
ncarbo	A dummy variable: shares of calories from carbohydrate 55-75% = 1, otherwise = 0	51,683	57.39	49.45	0	1
hcarbo	A dummy variable: shares of calories from carbohydrate >75% = 1, otherwise = 0	51,683	13.21	33.86	0	1
Drinking History (during the past year) (not surveyed for wave 1989)						
drink	A dummy variable: drink beer or any other alcoholic beverage = 1, otherwise = 0	45,891	0.39	0.49	0	1
drinkdaily (drinking frequency)	A dummy variable: drink daily or almost every day = 1, otherwise = 0	45,425	0.10	0.30	0	1
drink3to4w (drinking frequency)	A dummy variable: drink 3 to 4 times a week = 1, otherwise = 0	45,425	0.06	0.23	0	1

drink1to2w (drinking frequency)	A dummy variable: drink once or twice a week= 1, otherwise = 0	45,425	0.10	0.30	0	1
drink1to2m (drinking frequency)	A dummy variable: drink once or twice a month= 1, otherwise = 0	45,425	0.08	0.27	0	1
drink0to1m (drinking frequency)	A dummy variable: drink no more than once a month = 1, otherwise = 0	45,425	0.04	0.21	0	1
nodrink (drinking frequency)	A dummy variable: do not drink beer or any other alcoholic beverage = 1, otherwise = 0	45,425	0.62	0.49	0	1
Smoking History (not available for wave 1989)						
eversmoked	A dummy variable: ever smoked cigarettes (including rolled or manufactured) = 1, otherwise = 0	46,166	0.35	0.48	0	1
smokenow	A dummy variable: currently smoke = 1, otherwise = 0	46,017	0.34	0.47	0	1
dailysmokeamt (number of cigarettes smoke per day)	A continuous value	45,642	5.45	9.20	0	60
Other Explanatory Variables						
male (gender)	A dummy variable: male = 1, female = 0	59,356	0.55	0.50	0	1
age (in years)	A continuous value	59,356	35.75	11.20	16	59
age16_25 (age group)	A dummy variable: adults aged 16-24.99 = 1, otherwise = 0	59,356	0.24	0.43	0	1
age26_35 (age group)	A dummy variable: adults aged 25-34.99 = 1,	59,356	0.26	0.44	0	1

	otherwise = 0					
age36_45 (age group)	A dummy variable: adults aged 35-44.99 = 1, otherwise = 0	59,356	0.29	0.45	0	1
age46_59 (age group)	A dummy variable: adults aged 45-54.99 = 1, otherwise = 0	59,356	0.21	0.41	0	1
Area of Residence						
urban	A dummy variable: residents of urban neighbourhood or county town = 1, residents of suburban villages or rural villages = 0	59,356	0.29	0.45	0	1
Region of Residence						
liaoning (province)	A dummy variable: liaoning = 1, otherwise = 0	59,356	0.09	0.29	0	1
heilongjiang (province)	A dummy variable: heilongjiang = 1, otherwise = 0	59,356	0.07	0.25	0	1
jiangsu (province)	A dummy variable: jiangsu = 1, otherwise = 0	59,356	0.12	0.32	0	1
shandong (province)	A dummy variable: shandong = 1, otherwise = 0	59,356	0.10	0.31	0	1
henan (province)	A dummy variable: henan = 1, otherwise = 0	59,356	0.13	0.33	0	1
hubei (province)	A dummy variable: hubei = 1, otherwise = 0	59,356	0.12	0.32	0	1
hunan (province)	A dummy variable: hunan= 1, otherwise = 0	59,356	0.11	031	0	1
guangxi (province)	A dummy variable: guangxi= 1, otherwise = 0	59,356	0.13	0.34	0	1
guizhou (province)	A dummy variable: guizhou = 1, otherwise = 0	59,356	0.13	0.33	0	1
Marital Status						

married	A dummy variable: married = 1, otherwise = 0	59,356	0.76	0.43	0	1
nchildrenhh (number of children in each household)	A continuous value	59,356	0.85	0.89	0	6
child0	A dummy variable: no children in the household = 1, otherwise = 0	59,356	0.42	0.49	0	1
child1	A dummy variable: one child in the household = 1, otherwise = 0	59,356	0.36	0.48	0	1
child2m	A dummy variable: two or more children in the household = 1, otherwise = 0	59,356	0.22	0.41	0	1
Educational attainment						
noedu	A dummy variable: no qualification attained = 1, otherwise = 0	59,356	0.15	0.36	0	1
primary	A dummy variable: highest qualification is primary school = 1, otherwise = 0	59,356	0.21	0.41	0	1
lowermid	A dummy variable: highest qualification is lower middle school degree = 1, otherwise = 0	59,356	0.39	0.49	0	1
uppermid	A dummy variable: highest qualification is upper middle school degree = 1, otherwise = 0	59,356	0.16	0.37	0	1
techvoc	A dummy variable: highest qualification is technical or vocational degree = 1, otherwise = 0	59,356	0.05	0.22	0	1
university	A dummy variable: highest qualification is university or higher = 1, otherwise = 0	59,355	0.04	0.20	0	1

Occupation						
occup_1 (inlcuding senior professional/technical personnel (doctor, professor, lawyer, architect, engineer))	A dummy variable: occup_1 = 1, otherwise = 0	49,800	0.03	0.17	0	1
occup_2 (inlcuding junior professional/technical personnel (midwife, nurse, teacher, editor, photographer))	A dummy variable: occup_2 = 1, otherwise = 0	49,800	0.04	0.19	0	1
occup_3 (including administrator/executive/manager, factory manager, government official, section chief, director, administrative cadre, army officer, police officer)	A dummy variable: occup_3 = 1, otherwise = 0	49,800	0.05	0.21	0	1
occup_4 (including office staff (secretary, office helper))	A dummy variable: occup_4 = 1, otherwise = 0	49,800	0.04	0.20	0	1
occup_5 (including farmer, fisherman, hunter, logger)	A dummy variable: occup_5 = 1, otherwise = 0	49,800	0.49	0.50	0	1
occup_6 (including technical, skilled worker (foreman, craftsman))	A dummy variable: occup_6 = 1, otherwise = 0	49,800	0.08	0.28	0	1
occup_7 (including non-technical, non-skilled worker (labourer))	A dummy variable: occup_7 = 1, otherwise = 0	49,800	0.12	0.32	0	1
occup_8 (including driver, service worker (housekeeper, cook, waiter, doorkeeper, barber/beautician, counter sales, launderer, childcare))	A dummy variable: occup_8 = 1, otherwise = 0	49,800	0.11	0.31	0	1
occup_9 (including engagement in small commercial household business, handicraft, athlete, actor, musician, ordinary	A dummy variable: occup_9 = 1, otherwise = 0	49,800	0.04	0.20	0	1

soldier, policeman and other)						
Employment Position						
selfemp (self-employed, including owner-manager with employees, and independent operator with no employees)	A dummy variable: self-employed = 1, otherwise = 0	49,626	0.56	0.50	0	1
employee (including works for another person or enterprise, paid family worker)	A dummy variable: employee = 1, otherwise = 0	49,626	0.36	0.48	0	1
otherposi (including unpaid family worker, temporary worker, and other)	A dummy variable: emplother = 1, otherwise = 0	49,626	0.08	0.27	0	1
Type of Work Unit						
privatesec (including individual or private enterprise, three source invested enterprise, household business, family contract)	A dummy variable: privatesec= 1, otherwise = 0	49,330	0.63	0.48	0	1
government (including state enterprise or institute/governmental units)	A dummy variable: government = 1, otherwise = 0	49,330	0.21	0.41	0	1
typeother (including small collective enterprise (such as a township enterprise), and large collective (such as county, city or provincially owned enterprise), three-capital enterprise (owned by foreigners, overseas Chinese and joint venture), and others)	A dummy variable: wktypeother = 1, otherwise = 0	49,330	0.16	0.37	0	1
Number of Employees in Work Unit						
firm020	A dummy variable: firm020 = 1, otherwise = 0	50,823	0.25	0.43	0	1

(number of employees are 0 to 20)						
firm20100 (number of employees are 20 to 100)	A dummy variable: firm20100 = 1, otherwise = 0	50,823	0.11	0.32	0	1
firmmt100 (number of employees are more than 100)	A dummy variable: firmmt100 = 1, otherwise = 0	50,823	0.20	0.40	0	1
frimunk (the size of work unit is missing)	A dummy variable: firmunk = 1, otherwise = 0	50,823	0.44	0.50	0	1
secjob	A dummy variable: respondents have second job = 1, otherwise = 0	58,434	0.10	0.29	0	1
Wave (Survey Year)						
1989	A dummy variable: wave1989 = 1, otherwise = 0	59355	0.15	0.35	0	1
1991	A dummy variable: wave1991 = 1, otherwise = 0	59355	0.14	0.34	0	1
1993	A dummy variable: wave1993 = 1, otherwise = 0	59355	0.13	0.33	0	1
1997	A dummy variable: wave1997 = 1, otherwise = 0	59355	0.13	0.34	0	1
2000	A dummy variable: wave2000 = 1, otherwise = 0	59355	0.14	0.35	0	1
2004	A dummy variable: wave2004 = 1, otherwise = 0	59355	0.11	0.31	0	1
2006	A dummy variable: wave2006 = 1, otherwise = 0	59355	0.10	0.30	0	1
2009	A dummy variable: wave2009 = 1, otherwise = 0	59355	0.10	0.30	0	1

Source: Author's calculations using CHNS data.

Table 3.6 Daily Average Energy Requirements

Range of Daily Average Energy Requirements for Men aged 16-29				
Weight	**Basal Metabolic Rate (BMR) (kcal/kg)**	**Light Active Lifestyle (1.40-1.69 BMR) (kcal)**	**Moderate Active Lifestyle (1.70-1.99 BMR) (kcal)**	**Vigorous Active Lifestyle (2.00-2.40 BMR) (kcal)**
50	**29**	2030.00-2450.50	2,465.00-2,885.50	2900.00-3480.00
55	**28**	2156.00-2602.60	2,618.00-3,064.60	3080.00-3696.00
60	**27**	2268.00-2737.80	2,754.00-3,223.80	3240.00-3888.00
65	**26**	2366.00-2856.10	2,873.00-3,363.10	3380.00-4056.00
70	**25**	2450.00-2957.50	2,975.00-3,482.50	3500.00-4200.00
75	**24**	2520.00-3042.00	3,060.00-3,582.00	3600.00-4320.00
80	**24**	2688.00-3244.80	3,264.00-3,820.80	3840.00-4608.00
85	**23**	2737.00-3303.95	3,323.50-3,890.45	3910.00-4692.00
90	**23**	2898.00-3498.30	3,519.00-4,119.30	4140.00-4968.00
Daily Average Energy Requirements for Men aged 30-59				
Weight	**Basal Metabolic Rate (BMR) (kcal/kg)**	**Light Active Lifestyle (1.40-1.69 BMR) (kcal)**	**Moderate Active Lifestyle (1.70-1.99 BMR)**	**Vigorous Active Lifestyle (2.00-2.40 BMR) (kcal)**
50	**29**	2030.00-2450.50	2,465.00-2,885.50	2900.00-3480.00
55	**27**	2079.00-2509.65	2,524.50-2,955.15	2970.00-3564.00
60	**26**	2184.00-2636.40	2,652.00-3,104.40	3120.00-3744.00
65	**25**	2275.00-2746.25	2,762.50-3,233.75	3250.00-3900.00
70	**24**	2352.00-2839.20	2,856.00-3,343.20	3360.00-4032.00
75	**23**	2415.00-2915.25	2,932.50-3,432.75	3450.00-4140.00
80	**22**	2464.00-2974.40	2,992.00-3,502.40	3520.00-4224.00
85	**22**	2618.00-3160.30	3,179.00-3,721.30	3740.00-4488.00
90	**21**	2646.00-3194.10	3,213.00-3,761.10	3780.00-4536.00
Daily Average Energy Requirements for Women aged 16-29				
Weight	**Basal Metabolic Rate (BMR) (kcal/kg)**	**Light Active Lifestyle (1.40-1.69 BMR) (kcal)**	**Moderate Active Lifestyle (1.70-1.99 BMR)**	**Vigorous Active Lifestyle (2.00-2.40 BMR) (kcal)**

40	**27**	1512.00-1825.20	1,836.00-2,149.20	2160.00-2592.00
45	**26**	1638.00-1977.30	1,989.00-2,328.30	2340.00-2808.00
50	**25**	1750.00-2112.50	2,125.00-2,487.50	2500.00-3000.00
55	**24**	1848.00-2230.80	2,244.00-2,626.80	2640.00-3168.00
60	**23**	1932.00-2332.20	2,346.00-2,746.20	2760.00-3312.00
65	**22**	2002.00-2416.70	2,431.00-2,845.70	2860.00-3432.00
70	**22**	2156.00-2602.60	2,618.00-3,064.60	3080.00-3696.00
75	**21**	2205.00-2661.75	2,677.50-3,134.25	3150.00-3780.00
80	**21**	2352.00-2839.20	2,856.00-3,343.20	3360.00-4032.00
85	**21**	2499.00-3016.65	3,034.50-3,552.15	3570.00-4284.00
Daily Average Energy Requirements for Women aged 30-49				
Weight	**Basal Metabolic Rate (BMR) (kcal/kg)**	**Light Active Lifestyle (1.40-1.69 BMR) (kcal)**	**Moderate Active Lifestyle (1.70-1.99 BMR)**	**Vigorous Active Lifestyle (2.00-2.40 BMR) (kcal)**
40	**29**	1624.00-1960.40	1,972.00-2,308.40	2320.00-2784.00
45	**27**	1701.00-2053.35	2,065.50-2,417.85	2430.00-2916.00
50	**25**	1750.00-2112.50	2,125.00-2,487.50	2500.00-3000.00
55	**24**	1848.00-2230.80	2,244.00-2,626.80	2640.00-3168.00
60	**22**	1848.00-2230.80	2,244.00-2,626.80	2640.00-3168.00
65	**21**	1911.00-2306.85	2,320.50-2,716.35	2730.00-3276.00
70	**20**	1960.00-2366.00	2,380.00-2,786.00	2800.00-3360.00
75	**19**	1995.00-2408.25	2,422.50-2,835.75	2850.00-3420.00
80	**19**	2128.00-2568.80	2,584.00-3,024.80	3040.00-3648.00
85	**18**	2142.00-2585.70	2,601.00-3,044.70	3060.00-3672.00

Source: FAO/WHO/UNU (2001), author's own calculations.

Table 3.7 Average Nutrient Intakes for the Total Sample and the Working sample

Nutrient Intakes by Gender and Age			Total Sample (N = 59,356) Mean (Std. Dev.)	Total Working Sample (N = 50,823) Mean (Std. Dev.)
Energy Intake (kcal/day)	Men	<30	2,767.75 (818.73)	2,821.69 (821.07)
		30-59	2642.37 (788.12)	2,675.58 (789.69)
	Women	<30	2,351.08 (744.64)	2,407.70 (751.72)
		30-49	2,323.27 (716.56)	2,363.85 (724.51)
Protein (% of total energy)	Men	<30	12 (2.41)	12 (2.36)
		30-59	12 (2.58)	12 (2.55)
	Women	<30	12 (2.54)	12 (2.50)
		30-49	12 (2.59)	12 (2.55)
Fat (% of total energy)	Men	<30	24 (10.88)	24 (10.84)
		30-59	26 (11.10)	26 (11.02)
	Women	<30	26 (11.72)	25 (11.75)
		30-49	27 (11.53)	26 (11.50)
Carbohydrate (% of total energy)	Men	<30	63 (11.26)	64 (11.23)
		30-59	60 (12.34)	61 (12.30)
	Women	<30	62 (11.82)	63 (11.87)
		30-49	61 (11.78)	62 (11.74)

Source: Author's calculations using CHNS data.

Table 3.8 Gross Domestic Product (GDP) and *per capita* GDP by Region of Residence

Provinces	Nominal GDP (100 million yuan)				
	1997	**2000**	**2004**	**2006**	**2009**
North eastern					
Liaoning	3,490.66	4,669.06	6,872.65	9,251.15	15,212.49
Heilongjiang	2,708.46	3,253.00	5,303.00	6,188.90	8,587.00
Eastern coastal					
Jiangsu	6,680.34	8,582.73	15,403.16	21,645.08	34,457.30
Shandong	6,650.02	8,542.44	15,490.73	22,077.36	33,896.65
Central					
Henan	4,079.26	5,137.66	8,815.09	12,495.97	19,480.46
Hubei	3,450.25	4,276.32	6,309.92	7,581.32	12,961.10
Hunan	2,993.00	3,691.88	5,612.26	7,568.89	13,059.69
South western					
Guangxi	2,015.20	2,050.14	3,320.10	4,828.51	7,759.16
Guizhou	792.98	993.53	1,591.90	2,282.00	3,912.68
Provinces	**Nominal *per capita* GDP (yuan)**				
	1997	**2000**	**2004**	**2006**	**2009**
North eastern					
Liaoning	8,525	11,226	16,297	21,788	35,239
Heilongjiang	7,243	8,562	13,897	16,195	22,447
Eastern coastal					
Jiangsu	9,344	11,773	20,705	28,814	44,744
Shandong	7,590	9,555	16,925	23,794	35,894
Central					
Henan	4,430	5,444	9,470	13,313	20,597
Hubei	5,899	7,188	10,500	13,296	22,677
Hunan	4,643	5,639	9,117	11,950	20,428
South western					
Guangxi	4,356	4,319	7,196	10,296	16,045
Guizhou	2,215	2,662	4,215	5,787	10,309

Source: http://www.stats.gov.cn/english/statisticaldata/yearlydata/.

Table 3.9 SRHS by Gender, Area of Residence and Region of Residence for the Whole Sample (Longitudinal Dataset, pooled waves, 1991-2006)

		Liaoning	Heilongjiang	Jiangsu	Shandong	Henan	Hubei	Hunan	Guangxi	Guizhou	Total
Total	Excellent (%)	17	29	21	27	18	11	15	5	10	16
	Good (%)	58	50	58	57	56	64	61	64	67	60
	Fair (%)	21	18	19	15	23	21	21	28	20	21
	Poor (%)	4	3	2	1	3	4	3	3	3	3
	Total (%)	100	100	100	100	100	100	100	100	100	100
Male	Excellent (%)	19	31	23	27	19	12	17	6	11	17
	Good (%)	59	48	57	57	56	63	61	65	66	60
	Fair (%)	19	17	18	14	22	20	20	26	20	20
	Poor (%)	3	4	2	2	3	5	2	3	3	3
	Total (%)	100	100	100	100	100	100	100	100	100	100
Female	Excellent (%)	16	27	19	26	17	10	13	5	9	15
	Good (%)	57	51	59	57	57	66	63	62	68	60
	Fair (%)	23	19	20	15	23	21	21	30	20	22
	Poor (%)	4	3	2	2	3	3	3	3	3	3
	Total (%)	100	100	100	100	100	100	100	100	100	100

Urban	**Excellent (%)**	21	31	21	35	20	13	16	7	11	18
	Good (%)	57	48	57	49	55	61	61	65	66	58
	Fair (%)	19	19	21	14	23	22	20	25	20	21
	Poor (%)	3	2	1	2	2	4	3	3	3	3
	Total (%)	100	100	100	100	100	100	100	100	100	100
Rural	**Excellent (%)**	16	29	21	23	17	10	15	5	10	15
	Good (%)	58	50	59	60	56	66	61	63	67	61
	Fair (%)	22	17	18	15	23	20	21	29	20	21
	Poor (%)	4	4	2	2	4	4	3	3	3	3
	Total (%)	100	100	100	100	100	100	100	100	100	100

Source: Author's calculations using CHNS data.

Table 3.10 Response Rates across Nine Occupation Categories (Longitudinal Dataset, pooled all waves, 1989-2009)

	Response rate (%)				
Occupation category	**(a)**[1]	**(b)**[2]	**(c)**[3]	**(d)**[4]	**(e)**[5]
occup_1 (senior professional/technical personnel)	3	91	98	90	95
occup_2 (junior professional/technical personnel)	4	93	96	92	92
occup_3 (administrator/executive/manager, officers)	5	91	95	89	89
occup_4 (office staff)	4	94	98	93	93
occup_5 (farmer, fisherman, hunter, logger)	49	3	37	3	30
occup_6 (technical, skilled worker)	8	86	93	85	87
occup_7 (non-technical, non-skilled worker)	12	84	90	83	81
occup_8 (driver, service worker)	11	57	80	56	77
occup_9 (handicraft, literary/art worker and other)	4	36	63	35	55
Total (%)	100	41	63	40	57

Source: Author's calculations using CHNS data. Notes: For full details on each occupation category, see Table 3.5 in this chapter.
[1] Column (a) represents the response rate to each occupation category, i.e. the ratio between the number of people in each occupation category and the total number of people gives response on his occupation. [2] Column (b) represents the response rate of average monthly earnings in each occupation category, i.e. the ratio between the number of people gives response on his monthly income in each category and the total number of people in that category. [3] Column (c) represents the response rate of weekly hours worked in each occupation category, i.e. the ratio between the number of people gives response on his weekly working hours in each category and the total number of people in that category. [4] Column (d) represents the response rate of hourly income in each occupation category, i.e. the ratio between the number of people gives response on his hourly income in each category and the total number of people in that category. [5] Column (e) represents the response rate of the size of work unit in each occupation category, i.e. the ratio between the number of people gives response on his work unit's size in each category and the total number of people in that category.

Table 3.11 Response Rates across Three Employment Position Categories (Longitudinal Dataset, pooled all waves, 1989-2009)

	Response rate (%)				
Employment Position category	**(a)**[1]	**(b)**[2]	**(c)**[3]	**(d)**[4]	**(e)5**
selfemp (self-employed, including owner-manager with employees, and independent operator with no employees)	56	7	40	7	36
Employee (including works for another person or enterprise, paid family worker)	36	92	95	91	88
Otherpositon (including unpaid family worker, temporary worker, and other)	8	47	81	46	60
Total (%)	100	41	63	40	57

Source: Author's calculations using CHNS data.

[1] Column (a) represents the response rate of each employment position category, i.e. the ratio between the number of people in each employment position category and the total number of people gives response on his employment position.

[2] Column (b) represents the response rate of average monthly earnings in each employment position category, i.e. the ratio between the number of people gives response on his monthly income in each category and the total number of people in that category.

[3] Column (c) represents the response rate of weekly working hours in each employment position category, i.e. the ratio between the number of people gives response on his weekly working hours in each category and the total number of people in that category.

[4] Column (d) represents the response rate of hourly income in each employment position category, i.e. the ratio between the number of people gives response on his hourly income in each category and the total number of people in that category.

[5] Column (e) represents the response rate of the size of work unit in each employment position category, i.e. the ratio between the number of people gives response on his work unit's size in each category and the total number of people in that category.

Table 3.12 Distribution of the Employment Position across Agricultural Worker/Non Agricultural Worker (Longitudinal Dataset, pooled all waves, 1989-2009)

Employment Position category	**Agricultural Worker (%)**	**Non Agricultural Worker (%)**
selfemp (self-employed, including owner-manager with employees, and independent operator with no employees)	94	19
Employee (including works for another person or enterprise, paid family worker)	1	70
Otherpositon (including unpaid family worker, temporary worker, and other)	5	11
Total (%)	100	100

Source: Author's calculations using CHNS data.
Notes: see Notes Table 3.10 in this chapter.

Table 3.13 Distribution of the Occupation and the Employment Position across Urban/Rural Area (Longitudinal Dataset, pooled all waves, 1989-2009)

Occupation category	**Urban Area (%)**	**Rural Area (%)**
occup_1 (senior professional/technical personnel)	8	1
occup_2 (junior professional/technical personnel)	8	2
occup_3 (administrator/executive/manager, officers)	10	3
occup_4 (office staff)	11	2
occup_5 (farmer, fisherman, hunter, logger)	7	65
occup_6 (technical, skilled worker)	16	5
occup_7 (non-technical, non-skilled worker)	15	11
occup_8 (driver, service worker)	19	7
occup_9 (handicraft, literary/art worker and other)	6	4
Total (%)	100	100
Employment Position category	**Urban Area (%)**	**Rural Area (%)**
selfemp (self-employed, including owner-manager with employees, and independent operator with no employees)	20	69
Employee (including works for another person or enterprise, paid family worker)	74	22
Otherpositon (including unpaid family worker, temporary worker, and other)	6	9
Total (%)	100	100

Source: Author's calculations using CHNS data.
Notes: see Notes in Table 3.10 in this chapter.

Chapter 4 Health and Employment

4.1 Introduction

An individuals' health status is considered as an important index for assessing a nation's economic development, social advancement, health care and population quality (Chen, 2005). China has achieved impressive economic success since launching market-oriented reforms in 1979 and one of the reasons behind this is the relatively good human capital compared to that of other developing countries (CMH, 2005[53]; Banister, 2009). From this aspect, health, which is a condition to acquiring skills and be productive at work, has played an indispensable role in Chinese economic development.

The prevalence of obesity has become a global health problem in both developed and developing countries, including China (WHO, 2000). The increasing prevalence rate of obesity in the US and in many European countries has been well documented. Due to changes in diet habits and in lifestyle (for example, a high fat or protein intake and a decrease in physical activities), along with the rapid development of Chinese economy, increasing attention has been given to the growing issue of obesity in China (Du *et al.*, 2002). According to Gao *et al.* (2008), China accounts for one out of eight overweight and one out of 16 obese people in the world. In addition, based on the WHO definition, where overweight is defined as a BMI (kg/m^2) of 25-30 (includes 25) and obesity is defined as a BMI (kg/m^2) greater than or equal to 30, the population of overweight and obese individuals in China in 2005 increased to 226 and 34 million, respectively.

As argued by CMH (2005), China has been relying heavily on rapid accumulation of physical capital in recent years. Therefore, the Chinese government is called upon to increase investment in health capital as a precondition to increasing the labour productivity, thereby achieving sustainable economic growth. Moreover, the Seventeenth National Congress of the CPC in 2007 clearly put forward the task of improving the health of the whole nation, including providing everyone with access to basic medical and health services, and strengthening the responsibility and increasing the input of government (Xu, 2007). In this context, the underlying

[53] Recall from the Chapter 2 that 'CMH (2005)' refers to the Report of the Commission on Macroeconomics and Health (CMH) in China, 2005.

question is whether individuals' health would have an impact on their labour market performance such as on employment status.

A better understanding of the effect of health on employment is of crucial importance in estimating the costs of health problems to the economy, this is because if poor health negatively affect labour productivity and labour force participation, health limitations lead to economic loss in terms of production loss (Chirikos, 1993; Cai and Kald, 2006). So far, much of the work in studying this relationship has been undertaken in US and in European countries. Such work for developing countries is relatively scarce, especially for China. Therefore, one of the aims of this chapter is to contribute to the literature in estimating the impact of health on employment for the working-age population in China.

An issue in exploring the impact of health on employment is the measurement of health in general and the potential endogeneity of health in particular. The measurement of health has been described in Chapter 1. One advantage of this analysis is the use of multiple measures of health, i.e. both the self-reported health status and obesity (which is measured by BMI) to examine the link between health and employment.

To account for the potential endogeneity of health, some studies have employed an IV approach to instrument their measure of health such as Stern (1989), Bound *et al.* (1996) and Cawley (2004). This chapter also employs an IV approach to control for the potential endogeneity of health. Based on a recursive bivariate probit model, information on whether individuals were born in the period of the 'Great Chinese Famine' (also known as the 'Three Years of Natural Disasters'), 1958-1961, is used as one of the instruments for both the two health measures. This study adopts two different measures as an additional instrument for SRHS and for obesity, respectively. An area level SRHS measure - the prevalence of SRHS in the area in which the respondent lives, is used as another instrument for individual SRHS. Information on protein intake is used as an additional instrument in estimating the effect of overweight on employment. Therefore, the third advantage is being able to compare the results when assuming health is exogenously associated with employment to the results when treating health as an endogenous variable.

The data source used in this analysis is the second - seventh waves (1991, 1993, 1997, 2000, 2004 and 2006) of the CHNS, which has been described in Chapter 3. Unlike using the cross-sectional data in Pan and Qin (2010), this chapter utilises the panel nature of the data and therefore is able to control for individual unobserved factors which may bias the estimates of the impact of health on employment (Arulampalam and Booth, 1998; Garcia-Gomez and Nicolas, 2006; Cai, 2010; Garcia-Gomez *et al.*, 2010).

This chapter also provides evidence on the impact of health on employment for sub samples within the population, and is able to explore the differences across genders, ages and area of residence in China. The evidence when examining the influence of gender on relationship between health and employment from the existing literature are somewhat mixed. Some studies find evidence that the impact of health on employment is larger for women than for men, for example, in Australia (Cai and Kald, 2006; Cai, 2010) and in the US (Luft, 1975; Baldwin and Johnson, 1994). In contrast, Garcia-Gomez *et al.* (2010) find that the effect is larger for males than for females using the British Household Panel Survey. The effects of health on employment probabilities are expected to be different for urban and for rural residents. Recall from Chapter 2 that the employment reforms increase the employment pressure in urban area. In rural area where agriculture is still dominant, most of the workers are either self-employed or family-run enterprises[54]. From this aspect, health might be more important in terms of employment probabilities for urban residents than for rural residents. Another focus in this chapter is to examine the health effect on employment across the young (aged 16-45) and the older (aged 46-59) working-age Chinese. Researchers argue that it is important to cover the entire working-age group rather than a narrowly defined sample of individuals such as the retirement decisions to the older working-age population. To the author's knowledge, few studies have been undertaken in this area for the Chinese population[55], for example, Benjamin *et al.* (2003) and Pan and Qin (2010). Previous evidence from Australia finds the effect of health on labour force participation is larger for the older groups (Cai and Kald, 2004 and 2006; and Zhang *et al.*, 2009).

[54] Recall from Table 3.13 in Chapter 3 that about 70% of workers in rural area are self-employed, while about 20% of workers in urban area are self-employed.
[55] Chang and Yen (2011) find that mental health problem is negatively associated with labour force participation for the Taiwanese elderly.

This analysis therefore contributes to the literature by analyzing the health effect on employment for both young and older working-age population in China.

The rest of the chapter is organized as follows: the next section briefly reviews some issues regarding the potential endogeneity of health and weight-based discrimination. Section 4.3 reviews the results of previous empirical research. In Section 4.4, the empirical strategies along with more detailed descriptions of the variables used in the empirical analysis are provided. Section 4.5 discusses the empirical results. The final section summarises and draws some conclusions.

4.2 Background

Endogeneity of Health

Human capital theory implies that health might be endogenous to labour market behaviour (Becker, 1964; Grossman, 1972b; and Currie and Madrian, 1999). Individuals need to make investments into health to maintain or improve their health status over their lifetime. Since the investments require both time and material resources, the availability of resources might depend on the individual's labour supply (Cai and Kald, 2006; Cai, 2010). So far a number of studies have found evidence of the endogeneity of health on labour market activities, for example, Stern (1989), Cawley (2004), Cai and Kald (2006), Morris (2007) and Cai (2010). There are three main causes of endogeneity of health to employment: first, simultaneity, that is, employment could also have an impact on individuals' health. For example, employment pressure might be subject to a deterioration of health (Cai and Kald, 2006; and Cai, 2010). Alternatively, it is possible that unemployed individuals who have lower incomes are more likely to consume cheaper, unhealthy or even more fattening food (Morris, 2004 and 2007; and Greve, 2008). The second source of endogeneity is justification endogeneity, which arises from the use of measures of self-reported health measures (e.g. the self-reported health status)[56]. For example, individuals who are not employed might tend to report poor health status to justify their unemployment (Stern, 1989; Bound, 1991; Morris, 2004; Cai and Kald, 2006; Morris, 2007; Greve, 2008; Zhang *et al.*, 1999; and Cai, 2010). Although more objective health measures are less likely to lead to justification endogeneity, it is

[56] It is also note that self-reported health may also suffer from measurement errors (Bound, 1991; Morris, 2004, Cai and Kald, 2006; Zhang *et al.*, 1999).

possible that objective health measures are subject to endogeneity problem such as simultaneity (Bound, 1991; Cai and Kald, 2006; and Cai, 2010). The third source of endogeneity is unobserved factors, which are correlated with both health and employment, such as time preferences (Morris, 2004 and 2007; Greve, 2008; Zhang *et al.*, 1999; and Cai, 2010). It is possible that individuals who are obese have a higher time preference rate and so put less emphasis on their future health. It is also possible that individuals with a higher time preference rate are less likely to invest in their human capital and so are more likely to be employed rather than be out of the labour force (Morris, 2007). If health is treated as an exogenous variable in an employment equation, the estimated effect of health may be biased. Therefore, the potential endogeneity of health to employment should be examined in order to obtain efficient estimates of the effect of health on employment.

Weight-based Discrimination

The existing literature suggests that body weight exerts its impact on employment through two mechanisms. First, body weight might affect employment through the health channel (Pan and Qin, 2010). Having unhealthy body weight may limit the ability to work, which will decrease the individual productivity, thereby affecting employment (Greve, 2008). For example, being underweight is probably caused by malnutrition (Pan and Qin, 2010), while being overweight or obese might be strong indicators for a number of diseases such as hypertension, coronary heart diseases, type II diabetes, gout and pulmonary diseases (NHLBI, 1998; and WHO, 2000). Second, there might be discrimination against people who have unhealthy body weight. According to Morris (2004), weight-based discrimination can be considered as a situation in which, after controlling for all the factors (including health) that should affect employment, having unhealthy body weight still positively or negatively associated with employment. Individuals who have unhealthy body weight suffer discrimination mainly due to two factors: the prejudice or preference from employers, employees or customers (Becker, 1971); and the stereotyping by employers (Everett, 1990; Kristen, 2002; Morris, 2004 and 2007). From the personal aesthetic point of view, employers and employees may prefer to work with those who have a favourable body shape. Furthermore, some employers also believe that obesity is a sign of laziness and lack of self-control, while underweight people (especially underweight men) is often considered as lack of strength (Greve, 2008).

4.3 Empirical Evidence

When investigating the role played by health on labour force participation, many studies have focused on older working-age people and employment (for example, Costa, 1996; Bound *et al.*, 1999; Campolieti, 2002; Au *et al.*, 2005; Disney *et al.*, 2006; and Kalwij and Vermeulen, 2008), since understanding this relationship for older workers has important policy implications regarding the retirement decisions. Cai and Kald (2006) state that it is essential to keep older people in the labour market to meet the increasing demand for public expenditure on income support for the elderly. Moreover, encouraging older workers to remain active in the labour market enhances the labour supply to support national production for developed countries. Nonetheless, studies relating to the link between health and labour market outcomes for younger working-age people are less comprehensive (Currie and Madrian, 1999). Garcia-Gomez et al. (2010) argue that it is of great importance to focus on the entire working-age group rather than a narrowly defined sample of workers.

4.3.1 Evidence from the International Literature

The Impact of Self-reported Health Status, Illness or Disability on Employment

Stern (1989) examines the impact of disability on labour force participation for American adults (aged 25-65) using two data sets. The first data set is a sample of 3,052 adults from 1978 Survey of Disability and Work (SDW) and the second data set is a sample of 11,653 adults from the 1979 cohort of the Health Interview Survey (HIS79). This study uses three indicators to measure disability, including: '*limits*' (whether individuals claimed any health limitations), '*health*' (self-rated health from excellent to poor) and '*conditions*' (the number of symptoms or diseases that individuals have). Labour force participation is measured as whether individuals are employed or not, and a simultaneous equations discrete choice model is estimated by employing both two-step estimation and Maximum Likelihood Estimation (MLE) techniques. In addition, Stern (1989) also uses various symptoms or diseases as instruments to account for the potential endogeneity of disability. Three endogeneity tests are performed, including an asymptotic *t*-test from two-step procedure, a Wald test using maximum likelihood estimators and a Hausman (1978) test. The study concludes that all disability variables negatively affect labour force participation for American adults. It also found that only the MLE estimates for HIS79 present weak endogeneity of '*limits*'.

Cai and Kald (2006) use data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey, 2001, to estimate the impact of self-reported health status on labour force participation. This analysis is carried out for four population groups, men aged 15–49, men aged 50–64, women aged 15–49 and women aged 50–60. A dummy variable measuring the labour force participation status indicates whether individuals are in the labour force (including both employed and unemployed individuals) or not. An ordered variable (0 = poor, 1 = fair, 2 = good, 3 = very good, 4 = excellent) is used to measure the self-reported health status. To account for the potential endogeneity of self-assessed health, a simultaneous equation model which is similar to that in Stern (1989) is estimated using a full information maximum likelihood (FIML) estimation technique[57]. Three major findings are concluded in this study: (i) a positive relationship between better health and the probability of labour force participation is found for all four groups; (ii) for both genders, the effect of health on labour force participation is larger for the older groups than for the younger groups, and for both age groups, the effect is larger for women than for men; (iii) the hypothesis of endogeneity of health cannot be rejected for all groups.

Garcia-Gomez and Lopez-Nicolas (2006) examine the impact of a health shock on employment and income for the Spanish working-age population using panel data from the eight waves of the European Community Household Panel (ECHP), 1994-2001. This study uses various measures of employment and income. The four employment measures are the probability of being employed, the probability of being unemployed, the probability of being out of the labour force (which includes retirement, housework, inactivity and being a student) and the probability of being inactive (which is one of the elements included in the third employment measure). There are eight income measures in this study, including total household income, total individual income and three sources of income (income from labour, private

[57]Cai and Kald (2006) also apply the two-stage estimation method and confirm that both approaches provide consistent parameter estimates. However, they argue that the two-stage estimation method which is also applied in Stern (1989) is less efficient than the FIML method.

transfers and security transfers). The measure of health shock is based on a five-scaled responses to the survey question on self-reported health status (very good, good, fair, bad and very bad), and is defined as a drop from 'very good' or 'good' to 'fair', 'bad', or 'very bad'. The authors estimate the Average Treatment Effect on the Treated (ATET) using matching and a differences in differences technique to control for the endogeneity of health and unobserved heterogeneity. The authors conclude that: (i) suffering a health shock has a statistically significant and negative effect (-5%) on the probability of being employed and has a statistically significant and positive effect (3.5%) on the probability of being inactive; and (ii) negative changes in health significantly reduce total income since the increases in social security transfers does not offset the decrease in workers' income.

Using the pooled data from the Australian National Health Survey (NHS2001 and NHS2004), Zhang *et al.* (2009) explore the effects of four measures of chronic diseases (four dummy variables indicating the incidences of cardiovascular diseases, diabetes, mental diseases and other chronic diseases involving asthma, cancer and arthritis) on the probability of labour force participation. Similar to Cai and Kald (2006), the sample is divided into four groups (men aged 18-49, men aged 50-64, women aged 18-49 and women aged 50-64), and the same technique used in Cai and Kald (2006) is applied on a multivariate endogenous probit model to account for the potential endogeneity of the incidences of the diseases variables in this study. The results show that for men in both age groups, the probability of labour force participation is significantly and negatively affected by all measures of chronic disease[58], and the effects are larger for older men than for younger men. There are no significant effects of chronic diseases on probability of labour force participation for women in both age groups[59]. In addition, the null hypothesis of exogeneity of diseases measures is rejected for all groups.

Cai (2010) investigates the effect of self-reported health on labour force participation of Australia males (aged 25-64) and females (aged 25-59). Unlike Cai and Kald (2006), who use cross-sectional data from the HILDA, 2001, this study uses panel data from the first four waves of HILDA and, therefore is able to control for

[58] There is one exception that the effect of cardiovascular diseases is statistically insignificant for younger men.

[59] There is one exception that the measure of other chronic diseases has a significant and negative effect on the probability of labour force participation for older women.

unobserved heterogeneity. By applying the same measures of health and labour force participation and the FIML method, Cai (2010) concludes the following findings which are similar to Cai and Kald (2006): (i) having better health significantly increases the probability of labour force participation for both Australia men and women, and the effect is larger for women than for men; (ii) based on the Wald statistics, the hypothesis of exogeneity of health is rejected for both genders.

Garcia-Gomez *et al.* (2010) use panel data from the British Household Panel Survey (BHPS), 1991-2002, to investigate the impact of health on the labour market entries and exits between working-age males (aged 16-64) and females (aged 16-59). This study applies the discrete-time hazard method to estimate the effects of health on the hazard of employment and the hazard of non-employment. Employment covers individuals who are either employed or self-employed. Non-employment is defined as individuals whose current economic activity is 'unemployment, retirement, maternity leave, family care duties, being a student, being long-term sick or on a government training scheme' (Garcia-Gomez *et al.*, 2010, pp. 65). The authors use three measures of health: two constructed latent indices of health limitation and self-reported health status which are able to purge the reporting bias, and an index of psychological well-being called General Health Questionnaire (GHQ) index. By applying the discrete-time duration model, the authors confirm that: (i) a worsening in general health positively affects the hazard of non-employment and negatively affects the hazard of employment, and the effect is larger for males than for females; (ii) negative changes in mental health are associated with an increase in the hazard of non-employment, and the effect is larger for men than for women. However, negative changes in mental health do not have a statistically significant impact on the hazard of employment.

By applying the same empirical strategies as Garcia-Gomez and Lopez-Nicolas (2006), Garcia-Gomez (2011) explores the relationship between adverse health transitions and employment for working-age population across nine European countries (Belgium, Denmark, France, Greece, Ireland, Italy, The Netherlands, Portugal and Spain). The panel data used in this study also comes from the same survey as Garcia-Gomez and Lopez-Nicolas (2006), that is, ECHP, 1994-2001. Garcia-Gomez (2011) uses four employment measures and two measures of health shocks. The four employment measures are the probability of being employed (including self-employed), being unemployed, being early retired and being inactive. The first measure of health shock is the same as the measure in Garcia-Gomez and Lopez-Nicolas (2006). The second measure of health shock is defined as the onset of a chronic condition based on a survey question which asks individuals whether they have any chronic health problem or disease. Garcia-Gomez (2011) finds evidence that negative changes in health significantly decrease the probability of being employed, and increase the probability of being unemployed and being inactive for most of the nine European countries. In addition, the effect on the probability of being employed is largest for Ireland, Denmark and The Netherlands, whereas health shocks do not have statistically significant effect on employment for Italy and France.

The Impact of Body Weight on Employment

Both Morris (2004) and Morris (2007) explore the effect of obesity on employment for men and women using pooled data from two waves of the Health Survey for England (HSE, 1997 and 1998). Nonetheless, they use different instruments (base on the IV method) to account for the endogeneity of obesity. The samples used in Morris (2004) and Morris (2007) include 16,929 and 16,967 observations, of working age, respectively.[60] For both studies, the employment indicator is a dummy variable which equals to one if the individual is employed (including self-employed) and zero otherwise. There are two measures of obesity in Morris (2004). The first measure is a dummy variable which equals one if the individual has a BMI above 30 (in kg/m^2) and zero otherwise. The second measure is a dummy variable which equals one if the individual has a BMI above 35 (in kg/m^2) and zero otherwise. Morris (2004) uses a univariate probit model and a recursive bivariate probit model (IV regression) which is able to control for the potential endogeneity of obesity. The instruments used in the IV regression are two measures of individual physical activity, indicating whether or not the respondent has been for a brisk walk and has undertaken any types of physical activity for at least 15 minutes in the last four weeks, respectively. Morris (2004) finds evidence that: (i) there is a negative relationship between obesity and employment for both sexes. For both obesity

[60] In the sample of Morris (2004), 8,305 are male and 8,624 are female. In the sample of Morris (2007), 8,324 are male and 8,643 are female.

measures, the negative effect on employment is larger in magnitude for women than for men. For both men and women, the negative effect is larger for the second than for the first obesity measure; (ii) the null-hypothesis of exogeneity is rejected for both sexes and for both obesity measures, which indicates that failure to control for the endogeneity of obesity will underestimate the negative influence of obesity on employment. The obesity measure used in Morris (2007) is the same as the first measure in Morris (2004). Morris (2007) adopts three methods to account for the potential endogeneity of obesity, which are a univariate probit model, propensity score matching, and a bivariate probit model (IV regression). The instrument used in the IV regression is the prevalence of obesity in the respondents' living area. Moreover, the marginal effects, the ATET and the Local Average Treatment Effect (LATE) are estimated using the univariate probit model, propensity score matching and the IV regression, respectively. The key findings are: (i) the negative impact of obesity on the probability of being employed is statistically significant for both genders; (ii) for men, the null-hypothesis of exogeneity of obesity cannot be rejected, and the negative effect is similar across the three approaches. For women, the null-hypothesis of exogeneity is rejected. Both studies suggest that, under the hypothesis that individuals' health can be correctly controlled for in the employment model[61], there is evidence of discrimination against the obese.

In analysing the impact of BMI on the probability of being employed between genders, Greve (2008) draws on a panel data from the Danish Work Environment Cohort Study (DWECS) 1995 and 2000, including a sample of 3,618 men and 3,666 women, aged 18-60. BMI is calculated by the author using the self-reported height and weight in the Danish survey. To correct for measurement error in BMI, Greve (2008) applies a Bendixsen's formulation to predict the true BMI. The effect of a continuous of BMI value and a discrete measure of BMI are estimated with a probit model, a fixed effects logit model, and an IV model. The discrete measure of BMI contains four dummy variables indicates whether an individual is underweight (BMI less than 18.5), overweight (BMI between 25 and 30), obese (BMI of 30 or more) and having healthy weight (BMI between 18.5 and 25, which is used as the reference group). The author uses whether the interviewee's father or mother had been

[61] Both studies use self-reported health, acute illness, longstanding illness and mental health as control variables for individual's health.

prescribed medication for genetically determined diseases related to obesity between 1995 and 2000 as the instruments of respondent's BMI. The result shows a significant inverted U-shaped relationship between BMI and employment for Danish men and a significant negative relationship for Danish women. The Wald test of exogeneity reveals that BMI is exogenously related to employment for both genders.

Johansson *et al.* (2009) use a sample of 1,163 males and 1,170 females (both aged 30-54) from the Health 2000 in Finland population survey to study the effect of obesity on the probability of being employed (excluding the self-employed[62]). The authors use various indicators of individual body weight, including both continuous and discrete measures of BMI, Waist Circumference (WC) and fat mass, which are measured by the health professionals. The BMI classification in this study is the same as that in Greve (2008). For males, having high fat mass is defined as a percent body fat more than 25 and having high WC is defined as a WC more than 102cm. For females, having high fat mass is defined as a percent body fat more than 30 and having high WC is defined as a WC more than 88cm. The marginal effects are estimated based on a probit model. The results reveal that there is an inverted U-shaped relationship between BMI and employment for males, and no effect of BMI on employment for females. Fat mass is negatively associated with employment for males and both fat mass and WC have a statistically significant and negative impact on employment for females. Regarding the probit estimates when applying discrete measures of individual body weight, the results show that being underweight, having high WC and having high fat mass are negatively associated with employment for males. For females, all measures of obesity (including being overweight, being obese and having high WC/fat mass) have statistically significant and negative effects on employment. However, Johansson *et al.* (2009) point out that their study fails to account for the endogeneity of obesity due to lack of valid instruments.

4.3.2 Evidence from the Chinese Literature

In studying the impact of health on labour supply of the elderly in rural China, Benjamin *et al.* (2003) extract five health indicators from the panel dimension of the 1991, 1993 and 1997 waves of the CHNS. The first (binary) health indicator used in this paper, is the top two categories of the Self-Reported Overall Health Status

[62] Johansson *et al.* (2009) state that only 7% of the labour force in Finland is self-employed.

(SRHS) which can be collected from each of the three waves in the CHNS. They exploit IV methods to address the potential endogeneity of SRHS and attenuation bias. The second indicator is the BMI. The authors create two extremums which are 'Low BMI' and 'High BMI'. 'Low BMI' is defined as BMI less than 20, and 'High BMI' is defined as BMI greater than 30. The third indicator is the limitations on activities of daily living (ADL). However, the CHNS only collects data on the limitations on daily living activities for people aged over fifty years old. The last two health indicators are an index of Physical function limitations (PF) and a binary measure of Subsequent Death, respectively. The index of PF is derived from a series of five questions which ask information about the current state of individuals' physical functions such as heart, stomach, arms, legs, hearing and speaking. The authors utilize the index of PF's as instruments for the SRHS to account for the potential endogeneity and measurement error bias. The Subsequent Death indicator is used to validate other health indicators. Three labour supply indicators are measured as: i) positive weeks worked during the last year; ii) hours of work during the last year (including zero values); iii) hours spent on different types of work. Benjamin *et al.* (2003) also estimate both fixed-effects and random-effects model and apply non-parametric, reduced-form and structural analysis in addressing the difficulty of isolating age from cohort effects. Although they conclude that deteriorating health plays a small observable role in interpreting decreasing labour supply for Chinese elderly workers aged from 60 to 70, their findings suggest a statistically significant estimate of the impact of health on labour supply for the other working-age groups.

Based on the Urban Resident Basic Medical Insurance Survey (URBMI) 2008, Pan and Qin (2010) examine the impact of BMI on the probability of being employed for the Chinese working-age population. The sample is split by gender and age (men/women aged 18-30, men/women aged 30-40, men/women aged 40-50, men aged 50-60 and women aged 50-55). The authors calculate BMI based on the self-reported height and weight in the survey. The effect of a continuous measure of BMI and a discrete measure of BMI are estimated with the probit model and an IV model[63]. Instead of using the international classification of BMI such as Greve (2008) and Johansson *et al.* (2009), Pan and Qin (2010) use a BMI classification which is

[63] Pan and Qin (2010) only use the IV method to estimate the effect of BMI on employment between genders.

proposed by the Cooperative Meta-analysis Group of the Working Group on Obesity in China (C-WGOC) and is specifically for the Chinese population. The discrete measure of BMI contains four dummy variables indicates whether an individual is underweight (BMI less than 18.5), overweight (BMI between 24 and 28), obese (BMI of 28 or more) and having healthy weight (BMI between 18.5 and 24, which is used as the reference group). The instruments are the average BMI, the prevalence of underweight, the prevalence of overweight and the prevalence of obese in respondents' living area. There are four major findings: (i) there is a statistically significant and inverted U-shaped relationship between the continuous measure of BMI and employment for both genders; (ii) for men, being underweight is negatively associated with employment and being overweight is positively associated with employment. For women, being obese is negatively associated with employment, neither being underweight nor being overweight has a statistically significant effect on employment; (iii) based on the Wald statistics, BMI is exogenously related to employment for both genders; (iv) BMI is more sensitive in terms of the probability of being employed for people in age group 18-30 than those in other age groups. The authors suggest that their findings provide evidence of discrimination against underweight men and obese women. Nevertheless, they also argue that there are two shortcomings in their study: that the use of the self-reported height and weight in calculating BMI might produce measurement errors; and they are unable to deal with the omitted variables bias by using the cross-sectional data.

4.4 Model and Methodology

4.4.1 Model and Methodology - SRHS as a Measure of Heath

In order to estimate the impact of SRHS on employment for the Chinese working-age population, this chapter firstly adopts the pooled probit model. The panel nature of the CHNS is then utilized to estimate a Random Effects (RE) probit model and a Fixed Effects (FE) logit model[64] to control for unobservable individual heterogeneity. Furthermore, this study employs a recursive bivariate probit model, which draws heavily on model specification of Morris (2007), to account for the potential

[64] The author uses Stata programme to estimate the pooled probit, RE probit and FE logit. To the best knowledge of the author, Stata does not allow for the FE probit estimation.

endogeneity of SRHS. The following model specifications also draw heavily on Arulampalam and Booth (1998), and Cameron and Trivedi (2009).

Similar to the cross-sectional probit model which is used to estimate binary or dichotomous outcome variables, the pooled probit model also relies on a between comparison as all individual observations are pooled across time, and is as follows:

Assume, $P(y_{it} = 1|x_{it}) = \Phi(\beta_0 h_{it} + \beta_1 x_{it})$,

where, P = probability,

i = individuals,

t = time periods, i.e. waves 1991, 1993, 1997, 2000, 2004 and 2006,

y_{it} = a binary response indicator of employment,

h_{it} = SRHS indicator,

x_{it} = a vector of additional (either time-varying or time-invariant) explanatory variables that affect employment,

β_0 and β_1 = coefficients, and β_0 is a measure of the impact of SRHS on employment to be estimated from these,

Φ = the standard normal Cumulative Distribution Function (CDF).

The probit model is as follows:

$$y_{it} = \begin{cases} 1 \text{ if } y_{it}^* = \beta_0 h_{it} + \beta_1 x_{it} + \varepsilon_{0it} > 0, i = 1, \dots, n, t = 1, \dots, T \\ 0 \text{ otherwise} \end{cases}, \quad (4.1)$$

where, y_{it}^* = a latent variable of employment,

ε_{0it} = random errors which are assumed to follow the standard normal distribution with variance 1, and $\varepsilon_{0it} \sim N(0,1)$.

Under the weaker assumption that errors are independent across individuals, ε_{0it} is likely to be correlated over time for a given individual, thus cluster-robust standard errors are then used to correct for error correlation over time for a given individual.

By allowing for the panel nature of the data, a RE probit assumes that the heterogeneity across individuals is time-invariant, the error term ε_{0it} in Equation (4.1) can be decomposed as:

$$\varepsilon_{0it} = \kappa_{0i} + \mu_{0it} \quad (4.2)$$

where, κ_{0i} = the individual specific unobservable effect,

μ_{0it} = a random error term, which is independent of x_{it}, assumes $\mu_{0it} \sim IN(0, \sigma_{\mu_0}^2)$, where *IN* refers to Independent Normal Distribution.

In order to marginalize the likelihood, it assumes that $\kappa_i \sim IN(0, \sigma_{\kappa_0}^2)$ and is independent of the μ_{0it} and x_{it}. This implies that the correlation between two successive error terms for the same individual i is a constant, given by:

$$\rho_0 = corr(\varepsilon_{0i2}, \varepsilon_{0i1}) = \frac{\sigma_{\varepsilon_0}^2}{\sigma_{\varepsilon_0}^2 + \sigma_{\mu_0}^2} \tag{4.3}$$

A likelihood-ratio test of the significance of ρ_0 (H_0: $\rho_0 = 0$) compares the pooled probit estimator with the RE probit estimator (StataCorp, 2009). If $\rho_0 = 0$, the pooled probit parameter estimates will be equal to the RE probit estimates, then it is appropriate to use the pooled probit estimator. If $\rho_0 \neq 0$, then it suggests that the RE probit estimator would be more appropriate.

In the case of the FE logit model, assume, $P(y_{it} = 1|x_{it}) = \Lambda\ (\beta_0 h_{it} + \beta_1 x_{it})$, where Λ denotes the logistic CDF, x_{it}denotes a vector of time-varying explanatory variables, $\kappa_i$denotes an individual fixed effect which is independent of x_{it}, and $\mu_{it}$is a time-varying logit-distributed error term. Identification of the FE logit model relies on within-individual variations in the dependent variable and explanatory variables. Therefore those individuals who do not change employment status cannot be included in the FE logit estimation, neither the coefficients of the time-invariant explanatory variables, such as gender, are identified. Unlike the random effects methods, fixed effects methods help to control for omitted variable bias by having individuals serve as their own controls (Allison, 2009; Williams, 2011a and 2011b). Williams (2011b) states that if there are omitted variables, and these variables have time-invariant effect on the dependent variables, then fixed effect models may provide a means for controlling for omitted variable bias.

It is well known that in a linear regression model the estimated coefficient measures the effect on the average value of the depend variable for a unit change in the value of the explanatory variable (Gujarati, 1995). However, the coefficient estimate in a nonlinear model need to be transformed to obtain an estimate of the marginal effects, since the marginal effects give the change in predicted probability associated with the changes in the explanatory variables (Anderson and Newell, 2003).

In the probit model, the marginal effects for explanatory variables (i.e.$\partial E(y_{it})/\partial(\beta_0 h_{it} + \beta_1 x_{it})$) are given by:

$$\partial E(y_{it})/\partial(\beta_0 h_{it} + \beta_1 x_{it}) = (\beta_0 + \beta_1)_{it}\phi(\beta_0 h_{it} + \beta_1 x_{it}), \quad (4.4)$$

where, E = the predicted probability, given that $E(y_{it}) = \Phi(\beta_0 h_{it} + \beta_1 x_{it})$,

ϕ = the standard normal density function, where

$$\phi(\beta_0 h_{it} + \beta_1 x_{it}) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}(\beta_0 h_{it} + \beta_1 x_{it})^2\right),$$

β_{it} = the coefficient of the explanatory variable x_{it}.

In the logit case,

$$\partial E(y_{it})/\partial(\beta_0 h_{it} + \beta_1 x_{it}) = (\beta_0 + \beta_1)_{it} g(\beta_0 h_{it} + \beta_1 x_{it}), \quad (4.5)$$

where, g = the probability density function, where

$$g(\beta_0 h_{it} + \beta_1 x_{it}) = \frac{exp(\beta_0 h_{it} + \beta_1 x_{it})}{[1 + exp(\beta_0 h_{it} + \beta_1 x_{it})]^2}.$$

Dummy (or categorical) variables can be normalized to equal to zero at the desired reference point, $\beta_0 h_{it} + \beta_1 x_{it}$ simplifies to c, and $\phi(\beta_0 h_{it} + \beta_1 x_{it})$ simplifies to $\phi(c)$, where c is the estimated constant term. Therefore, the marginal (or discrete) effect of a dummy variable (i.e. the difference between the predicted probability with and without that dummy variable equals to one) is given by:

$$E(y_{it}|d = 1) - E(y_{it}|d = 0) = \phi(c + d) - \phi(c), \quad (4.6)$$

where, d = the estimated parameter for the dummy variable.

In the logit case,

$$E(y_{it}|d = 1) - E(y_{it}|d = 0) = g(c + d) - g(c), \quad (4.7)$$

Theoretical speaking, the marginal effect of an explanatory variable gives the difference between the predicted probability of being employed with and without that explanatory variable equals to one, holding other explanatory variables at their reference groups.

The above methods assume that SRHS conditional on the covariates is independent of employment, i.e. the SRHS is exogenously related to employment. If SRHS is endogenous then the conditional independence assumption is violated, and the marginal effects based on the above models are biased, which results in an unreliable estimate of the effect of SRHS on employment. Therefore, this study uses an IV regression approach based on a recursive bivariate probit model where the dependent variable is binary and an independent variable is binary and endogenous. Based on Morris (2007), the model is as follows:

$$y_{it}^{*} = \beta_2 h_{it}^{*} + \beta_3 x_{it} + \varepsilon_{1it} \quad (4.8)$$

$$h_{it}^{*} = \beta_4 z_{it} + \beta_5 x_{it}^{h} + \varepsilon_{2it} \quad (4.9)$$

$$E[\varepsilon_1] = E[\varepsilon_2] = 0 \quad (4.10)$$

$$Var[\varepsilon_1] = Var[\varepsilon_2] = 1 \quad (4.11)$$

$$Cov[\varepsilon_1, \varepsilon_2] = \rho_1 \quad (4.12)$$

where, h_{it}^{*} = an unobserved latent variable such that $h_{it} = \begin{cases} 1 \text{ if } h_{it}^{*} > 0 \\ 0 \text{ if } h_{it}^{*} \leq 0 \end{cases}$

z_{it} = a vector of IVs that are correlated with h_{it} but not ε_1,

x_{it}^{h} = a vector of additional explanatory variables that affect h_{it}^{*},

$\beta_2 \sim \beta_5$ = coefficients, and the coefficients of interest is β_2,

ρ_1 = the correlation between the error terms (i.e. ε_1 and ε_2) in Equation (4.8) and (4.9).

A Wald test of the significance of ρ_1 is a direct test of the endogeneity of h_{it}^{*} in Equation (4.8). If $\rho_1 = 0$ then it is appropriate to adopt the univariate probit model. If $\rho_1 \neq 0$ then SRHS is endogenous, the univariate probit estimates are biased, which indicates that the recursive bivariate probit model should be adopted.

The method of computing marginal effects is similar to that in the pooled probit model. The difference is that the estimated coefficients used in Equation (4.9) are β_2 and β_3 instead of β_0 and β_1. Therefore the marginal effects in this case take into account the endogeneity of SRHS.

Applying an IV method to control for potential endogeneity of SRHS requires a variable that is correlated with and exogenous to SRHS but uncorrelated with employment. Morris (2007) states that instruments should fit two characteristics: first, it must be a non-weak predictor of h_{it} conditional on x_{it} (i.e. $\beta_4 \neq 0 | x_{it}$); it is possible to test this characteristic by examining the significance of z in the Equation (4.9); second, it must be uncorrelated with ε_{1it} (i.e. $\text{Cov}(z_{it}, \varepsilon_{1it}) = 0$), however, this is not possible to test directly and this assumption must be maintained. The validity of the variables (which are considered as potentially suitable instruments) will be examined later in this section.

The data used in studying the impact of health on employment are the six waves of the CHNS which include 1991, 1993, 1997, 2000, 2004 and 2006.[65] The analysis is based on a sample of working-age adults (16-59 for male, 16-49 for female), and screens out the observations that have missing values for general demographics[66]and self-reported health status. In addition, those who are still at school or undertaking full-time study are also excluded from the sample (i.e. 6% of sample is excluded), because they might not have established work patterns (Haveman *et al.*, 1994). Thus the sample includes 15,272 individuals pooled across the six waves (for each $i, t = 1, \dots, T$, where $T \leq 6$ in a case of the unbalanced panel), that is, a total of 40,723 observations.

This chapter uses an unbalanced panel data. It is generally argued that a balanced panel is preferred over an unbalanced panel since it allows an observation of the same individual in every time period, which reduces the noise caused by individual heterogeneity. Some researchers also point out that consideration should be given to the potential panel bias occurring due to attrition in research analysis, such as Russell (1981) and Benjamin *et al.* (2003). Poking *et al.* (2009) document the extent of attrition of the CHNS - if response rate was based on those who participated in previous survey years remaining in the current survey, the response rate was about 88%; and would be about 69% if response rate was based on those who participated in first survey (CHNS, 1989) and remained in survey 2006. The authors also list three main causes of attrition, which have also described in Section 3.1 of Chapter 3 (on p. 35): (i) participants who left in one survey year (and may also move back in a later year) due to travel, hours of work or refusal to participate the survey; (ii) since 1997 new participants were added as replenishment samples if participants have formed a new household or joined another household or there are fewer than twenty households in a survey site; and (iii) Liaoning province was replaced by

[65] Recall from Chapter 3 that the first wave and the latest wave are not used because both waves did not survey the self-reported health status which is one of the health measures of this thesis.
[66] Recall from Chapter 3 that general demographic information in this thesis is considered as information on gender, age, marital status, area of residence, region of residence, educational attainment and current working status (presently working or not working).

Heilongjiang province in 1997 (and rejoined in the following survey year, in 2000), because it was unable to participate in the survey for natural disaster, political and administrative reasons.[67] Accordingly, if using a balanced data, individuals who living in Liaoning or Heilongjiang, that is, around 20% of the sample, would be dropped from the analysis. Although there is no systematic reason for not using unbalanced data and this approach is adopted throughout, the results should be discussed with caution and awareness of potential influence of panel attrition bias.

The sample is split by sex, by urban/rural area and by age group. Young working-age adults are defined as adults who are aged between 16 and 45, and older working-age adults are defined as adults who are aged 46 and over. Separate analyses are undertaken for 7,818 males and 7,454 females (22,411 and 18,312 observations, respectively), for 4,905 urban and 10,367 rural residents (11,407 and 29,316 observations, respectively), and for 11,204 young and 4,068 older working-age adults (31,801 and 8,922 observations, respectively).

As explained in Chapter 3 the binary response indicator y_{it} denotes employment, i.e. a dummy variable, *employed*, which equals to one if an individual is presently working (including self-employed), and equals to zero otherwise[68]. The SRHS indicator includes four dummy variables *excellent*, *good*, *fair* and *poor* (the reference category), which take value of one if an individual has excellent health, has good health, has fair health or has poor health, respectively. Since the bivariate probit model requires that both the dependent variable and the endogenous variable take a binary form (Greene, 2011), the dummy variable, *poor*, is used to measure individual health. Further justification for the focus on poor health as the dependent variable in the bivariate probit model will be provided as part of the discussion of the empirical results in the next section.

The other explanatory variables have been found to affect employment status (Mullahy and Sindelar, 1996; Tekin, 2002; and Pelkowski and Berger, 2004) include gender (*male*), age, educational attainment, urban/rural area (*urban*), region of

[67] For more detail, see Popkin *et al.* (2009), p. 3.
[68] Recall from Chapter 3 that the employment indicator used in this thesis is consistent with that in Jones (2011).

residence, marital status (*married*), and presence of children in the household.[69] A more comprehensive discussion of the explanatory variables is provided in Chapter 3.[70] Four dummy variables capture age effects include *age16_25* (the reference category), *age26_35*, *age36_45* and *age46_59*. Six binary indicators record the highest educational level achieved include *noedu* (the reference category), *primary*, *lowermid*, *uppermid*, *techvoc* and *university*. The following nine dummy variables control for region of residence (i.e. province) are *liaoning* (the reference category), *heilongjiang*, *jiangsu*, *shandong*, *henan*, *hubei*, *hunan*, *guangxi* and *guizhou*. *child0* (the reference category). Three dummy variables, *child0* *child1* and *child2m*[71], indicate the presence of children in the household. Dummy variables for the survey year are also included: *1991* (the reference category), *1993*, *1997*, *2000*, *2004* and *2006*.

To account for the potential endogeneity of health, some studies use long-term health condition and physical functioning limitations as instruments for self-reported health (Stern, 1989; Bound *et al.*, 1996; Cai and Kald, 2006; Cai, 2010). It is not possible to use either of these variables here because information on long-term health conditions is not surveyed in all the selected waves[72] and information on physical functioning limitations is only surveyed for Chinese adults aged 55 and over.

This chapter considers a number of alternative variables as potentially suitable instruments for SRHS. Recall from Chapter 4 that three dummy variables are created according the energy requirement range proposed by FAO/WHO/UNU (2001), including *lenergy*, *nerengy* and *henergy*, which takes value of one if individual daily average energy intake is below, within, and above the FAO/WHO/UNU criteria, respectively. Similarly, based on the WHO/FAO criteria, nine dummy variables which reflect the levels of the three macronutrient intakes are created (*lprotein*,

[69] It is worth to mention that since the CHNS 1991 did not survey the individuals' nationality (minority status), the estimation fails to control for the effect of individuals' nationality.
[70] Table 3.5 in Chapter 3 gives descriptive statistics for both the dependent variable and the explanatory variables.
[71] Recall from Chapter 3, *child0* takes the value one if there are no children in the household, and zero otherwise; *child1* takes the value one if there is one child in the household, and zero otherwise; and *child2m* takes the value one if there are two or more children in the household, and zero otherwise.
[72] The information on long-term health condition is based on a survey question - whether individuals have any of health problems such as goiter, angular stomatitis, blindness in one eye (or both eyes), loss of one arm (or both arms) or use of one arm (or both arms), and loss of one leg (or both legs) or use of one leg (or both legs). Although the question can be found in the CHNS 1993, but for some reason the corresponding variable is not included in the original longitudinal dataset.

nprotein, *hprotein*, *lfat*, *nfat*, *hfat*, *lcarbohydrate*, *ncarbohydrate*, and *hcarbohydrate*). However, these variables are not found to be suitable instruments for individual SRHS due to one or both of the following reasons. First, they do not have a statistically significant effect on health. Second, they are correlated with employment. This was explored by examining their significance in the Equation (4.8).

In this chapter, the first instrument used for SRHS is a dummy variable, *famine* which takes value of one if an individual was born in the period of the 'Great Chinese Famine' (1958-1961). This study also adopts an area level SRHS measure - the prevalence of SRHS health status in the area in which the respondent lives, as another instrument in for individual SRHS. It is defined as the average value of the SRHS in the area (or community) in which the respondent lives, *ivsrhs_t4* (where the individual level of SRHS value is defined as: 1 = poor, 2 = fair, 3 = good, and 4 = excellent). Recall from Chapter 4 that the CHNS surveyed nine provinces. In each province, both cities and counties are surveyed. Urban neighbourhoods and suburban villages within the cities, and county towns and rural villages within counties are then randomly selected. In total there are 216 sampling unities (or communities) from nine provinces, including 36 urban neighbourhoods, 36 suburban villages, 36 county towns and 108 rural villages. Based on community-level, the mean number of sample observations per area is 173, and ranges from 20 to 335. The use of the area-based SRHS as an instrument for individual SRHS relies on a 'peer group effect'.[73] The two instruments are included in the Equation (4.9). According to Cai and Kalb (2006) and Cai (2010), the other explanatory variables which have been found to affect self-reported health status, including all the other variables in employment equation except the presence of children in the household, are also included in the health equation (Equation (4.9))[74].

The average employment rates against SRHS for each of the sample groups are tabulated in Table 4.1a. The average employment rate of each health status category is defined as a ratio between the number of people who are presently working in each health status category and the total number of people in that health status category. It

[73] The 'peer group effect' has been well documented in Morris (2007) where the area level obesity is used as the instrument for individual obesity.

[74] The significance levels of the variables on the presence of children in the household are also tested by including them in the health equation. The result shows that they do not have a statistically significant effect on health.

shows that the average employment rate of *poor* health status category is the lowest among the four SRHS categories for all the sample groups. For example, the employment rate of *poor* health category for men and women are 72% and 77%, respectively. The average employment rates of the other three health status categories are similar to each other at between 86%-89%. In addition, the employment rates of the these three health status categories are larger for the male, the rural, and the young working-age sample than for the female, the urban, and the older working-age sample, respectively. Table 4.1a also shows the distribution of SRHS (in parentheses) for each of the sample groups. For all the sample groups, *good* health status accounts for the highest proportion of the population health status categories (more than a half), while the *poor* health status typically accounts for less than 5% (except the older working-age sample).

Table 4.1b provides descriptive statistics across the sample groups for the other explanatory variables and the instruments. To summarise, 4% of men and 3% of women (3% of young and 4% of older people) have a degree qualification. Urban residents are, on average, better educated than rural residents, with 9% of urban residents and 1% of rural residents having a degree qualification. In rural area, 26% of adults have two or more children in their household, which is more than twice as many as the urban residents. For the older working-age (aged 46-59) sample, 4% was born in the period of the famine (1958-1961). For the other sample groups, around 6% - 8% were born in that period.[75] The mean value of the area-based SRHS measure for all the sample groups is very close to 3, i.e. good health status. This is consistent with the statistics from Table 4.1a that 53%-62% of sample has good health status.

4.4.2 Model and Methodology - BMI as a Measure of Heath

This chapter uses the same models in estimating the effect of BMI on employment for Chinese working-age adults: the pooled probit model, the RE probit model, and the FE logit model. However, the health indicator, BMI, is added to the employment equation.

[75] For example, respondents in survey 1991 are roughly aged between 30 and 33 if they were born between 1958 and 1961, and are roughly aged between 45 and for respondents in survey 2006.

Based on the same waves of the CHNS as in Section 4.4.1, the sample in this sub-section includes 14,206 individuals pooled across the six waves (for each i, $t = 1, \dots, T$, where $T \leq 6$ in a case of the unbalanced panel), that is, a total of 35,791 observations. The sample is also split by sex, by urban/rural area and by young/older working-age adults. Separate analyses undertake for 7,275 males and 6,931 females (19,248 and 16,543 observations, respectively), for 4,597 urban and 9,609 rural residents (10,114 and 25,677 observations, respectively), and for 9,571 young and 4,635 older working-age adults (27,584 and 8,207 observations, respectively).

This section uses various classifications of BMI to measure overweightness/obesity: (i) the WHO classification of BMI; (ii) the C-WGOC[76] classification of BMI; (iii) the Bao *et al.* (2008)'s classification of BMI; and (iv) the Tuan *et al.* (2008)'s classification of BMI. The WHO classification of BMI is an international BMI classification, therefore it is used as the primary health indicator in estimating the effect of BMI on employment. The other three types of classification of BMI are proposed specifically for the Chinese population, therefore their effects on employment are also examined as a comparison. Recall from Chapter 3 that four dummy variables are created in accordance with the WHO classification of BMI: *uweight_who* (BMI<18.5), *nweight_who* (BMI 18.5-25), *oweight_who* (BMI 25-30), and *obese_who* (BMI>=30). Four dummy variables are created based on the C-WGOC criteria: *uweight_cwgoc* (BMI<18.5), *nweight_cwgoc* (BMI 18.5-24), *oweight_cwgoc* (BMI 24-28), and *obese_cwgoc* (BMI>=28).

As shown in Table 4.2a, the average employment rates by BMI categories using the C-WGOC classification are similar to those using the WHO classifications. Individuals with normal weight have the highest average employment rate for all the sample groups, except for female and the older working-age sample which underweight individuals have the largest average employment rate. Obese individuals have the lowest average employment rate for all the sample groups except for men.

[76] The C-WGOC is the abbreviation for the Cooperative Meta-analysis Group of the Working Group on Obesity in China.

Bao *et al.* (2008) and Tuan *et al.* (2008) only propose the optimal BMI cutoff points for overweight[77]. In other words, individuals cannot be distinguished by whether they are underweight, healthy weight, overweight or obese. Therefore, instead of creating four dummy variables, only one dummy variable is created for each of these two types of BMI classification, that is, *oweight_bao* (take the value of one if male BMI>=24.5 or female>=24.7, and zero otherwise) and *oweight_tuan* (take the value of one if male BMI>=22.5 or female>=23.5, and zero otherwise), respectively. For the purpose of comparison, two additional dummy variables are also created based on the WHO and the C-WGOC classifications, respectively, i.e. *oweight_who2* (take the value of one if BMI>=25, and zero otherwise) and *oweight_cwgoc2* (take the value of one if BMI>=24, and zero otherwise). As mentioned earlier, since the bivariate probit model requires that the endogenous variable takes a binary form, the dummy variable, *oweight_who2*, is therefore used as a combined measure of obesity.

The other explanatory variables included in the employment equation of this analysis are the same as those in Section 4.4.1. Following Morris (2004) and Morris (2007), the dummy variables for SRHS are also added to the employment equation as an additional health indicator. Theoretically speaking, if an individual health can be correctly controlled for in the employment equation, and being underweight or being overweight still positively or negatively associated with employment, then there is evidence of discrimination against people who have unhealthy body weight. However, it is only true under an assumption that SRHS completely captures the impact of health on productivity.[78] In order to explore this issue, separate models with and without controlling for SRHS are estimated in this analysis.

In taking account of the potential endogeneity of overweight or obesity, some studies use information on BMI of interviewee's parents, siblings or biological children (Cawley, 2004; Norton and Han, 2007 and 2008; and Greve, 2008), the degree of physical activities undertaken by the respondent (Morris, 2004), an area level of BMI (Morris, 2007). Information on physical activities as instruments for overweight since information on physical activities is not surveyed in all the selected waves in

[77] The overweight in Bao*et al.* (2008) and Tuan *et al.* (2008) also include obesity.

[78]Besides using the self-reported health, both Morris (2004) and Morris (2007) further use acute illness, longstanding illness and mental health as explanatory variables for individuals' health.

the CHNS so is not utilized.[79] Further, no information is available on the BMI of interviewee's parents, siblings or biological children. Other variables that can be considered as potentially suitable instruments for overweight include dummy variables for individual daily average nutrient intakes. Additionally, some area-based health measures are also created and are considered as potentially suitable instruments, including *ivbmi_t4* and *ivoweight_t4*. The variable, *ivbmi_t4*, is defined as the average value of BMI in the area in which the respondent lives. The variable, *ivoweight_t4*, is defined as the prevalence of overweight (BMI above 25) in the area in which the respondent lives. Further investigation suggests that none of these variables can be used as suitable instruments for overweightness. For example, *ivbmi_t4*, *ivoweight_t4*, *henergy*, *hfat* and *hcarbo* cannot be used as instruments for overweightness since they are strongly correlated with employment for almost all the sample groups.

In this chapter, the dummy variable, *famine*, is used as an instrument for overweightness in estimating the effect on employment. Information on protein intake is also used as an additional instrument for overweight. It is a dummy variable, *hprotein*, which takes value of one if an individual's daily average protein intake is above the WHO/FAO criteria.[80] Table 4.2b provides descriptive statistics for this variable. It shows that the percentage of people whose daily average protein intake is above WHO/FAO criteria accounts for 10% for the total sample. Urban residents have the highest percentage which accounts for 15%, whereas rural residents have the lowest (7%).

4.5 Empirical Results

4.5.1 The Impact of SHRS on Employment

The marginal effects of SRHS impact on employment using the pooled probit (column 2), the RE probit (column 3) and the FE logit models (column 4) are shown in Tables 4.3-4.6. A total of 12,752 individuals are dropped from the fixed effects specification since there is no variation in y_{it} over t, leading to a loss of 31,291 observations of the original 40,723 observations. Additionally, the time-invariant

[79] The information on physical activities is not surveyed in CHNS, 1991, and 1993.
[80] More detailed information of the variables on individual daily nutrient intakes is provided in Chapter 3.

explanatory variables, *male*, *urban*, and the region dummy variables cannot be estimated in this specification. Large declines of the observations are also found in the FE logit estimates of Tables 4.4-4.6, thus the discussion of the results will not focus on the FE method. Allison (2009) warns that the FE coefficients might be imprecise and have large standard errors in a case where within-individual variation is small relative to the between-individual variation. Cai (2010) also argues that although the FE model can be estimated, the reliability of the results is still questionable due to the large number of individuals excluded. Table 4.7a and Table 4.7b show the transition probabilities of employment and of SRHS, respectively. In Table 4.7a, there is considerable persistence of work status from wave to wave. Indeed, 51% of those who did not work in one wave also did not work at next wave, while 91% of those who worked during one wave also worked at the next wave. In Table 4.7b, a considerable persistence is also found in good health status. 62% of those who reported good health status one wave also reported the same health status at next wave. As for the transition probabilities of the other three levels of health status, 26%/32%/23% of those who reported excellent/fair/poor status one wave also reported the same health status at next wave.

The Impact of SHRS on Employment - Total Sample

The marginal effects of the employment model for the whole sample are shown in Table 4.3. The coefficient estimates of health variables are the same in sign but are smaller in magnitude across the pooled probit and the RE probit model, suggesting that accounting for unobservable factors reduce influence of health. Additionally, the FE logit estimates are almost the same in sign and are similar in magnitude comparing to the pooled probit estimates. Table 4.3 shows that all the SRHS variables are significantly different from zero at the 1% level. The likelihood-ratio test of the significance of ρ_0 (H_0: $\rho_0 = 0$) shown at the bottom of Table 4.3 suggests that the null hypotheses of ρ_0 can be rejected (i.e. p-value = 0.00), therefore the results from the RE probit model are preferred to the pooled probit model.

According to the third column of Table 4.3 (the RE probit estimates), holding other explanatory variables at their reference groups, individuals who have good health status are 9.0 percentage points more likely than individuals who have poor health status to be employed. Individuals who have excellent/fair health status are 8.6/8.8

percentage points more likely than individuals who have poor health status to be employed. This indicates that better health is positively associated with the probability of being employed for the Chinese population. This finding is consistent with that of the most existing studies, for example, Stern (1989) for the US population and that in Garcia-Gomez and Lopez-Nicolas (2006), the Average Treatment Effect on the Treated (ATET) based the matching and differences in differences method shows that suffering a health shock[81] reduce the probability of being employed by around 5 percentage points for the Spanish working-age population[82]. However, none of these estimates are directly comparable. This is mainly because authors use different methods or different measures of employment. From Table 4.3 it also can be seen that there is a positive effect of any health status relative to the poor health status, and these marginal effect estimates are similar in terms of magnitude. This result, which is consistent with the statistical analysis in Table 4.1a, suggests that changes between excellent, good and fair health status would not change the probability of being employed.

As for the marginal effects from other non-health explanatory variables, men are 6.0 percentage points more likely than women to be employed. Individuals age 26-35 and age 36-45 are 5.0 and 5.7 percentage points more likely to be employed than individuals age 16-25, respectively. Individuals with technical or vocational degree qualifications (*techvoc*) and with university or higher degree qualifications (*university*) are 6.0 and 8.0 percentage points more likely than individuals with no qualifications, respectively. Urban residents are 9.2 percentage points less likely to be employed than rural residents. There are also some regional differences in employment within China. For individuals who are resident in Jiangsu and Guizhou, relative to individuals who are resident in Liaoning, the predicted probability of being employed will increase by approximately 6 percentage points. Married people are 4.4 percentage points more likely than people who are single, divorced, widowed or separated to be employed. The presence of children in the household has small effect on employment status. Only *children2m* (having two or more children in the household) has a statistically significant effect on employment, with a marginal

[81] Recall from Section 4.3 that in their study a health shock is defined as a drop of the SRHS from 'very good' or 'good' to 'fair', 'bad', or 'very bad'.
[82] See Table 2 in Garcia-Gomez and Lopez-Nicolas (2006), p. 1003.

effect of 1.0 percentage points. The year effects in Table 4.3 reflect the declining trend in employment as that identified in Figure 3.3 (Chapter 3).

The Impact of SHRS on Employment - Men versus Women

The marginal effects for men and women estimated separately are shown in Table 4.4. The estimated effects of the explanatory variables for both genders are the same in sign across pooled probit model and RE probit model, again after controlling for unobservable individual heterogeneity RE probit estimates are smaller in magnitude comparing to the pooled probit. The likelihood-ratio test of the significance of ρ_0 (H_0: $\rho_0 = 0$) shown at the bottom of Table 4.4 also suggests that the null hypotheses of ρ_0 can be rejected, therefore the estimates of the RE probit model are preferred to the pooled probit model for both men and women. The RE probit estimates for both men and women in Table 4.4 show that all the SRHS variables are significantly different from zero at the 1% level. For males, being in excellent, good and fair health increases probability of being employed by approximately 12.0 percentage points relative to those in poor health. Again, the marginal effect estimates of these health statuses are similar to each other in terms of magnitude for both genders. However, the positive effects for women are smaller in magnitude (3.8-4.7 percentage points) than those for men, which indicate that the positive effect of better self-reported health status on the predicted probability of being employed is larger for men than for women. This conclusion is similar to that in Garcia-Gomez *et al.* (2010) who also use panel data to examine the SRHS on employment for the UK working-age males and females. Unfortunately, the coefficient estimates are not directly comparable due to the different methods used. However, the empirical results for the Australia population reach the opposite conclusion. For example, in Cai and Kald (2006), a deterioration of health (from one level to a lower level of health status) reduces the probability of labour force participation by 0.5-0.7 percentage points for Australian men, and by 1.6-1.8 percentage points for Australian women, respectively[83]. Further, Cai (2010) finds a deterioration of health (from one level to a lower level of health status) reduces the probability of labour force

[83]See Table 5 in Cai and Kald (2006), p. 254.

participation by 1.6-2.4 percentage points for Australian men, and by 3.3-3.7 percentage points for Australian women, respectively[84].

Some of the marginal effect estimates of the other explanatory variables are also reported in Table 4.4-Table 4.6 because they are found to have different effects on employment between the sample groups. In Table 4.4, the predicted probability of being employed for males who have held the university or higher degree qualification/technical or vocational degree qualification is increased by 6.4/5.1 percentage points comparing to those who have no qualifications. For women, the predicted probability of being employed for holding the university or higher degree qualification/technical or vocational degree qualification is increased by 12.7/8.3 percentage points comparing to those who have no qualifications. This suggests that educational attainment is more important for women than for men in determining employment in China. Furthermore, married men are more likely to be employed than the base category. But marital status does not have statistically significant effect on the probability of being employed for women.

The Impact of SHRS on Employment - Urban versus Rural

The marginal effects for the urban and the rural sample are presented in Table 4.5. The result of the likelihood-ratio test of the significance of ρ_0 (H_0: $\rho_0 = 0$) rejects the null hypotheses of ρ_0, which implies that the estimates of the RE probit model are preferred to the pooled probit model for both urban and rural sample. According to the RE probit estimates, all the SRHS variables for both urban and rural sample are statistically significant at the 1% significance level. However, these positive effects for rural sample are smaller in magnitude than those for urban sample. For the urban residents, being in excellent and fair health increase the probability of employment by 12.5 and 13.0 percentage points relative to the poor health counterparts, respectively. In addition, being in good health are 14.0 percentage points more likely to be employed than individuals with poor health status. For the rural residents, being in excellent, good, or fair health are about 7.3 percentage points more likely to be employed than those with poor health status. This result is consistent with expectations and suggests that health (as measured by the SRHS) is more important in terms of employment for the urban citizens than for the rural residents. As

[84]See Table 5 in Cai (2010), p. 86.

discussed earlier, this finding might, in part, be explained by the increasing employment pressure in the urban area, plus the fact that a large share of labour force in rural area is self-employed.

The predicted probability of being employed for the urban residents, who have held the university or higher degree qualification, is 21.3 percentage points higher compared to those who have no qualifications. And those with technical or vocational degree qualification are also more likely to be employed relative to those who have no qualifications. Whereas for the rural residents, the predicted probability of being employed for holding the university or higher degree qualification is increased by less than 5.0 percentage points compared to those who have no qualifications. Therefore, having higher level of qualifications is much more important in terms of employment for the urban residents than for the rural residents. This probably also a consequence of the nature of employment opportunity between the areas discussed in Chapter 3.

The Impact of SHRS on Employment - Young versus Older

Table 4.6 shows the marginal effects for the young (aged 16-45) and the older (aged 46-59) working-age Chinese population. The likelihood-ratio test of the significance of ρ_0 (H_0: $\rho_0 = 0$) shown at the bottom of Table 4.6 suggests that the null hypotheses of ρ_0 can be rejected, therefore the results from the RE probit analysis are preferred to the pooled probit analysis for both young and older sample. According to the RE probit estimates, all the SRHS variables for both the young and the older working-age Chinese are positive and are statistically significant at the 1% significance level. For the young working-age sample, individuals who have excellent health status are 6.9 percentage points more likely to be employed than those who have poor health status; and individuals who have good and who have fair health status are 7.2 and 7.7 percentage points more likely to be employed than those who have poor health status, respectively. For the older working-age sample, individuals who have excellent health status are 11.8 percentage points more likely to be employed than those who have poor health status; and individuals who have good and who have fair health status are 13.8 and 12.0 percentage points more likely to be employed than those who have poor health status, respectively. These results suggest that, health, as measured by the SRHS, is a relative more important factor in terms of employment for the older than for the young working-age Chinese. This finding is consistent with that in Cai and Kald (2006) and that in Zhang *et al.* (2009)[85]. For example, in Cai and Kald (2006), a deterioration of health (from one level to a lower level of health status) reduces the probability of labour force participation by 0.5-0.7 percentage points for Australian men aged 15-49 (1.6-1.8 percentage points for women aged 15-49), and by 4.6-7.0 percentage points for Australian men aged 50-64 (5.8-8.8 percentage points for women aged 50-60), respectively[86].

There are also some differences in the effect of education by age. Young individuals with a university or higher degree qualification/technical or vocational degree qualification is increased by 4.3/6.6 percentage points comparing to those who have no qualifications. For the older working-age sample, the predicted probability of being employed for both holding the university or higher degree qualification, and holding technical or vocational degree qualification are increased by 10.1 percentage points comparing to those who have no qualifications. This suggests that having higher level qualifications has less influence on employment for the young than for the older working-age Chinese. Recall from Chapter 2 that the Cultural Revolution began in 1966, where almost all schools and universities were closed. Although schools and universities were reopened in the early 1970s, students spent much of their school time studying politics instead of receiving formal education. During the time when schools were closed, students were given diplomas although they did not receive formal education. In this study, more than 90% of individuals in the older working-age group have received degree qualification before the Cultural Revolution.[87] In other words, for the young working-age sample, the marginal effects for education may be affected by the Cultural Revolution.[88]

Addressing the Potential Endogeneity of SRHS

The IV regression results are shown in Table 4.8. There is a statistically significant effect of at least one of instruments on having poor health status for all the sample groups. This implies that both instruments, in broader terms, satisfy the non-

[85] In Zhang *et al.* (2009), the larger health effect on labour force participation for older working-age group is found only for men (no statistically significant effect for women). However, the coefficient estimates are not directly comparable since a different measure of health is used.
[86] See Table 5 in Cai and Kald (2006), p. 254.
[87] Assuming people get degree level qualifications at age of 20.
[88] However, further analysis needs to be done in terms of estimating the return for those in school during 1966-1970.

weakness requirement (Morris, 2007), although the instrument, *famine*, does not have a statistically significant effect on having poor health status for men, the urban, and the older working-age sample. The area level SRHS is negatively associated with individual self-reported health status for all the sample groups.

The top part of Appendix 4.1 shows the impact of the instruments on employment when applying SRHS to measure health. Neither of the two instruments (or least one of the instruments) has a statistically significant effect on employment for all the sample groups, indicating that there is no clear relationship between the instruments and the probabilities of being employment. As can be seen from Table 4.8, a Wald test this study fails to reject the null hypothesis ($H_0: \rho = 0$) of exogeneity of SRHS for all the sample groups except for women. Assuming the instruments are valid, it suggests that for all the sample groups except for women, having poor health status is exogenously related to the probability of being employed, and the RE probit estimates remain preferable. For women, the bivariate probit estimate is preferable, and the marginal effect estimate implies that women who have poor health status are about 21.9 percentage points less likely to be employed than those who have better health status, which is considerably large (compared to marginal effect estimate based on the RE probit).

4.5.2 The Impact of BMI on Employment

Based on the WHO classification and the C-WGOC classification criteria, the marginal effects of the impact of BMI on employment from the pooled probit, the RE probit and the FE logit models are shown in Tables 4.9-4.12. Moreover, by applying the four cutoff points of BMI for overweight[89], the marginal effects from the preferred model (the RE probit model) for all the sample groups are shown in Table 4.13. The FE logit estimation results in Table 4.9 show that 11,981 individuals are dropped since there is no variation in y_{it} over t, leading to a loss of 27,754 observations of the original 35,791 observations. Similarly, large declines of the observations are also found in the FE logit estimates of Tables 4.10-4.12. Table 4.14a and Table 4.14b show the transition probabilities of BMI categories based on the WHO and the C-WOGOC classification, respectively. In both tables, there is considerable persistence of BMI categories from wave to wave. For instance, in

[89] The overweight used here includes obesity.

Table 4.14a, 44.1%, 87.3%, 71.8% and 70.9% of those who are underweight, normal weight, overweight and obese in one wave also in the same BMI category at the next wave, respectively. Recall from Section 4.5.1 that the reliability of the FE results might questionable due to the large number of individuals excluded, thus the results discussion will not focus on the FE method. The likelihood-ratio tests of the significance of ρ_0 (H_0: $\rho_0 = 0$) in Tables 4.9-4.12 suggest that the null hypotheses of ρ_0 can be rejected, therefore the results from the RE probit analysis are preferred to the pooled probit analysis for all the sample groups.

The Impact of BMI on Employment - Total Sample

In Table 4.9 (WHO classification), individuals who are underweight are 1.4 percentage points less likely than normal weight individuals to be employed, and the result is significantly different from zero at the 5% level. Individuals who are overweight and who are obese are 1.8 and 2.5 percentage points less likely than normal weight individuals to be employed, respectively. The former is significantly different from zero at the 1% level, and the latter is significantly different from zero at the 5% level. The results indicate that either being underweight or being overweight/obese is negatively associated with the probability of being employed for the Chinese population. Although the marginal effect estimates of the three BMI categories based on the C-WGOC criteria are larger than those using the WHO criteria, the changes are small in terms of magnitude. Compared to the result in Table 4.3, after controlling for BMI the statistically significant and positive effect of the SRHS on employment is reduced, implying that BMI accounts for some of the influence of SRHS.

The Impact of BMI on Employment - Men versus Women

The marginal effects for men and women are separately shown in Table 4.10. For men (WHO classification), being underweight are 2.7 percentage points less likely to be employed than those with normal weight, which is significantly different from zero at the 1% level. Being overweight or being obese does not have a statistically significant effect on employment for men. However, being overweight is negatively associated to employment when applying the C-WGOC classification. Overweight women (WHO classification) are 2.9 percentage points less likely to be employed than normal weight women (significantly different from zero at the 1% level). Obese women are 4.3 percentage points less likely to be employed than normal weight women (significantly different from zero at the 5% level). However, being underweight does not have a statistically significant effect on employment for women. The results in Table 4.10 suggest that being underweight is more important in terms of employment for men, while being overweight or obese has influence for women. This finding is consistent with Johansson *et al.* (2009) who find that being underweight is significantly and negatively associated with employment for Finnish men (with a marginal effect of −40.9 percentage points[90]), while either being overweight or being obese is negatively associated with employment for Finnish women (with marginal effects of -6.7 and -9.9 percentage points, respectively[91]). Based on the same cutoff points of BMI (C-WGOC classification) as those in Pan and Qin (2010), this study finds that either being underweight or being overweight has a statistically and negative impact on employment for Chinese women. Additionally, Pan and Qin (2010) find that being obese is negatively associated with employment for Chinese women (with a marginal effect of -7.5 percentage points).

The Impact of BMI on Employment - Urban versus Rural

The marginal effects for the urban and the rural sample are separately presented in Table 4.11. For the rural sample, being underweight is negatively associated with the probability of being employed. In addition, overweight rural residents (WHO classification) are 1.6 percentage points less likely to be employed than the normal weight rural residents. For urban sample, however, none of the BMI categories has a statistically significant effect on the probability of being employed. This suggests that health, as measured by BMI, is strongly related to employment for the rural sample but not for the urban sample. In addition, being obese does not have a statistically significant effect on employment for rural residents when applying the WHO classification criteria, but it becomes statistically significant (with a marginal effect of −2.4 percentage points) when applying the C-WGOC classification criteria. As mention in Section 4.2, obesity can be sign of laziness, while being underweight is associated with as lack of strength. In rural areas, agricultural (physical work) is

[90]See Table 10 in Johansson *et al.* (2009), p. 43.
[91]See Table 9 in Johansson *et al.* (2009), p. 42.

still dominant, therefore rural residents who have unfavourable body shape may be more likely to suffer discrimination than urban residents.

The Impact of BMI on Employment - Young versus Older

Table 4.12 shows the marginal effects for the young (aged 16-45) and the older (aged 46-59) working-age sample. For the young working-age sample, individuals who are underweight (WHO classification) are 1.9 percentage points less likely to be employed than those who are normal weight, while individuals who are overweight or who are obese do not have statistically significant effects on employment. For the older working-age sample, only being overweight (WHO classification) has a statistically significant and negative effect on employment (with a marginal effect of −3.7 percentage points). Being obese does not have a statistically significant effect on employment for the older sample when applying the WHO classification criteria, whereas it becomes statistically significant (with a marginal effect of −4.1 percentage points) when applying the C-WGOC classification criteria. The results suggest that being underweight is a more important influence for the young than for the older working-age Chinese population, whereas being overweight is more important for the older. This finding is inconsistent with that in Pan and Qin (2010). By splitting the total sample into eight age-gender groups, they find some obesity effects for both genders in the two young working-age groups, but there is no statistically significant effect for both genders in the two older working-age groups.

The Impact of BMI on Employment -using Different Cutoff points of BMI for Overweight

Table 4.13 compares the four measures of BMI discussed in Section 4.4.2. For the total sample, being overweight is negatively associated with the probability of being employed using all the cutoff points of BMI. For men, being overweight does not have a statistically significant effect on employment when applying the WHO criteria, while it has a statistically significant effect on employment when applying other three measures (and the corresponding marginal effects are similar in terms of magnitude). For the young working-age sample, being overweight is negatively associated with the probability of being employed using both the C-WGOC's and the Tuan *et al.* (2008)'s cutoff points (again the corresponding marginal effects are similar in terms of magnitude), but no effects on employment using the WHO's and the Bao *et al.* (2008)'s cutoff points. Table 4.13 also shows that the effect of BMI on employment is largely unchanged when comparing alternative BMI classifications for Chinese population with using the WHO (international) classification. Therefore using the WHO classification of BMI as the prime health indicator is appropriate in estimating the relationship between BMI and employment for Chinese population.[92]

The Impact of BMI on Employment - without versus with Controlling for SRHS

As mentioned in Section 4.2, the effect of body weight on employment can work through the health channel and (or) due to the presence of the weight-based discrimination. Table 4.15 shows the marginal effects estimates both with and without controlling for SRHS. For comparison, the results including the variables for SRHS which are presented in the bottom part of Table 4.15 are the same to those in Tables 4.9-4.12. The inclusion of SRHS has a small influence on the marginal effects of BMI for all the sample groups, suggesting that the relationship between BMI and employment reflects discrimination against people who have unhealthy body weight in China. This, however, is only true if SRHS completely captures the impact of health on productivity is maintained.

Addressing the Potential Endogeneity of Overweight

The IV regression results are shown in Table 4.16. There is a statistically significant effect of at least one of the instruments on overweight for all the sample groups except for the urban, and the older working-age sample. This implies that both the instruments, in broader terms, satisfy the non-weakness requirement. However, for both the urban and the older working-age samples, neither of the two instruments has a statistically significant effect on overweightness. Therefore, neither *famine* nor *hprotein* can be used as a suitable instrument (or suitable instruments when including them simultaneously) in examining the potential endogeneity of obesity when constrained to these sample groups.

The bottom part of Appendix 4.1 shows the impact of the instruments on employment when applying overweightness as a measure of health. Neither of the two instruments has a statistically significant effect on employment for all the sample

[92] It is worth to note that this robustness test of using different BMI classifications ignores the cutoff points for underweight and obese.

groups except for the rural and the older working-age sample (which at least one of the two instruments does not has a statistically significant effect on employment), indicating that there is no relationship between the instruments and employment. As can be seen from the Wald statistics in Table 4.16, the null hypothesis ($H_0: \rho = 0$) of exogeneity of obesity cannot be rejected for the young working-age sample but it is rejected for men, women, the rural and the total sample. Therefore, assuming the instruments are valid, for men, women, the rural and the total sample, being overweight is endogenously related to employment, and the bivariate probit estimates are preferable. For the young working-age sample, the RE probit estimates remain preferable. For men, those overweight are 7.6 percentage points less likely to be employed than those with normal weight. For women, being overweight are 14.2 percentage points less likely to be employed than those with normal weight. Based on an IV method, Morris (2007) finds that there is a statistically significant and negative relationship between obesity (BMI over 30) and employment for both British men (with a marginal effect of −8.4 percentage points) and British women (with a marginal effect of 21.3 percentage points). Although the coefficient estimates are not directly comparable since Morris (2007) uses different measures of BMI category[93], the result still suggests that the negative effect of BMI on employment is larger for women than for men, which is consistent to the finding in Morris (2007).

4.6 Conclusions

This chapter contributes to the literature by examining the relationship between health and employment for the working-age population in China. This chapter adopts two alternative measure of health one based on self-reported health status and the other on BMI which is based on various cutoff points for overweight and obesity, respectively.

Besides using the pooled probit model, this chapter also employs both the RE probit model and the FE logit to take into account the panel nature of the data and finds that failure to control for individual unobserved factors would bias the estimates of the effect of health on employment. Comparing with the marginal effects estimates after pooled probit, estimates after RE probit are similar in terms of sign but smaller in terms of magnitude.

[93] The author classifies individuals' weight as either obese or not obese.

There is a statistically significant and positive relationship between having better SRHS and the probability of being employed for the Chinese working-age adults, which is consistent with the international literature. This finding is also in line with the target of the Seventeenth National Congress of the CPC on improving the health of the whole nation, thereby providing empirical evidence to the policy makers (such as government and institutions) in making reasonable and effective health improvement policies. The estimated results suggest that any changes among the excellent, good and fair health status would not significantly change the probabilities of being employed. For example, according to Table 4.3, improving health status from poor to fair would increase the employment probability by 8.8 percentage points, which is similar to the marginal effect (with a value of 9.0 percentage points) of changing from poor to good health status on the employment probability. Therefore, further government policy on improving population health should focus on individuals who have poor health status. The effect of SRHS on employment is larger for men than for women, which is consistent with the finding in Garcia-Gomez *et al.* (2010). As expected, the effect for urban residents is found to be larger than the effect for their counterparts. Moreover, the finding suggests that the effect of SRHS on employment is larger for the older working-age population than for the young working-age population, which is consistent with the international literature.

Having an unhealthy BMI is found to be negatively associated with employment for the Chinese population. Although the increase in overweight or obesity in China is not as rapid as in USA, UK and Australia, there is a growing trend of overweight and obesity in China could have a negative effect on the Chinese labour market. Furthermore, controlling for the effect of SRHS only slightly reduce the marginal effects of BMI on the probability of being employed, suggesting that there is evidence of discrimination against people who have unhealthy body weight in China, at least if the assumption that SRHS completely captures the impact of health on productivity is maintained. There is a statistically significant and negative relationship between being underweight and employment for men. For women either being overweight or being obese is negatively correlated with employment. Therefore, there is evidence on discrimination against underweight Chinese men and overweight/obese Chinese women in employment decisions. Furthermore, while either being underweight or being overweight is negatively correlated with employment for rural residents, there is no effect on employment for the urban sample.

Based on the Wald statistics, this chapter fails to reject the null hypothesis of exogeneity of SRHS for all the sample groups except for women. In contrast, the Wald tests of exogeneity of overweight suggest that being overweight is endogenously related to employment for the total sample, the rural residents and for both genders, but not for the young working-age sample. Failure to account for the endogeneity of overweight is found to be underestimate its impact on employment in China, which is consistent with the finding in Morris (2004) and (2007) based on UK data.

There are some shortcomings in this chapter which are worth noting. Firstly, this chapter fails to find suitable instruments in addressing the potential endogeneity of overweight for the urban or the older working-age sample. Secondly, it is not possible to fully test the validity of exclusion restriction for instrumental variable estimation. To the author's knowledge there is still no formal test for instrument validity for the bivariate probit model. Thirdly, the health measures used in this chapter are not without their drawbacks. For example, the nature of self-reported health status might introduce justification bias in employment analysis, while even the objective BMI measure cannot efficiently distinguish between muscle and fat, thereby producing measurement error. Fourthly, the dependent variable used in this chapter simply relies on a binary choice, i.e. either being employed or not. On one hand, the employment status cannot distinguish self-employed, paid employed, job seeking and being out of the labour force. On the other hand, the effect of health on hours of work is not captured. For example, being in unfavourable health status or BMI category might reduce an individual hours of work, but they may remain employed. Accordingly, further analysis in addressing the above issues is needed.

Table 4.1a Average Employment Rate by Self-reported Health Status (pooled waves, 1991-2006)

Variables	Average Employment Rate (%)[1]						
	Total (N=40,723)	**Men (N=22,411)**	**Women (N=18,312)**	**Urban (N=11,407)**	**Rural (N=29,316)**	**Young (N=31,801)**	**Older (N=8,922)**
Self-reported Health Status							
excellent	87	90	83	83	89	88	82
(Sample Size: %)	(16%)	(17%)	(15%)	(18%)	(15%)	(17%)	(11%)
good	89	92	86	85	91	90	86
(Sample Size: %)	(60%)	(60%)	(60%)	(58%)	(61%)	(62%)	(53%)
fair	86	89	83	79	89	88	81
(Sample Size: %)	(21%)	(20%)	(22%)	(21%)	(21%)	(19%)	(30%)
poor	74	72	77	60	79	77	71
(Sample Size: %)	(3%)	(3%)	(3%)	(3%)	(3%)	(2%)	(6%)
Total	88	90	85	82	90	89	83
(Sample Size: 100%)	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)

Source: Author's calculations using CHNS data.

[1] Average employment rate of each health status category (%) = $\frac{\text{The number of people who are presently working in each health status category}}{\text{Total number of people in that health status category}} \times 100\%$

Table 4.1b Summary Statistics (pooled waves, 1991-2006)

Variables	Mean (Standard Deviation)						
	Total	Male	Female	Urban	Rural	Young	Older
Other Explanatory Variables							
male (%)	0.55 (0.50)	-	-	0.54 (0.50)	0.55 (0.50)	0.50 (0.50)	0.75 (0.44)
age (in years)	36.68 (10.74)	38.35 (11.65)	34.63 (9.11)	37.41 (10.27)	36.40 (10.91)	-	-
age16_25 (%)	0.20 (0.40)	0.18 (0.39)	0.22 (0.41)	0.16 (0.37)	0.21 (0.41)	-	-
age26_35 (%)	0.27 (0.45)	0.25 (0.43)	0.31 (0.46)	0.29 (0.45)	0.27 (0.44)	-	-
age36_45 (%)	0.31 (0.46)	0.27 (0.44)	0.35 (0.48)	0.33 (0.47)	0.30 (0.46)	-	-
age46_59 (%)	0.22 (0.41)	0.30 (0.46)	0.12 (0.33)	0.22 (0.41)	0.22 (0.41)	-	-
noedu (%)	0.15 (0.35)	0.10 (0.31)	0.20 (0.40)	0.06 (0.23)	0.18 (0.39)	0.12 (0.33)	0.24 (0.43)
primary (%)	0.22 (0.42)	0.23 (0.42)	0.22 (0.42)	0.11 (0.31)	0.27 (0.44)	0.20 (0.40)	0.30 (0.46)
lowermid (%)	0.39 (0.49)	0.41 (0.49)	0.36 (0.48)	0.36 (0.48)	0.40 (0.49)	0.43 (0.49)	0.25 (0.43)
uppermid (%)	0.16 (0.36)	0.17 (0.37)	0.14 (0.35)	0.26 (0.44)	0.11 (0.32)	0.17 (0.37)	0.12 (0.32)
techvoc (%)	0.05 (0.22)	0.05 (0.22)	0.05 (0.21)	0.12 (0.32)	0.03 (0.15)	0.05 (0.22)	0.05 (0.22)
university (%)	0.03 (0.18)	0.04 (0.20)	0.03 (0.16)	0.09 (0.29)	0.01 (0.11)	0.03 (0.18)	0.04 (0.19)
urban (%)	0.28 (0.45)	0.28 (0.45)	0.28 (0.45)	-	-	0.28 (0.45)	0.28 (0.45)
married (%)	0.79 (0.40)	0.78 (0.41)	0.81 (0.39)	0.80 (0.40)	0.79 (0.41)	0.75 (0.43)	0.94 (0.24)
child0 (%)	0.41 (0.49)	0.45 (0.50)	0.36 (0.48)	0.45 (0.50)	0.39 (0.49)	0.33 (0.47)	0.71 (0.45)
child1 (%)	0.37 (0.48)	0.35 (0.48)	0.40 (0.49)	0.44 (0.50)	0.35 (0.48)	0.41 (0.49)	0.23 (0.42)
child2m (%)	0.22 (0.41)	0.20 (0.40)	0.24 (0.43)	0.11 (0.32)	0.26 (0.44)	0.26 (0.44)	0.06 (0.24)
Instruments[1]							
famine	0.07 (0.25)	0.06 (0.27)	0.08 (0.26)	0.08 (0.27)	0.06 (0.24)	0.07 (0.26)	0.04 (0.19)
ivsrhs_t4	2.89 (0.19)	2.89 (0.19)	2.89 (0.19)	2.91 (0.21)	2.88 (0.18)	2.89 (0.19)	2.89 (0.19)
No. of observations	40,723	22,411	18,312	11,407	29,316	31,801	8,922

Source: Author's calculations using CHNS data.

Notes: For full details of the other explanatory variables, see Table 3.5 in Chapter 3.

[1] The two instruments used in estimating the impact of SRHS on employment: *famine*, whether individuals were born in the famine period; *ivsrhs_t4*, average value of the SRHS in the area in which the respondent lives (1 = poor, 2 = fair, 3 = good, and 4 = excellent).

Table 4.2a Average Employment Rates by BMI Categories (pooled waves, 1991-2006)

BMI Categories	Average Employment Rate (%)[1]						
	Total (N=35,791)	Men (N=19,248)	Women (N=16,543)	Urban (N=10,114)	Rural (N=25,677)	Young (N=27,584)	Older (N=8,207)
WHO Classification[2]:							
uweight_who	87	87	87	80	90	87	89
(Sample Size: %)	(7%)	(6%)	(7%)	(7%)	(7%)	(7%)	(5%)
nweight_who	88	91	85	82	91	89	85
(Sample Size: %)	(75%)	(76%)	(75%)	(70%)	(77%)	(77%)	(70%)
oweight_who	83	87	78	79	85	87	75
(Sample Size: %)	(16%)	(16%)	(16%)	(21%)	(14%)	(14%)	(23%)
obese_who	80	87	73	77	82	83	74
(Sample Size: %)	(2%)	(2%)	(2%)	(2%)	(2%)	(2%)	(2%)
C-WGOC Classification[3]:							
uweight_cwgoc	87	87	87	80	90	87	89
(Sample Size: %)	(7%)	(6%)	(7%)	(7%)	(7%)	(7%)	(5%)
nweight_cwgoc	89	91	86	83	91	89	86
(Sample Size: %)	(67%)	(68%)	(67%)	(61%)	(77%)	(77%)	(70%)
oweight_cwgoc	84	87	80	80	86	87	77
(Sample Size: %)	(21%)	(21%)	(22%)	(26%)	(14%)	(14%)	(23%)
obese_cwgoc	81	87	75	79	83	85	74
(Sample Size: %)	(5%)	(5%)	(4%)	(6%)	(2%)	(2%)	(2%)
Total	87	90	84	82	90	89	83
(Sample Size: 100%)	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)

Source: Author's calculations using CHNS data.

[1] Average employment rate of each BMI category (%) = $\frac{\text{The number of people who are presently working in each BMI category}}{\text{Total number of people in that BMI category}} \times 100\%$

[2] Recall from Chapter 4 the WHO criteria of BMI cutoffs are: underweight (BMI<18.5); healthy weight (BMI 18.5-25); overweight (BMI 25-30); and obese (BMI>=30).

[3] Recall from Chapter 4 the C-WGOC criteria of BMI cutoffs are: underweight (BMI<18.5); healthy weight (BMI 18.5-24); overweight (BMI 24-28); and obese (BMI>=28).

Table 4.2b Summary Statistics (pooled waves, 1991-2006)

Variables	Mean (Standard Deviation)						
	Total	Male	Female	Urban	Rural	Young	Older
Instruments *hprotein*[1]	0.10 (0.29)	0.10 (0.29)	0.10 (0.30)	0.15 (0.36)	0.07 (0.26)	0.10 (0.30)	0.09 (0.29)
No. of observations	34,784	18,661	16,123	9,861	24,923	26,743	8,041

Source: Author's calculations using CHNS data.

[1] Dummy variable, *hyprotein*, indicating whether an individual's daily average protein intake is above the WHO/FAO criteria.

Table 4.3 The Impact of SRHS on Employment: Pooled Probit, RE Probit and FE Logit - The Total Sample

		The Total Sample	
Variables	**Pooled Probit**	**RE Probit**	**FE Logit**
SRHS Variables			
excellent	0.1042***	0.0858***	0.0869***
	(0.0125)	(0.0113)	(0.0296)
good	0.1077***	0.0902***	0.1106***
	(0.0120)	(0.0109)	(0.0272)
fair	0.1012***	0.0876***	0.1208***
	(0.0121)	(0.0109)	(0.0274)
Non-SRHS Variables			
male	0.0716***	0.0602***	
	(0.0038)	(0.0035)	
age26_35	0.0549***	0.0495***	0.1366***
	(0.0064)	(0.0058)	(0.0230)
age36_45	0.0646***	0.0566***	0.1596***
	(0.0065)	(0.0061)	(0.0323)
age46_59	0.0073	0.0065	0.0593
	(0.0076)	(0.0073)	(0.0460)
primary	0.0083	0.0062	-0.0143
	(0.0068)	(0.0059)	(0.0316)
lowermid	-0.0088	-0.0068	0.0147
	(0.0067)	(0.0059)	(0.0346)
uppermid	0.0135*	0.0173***	0.1143***
	(0.0077)	(0.0064)	(0.0406)
techvoc	0.0740***	0.0604***	0.1801***
	(0.0077)	(0.0063)	(0.0538)
university	0.1000***	0.0797***	0.2556***
	(0.0076)	(0.0059)	(0.0568)
urban	-0.0955***	-0.0924***	
	(0.0052)	(0.0049)	
heilongjiang	0.0294***	0.0279***	
	(0.0084)	(0.0080)	
jiangsu	0.0665***	0.0595***	
	(0.0077)	(0.0069)	
shandong	0.0015	0.0015	

	(0.0086)	(0.0083)	
henan	0.0365***	0.0357***	
	(0.0078)	(0.0073)	
hubei	0.0107	0.0107	
	(0.0084)	(0.0080)	
hunan	-0.0561***	-0.0558***	
	(0.0096)	(0.0092)	
guangxi	0.0559***	0.0494***	
	(0.0076)	(0.0071)	
guizhou	0.0628***	0.0577***	
	(0.0077)	(0.0069)	
married	0.0482***	0.0444***	0.1363***
	(0.0061)	(0.0057)	(0.0241)
children1	0.0019	0.0001	-0.0190
	(0.0039)	(0.0033)	(0.0133)
children2m	0.0158***	0.0102**	-0.0038
	(0.0054)	(0.0045)	(0.0194)
1993	-0.0028	-0.0010	0.0025
	(0.0028)	(0.0016)	(0.0115)
1997	-0.0352***	-0.0207***	-0.0967***
	(0.0035)	(0.0023)	(0.0216)
2000	-0.0854***	-0.0602***	-0.2734***
	(0.0045)	(0.0037)	(0.0356)
2004	-0.2114***	-0.1860***	-0.5736***
	(0.0060)	(0.0063)	(0.0304)
2006	-0.2159***	-0.1920***	-0.5896***
	(0.0063)	(0.0067)	(0.0304)
Pseudo-R^2	0.1713		
Likelihood ratio test (*p*-value)		1030.45 (0.00)	
Observations	40,723	40,723	9,432

Note: Only estimation results for marginal effects are reported. For dummy variables, the reported coefficient is the effect of changing from dummy from 0 to 1. Under the RE probit and the FE logit analysis, standard errors are in parentheses. Under the pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 4.4 The Impact of SRHS on Employment[1]: Pooled Probit, RE Probit and FE Logit - Men vs. Women

	Men			Women		
Variables	**Pooled Probit**	**RE Probit**	**FE Logit**	**Pooled Probit**	**RE Probit**	**FE Logit**
SRHS Variables						
excellent	0.1418***	0.1189***	0.1312***	0.0483***	0.0383**	0.0242
	(0.0163)	(0.0151)	(0.0370)	(0.0186)	(0.0168)	(0.0444)
good	0.1432***	0.1218***	0.1548***	0.0560***	0.0456***	0.0442
	(0.0158)	(0.0148)	(0.0336)	(0.0177)	(0.0159)	(0.0422)
fair	0.1326***	0.1160***	0.1605***	0.0537***	0.0467***	0.0588
	(0.0158)	(0.0148)	(0.0340)	(0.0178)	(0.0161)	(0.0427)
Non-SRHS Variables						
primary	0.0280***	0.0243***	0.0618	-0.0158	-0.0187**	-0.0882**
	(0.0090)	(0.0078)	(0.0399)	(0.0102)	(0.0094)	(0.0436)
lowermid	0.0055	0.0046	0.0360	-0.0310***	-0.0264***	0.0022
	(0.0090)	(0.0079)	(0.0444)	(0.0100)	(0.0092)	(0.0518)
uppermid	0.0226**	0.0248***	0.1525***	-0.0056	-0.0011	0.0655
	(0.0099)	(0.0083)	(0.0473)	(0.0120)	(0.0105)	(0.0674)
techvoc	0.0621***	0.0513***	0.1889***	0.0832***	0.0688***	0.1681*
	(0.0104)	(0.0086)	(0.0626)	(0.0117)	(0.0098)	(0.0942)
university	0.0774***	0.0636***	0.2000***	0.1265***	0.0996***	0.4095***
	(0.0105)	(0.0082)	(0.0690)	(0.0102)	(0.0085)	(0.0972)
married	0.0705***	0.0615***	0.1612***	0.0113	0.0135	0.0595
	(0.0077)	(0.0073)	(0.0283)	(0.0094)	(0.0088)	(0.0424)
Pseudo-R^2	0.1984			0.1543		
Likelihood ratio test (*p*-value)		431.01 (0.00)			519.30 (0.00)	
Observations	22,411	22,411	4,579	18,312	18,312	4,853

Note: Only estimation results for marginal effects are reported. For dummy variables, the reported coefficient is the effect of changing from dummy from 0 to 1. Under the RE probit and the FE logit analysis, standard errors are in parentheses. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: age group, urban/rural, region of residence, presence of children in the household, and year.

Table 4.5 The Impact of SRHS on Employment[1]: Pooled Probit, RE Probit and FE Logit - Urban vs. Rural

	Urban			Rural		
Variables	**Pooled Probit**	**RE Probit**	**FE Logit**	**Pooled Probit**	**RE Probit**	**FE Logit**
SRHS Variables						
excellent	0.1426***	0.1252***	0.0845	0.0893***	0.0734***	0.0833**
	(0.0263)	(0.0265)	(0.0529)	(0.0138)	(0.0118)	(0.0349)
good	0.1589***	0.1395***	0.1041**	0.0871***	0.0735***	0.1062***
	(0.0254)	(0.0255)	(0.0500)	(0.0133)	(0.0114)	(0.0321)
fair	0.1420***	0.1302***	0.1177**	0.0856***	0.0738***	0.1176***
	(0.0256)	(0.0257)	(0.0505)	(0.0133)	(0.0115)	(0.0323)
Non-SRHS Variables						
lowermid	0.0649***	0.0602***	-0.1759**	-0.0137**	-0.0087*	0.0531
	(0.0227)	(0.0230)	(0.0853)	(0.0063)	(0.0053)	(0.0362)
uppermid	0.1087***	0.1122***	-0.0636	-0.0081	-0.0001	0.1363***
	(0.0231)	(0.0231)	(0.0981)	(0.0079)	(0.0064)	(0.0463)
techvoc	0.2110***	0.1977***	0.0012	0.0179*	0.0157*	0.1987***
	(0.0229)	(0.0227)	(0.1147)	(0.0105)	(0.0084)	(0.0720)
university	0.2289***	0.2134***	0.1107	0.0608***	0.0476***	0.2042*
	(0.0232)	(0.0226)	(0.1179)	(0.0109)	(0.0067)	(0.1074)
children1	0.0242***	0.0221***	0.0286	-0.0103**	-0.0094***	-0.0435***
	(0.0083)	(0.0080)	(0.0241)	(0.0044)	(0.0035)	(0.0158)
children2m	-0.0057	-0.0059	-0.0072	0.0153***	0.0091**	-0.0117
	(0.0143)	(0.0137)	(0.0375)	(0.0055)	(0.0043)	(0.0223)
Pseudo-R^2	0.1702			0.1791		
Likelihood ratio test (*p*-value)		320.50 (0.00)			601.12 (0.00)	
Observations	11,407	11,407	3,210	29,316	29,316	6,222

Note: Only estimation results for marginal effects are reported. For dummy variables, the reported coefficient is the effect of changing from dummy from 0 to 1. Under the RE probit and the FE logit analysis, standard errors are in parentheses. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the model but not reported: gender, age group, marital status, region of residence, and year.

Table 4.6 The Impact of SRHS on Employment[1]: Pooled Probit, RE Probit and FE Logit - Young vs. Older

	Young (aged 16-45)			Older (aged 46-59)		
Variables	Pooled Probit	RE Probit	FE Logit	Pooled Probit	RE Probit	FE Logit
SRHS Variables						
excellent	0.0893***	0.0692***	0.0748*	0.1195***	0.1181***	0.0179
	(0.0157)	(0.0134)	(0.0426)	(0.0217)	(0.0226)	(0.0156)
good	0.0931***	0.0724***	0.0857**	0.1355***	0.1375***	0.0391**
	(0.0153)	(0.0131)	(0.0401)	(0.0194)	(0.0206)	(0.0160)
fair	0.0951***	0.0765***	0.1040**	0.1125***	0.1204***	0.0412**
	(0.0154)	(0.0132)	(0.0404)	(0.0196)	(0.0208)	(0.0164)
Non-SRHS Variables						
primary	0.0067	0.0047	-0.0059	0.0040	-0.0000	-0.0117
	(0.0077)	(0.0064)	(0.0486)	(0.0128)	(0.0126)	(0.0174)
lowermid	-0.0142*	-0.0123**	0.0628	-0.0117	-0.0123	-0.0213
	(0.0074)	(0.0061)	(0.0514)	(0.0132)	(0.0135)	(0.0197)
uppermid	0.0026	0.0044	0.1639***	0.0426***	0.0482***	0.0089
	(0.0084)	(0.0068)	(0.0574)	(0.0148)	(0.0137)	(0.0289)
techvoc	0.0570***	0.0427***	0.2297***	0.1102***	0.1014***	0.0500
	(0.0086)	(0.0068)	(0.0698)	(0.0164)	(0.0132)	(0.0664)
university	0.0910***	0.0660***	0.3772***	0.1094***	0.1013***	-0.1262
	(0.0078)	(0.0062)	(0.0564)	(0.0196)	(0.0142)	(15.3665)
Pseudo-R^2	0.1616			0.1843		
Likelihood ratio test (*p*-value)		650.96 (0.00)			281.83 (0.00)	
Observations	31,801	31,801	5,875	8,922	8,922	1,542

Note: Only estimation results for marginal effects are reported. For dummy variables, the reported coefficient is the effect of changing from dummy from 0 to 1. Under the RE probit and the FE logit analysis, standard errors are in parentheses. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: gender, marital status, urban/rural, region of residence, presence of children in the household, and year.

Table 4.7a Transition Probabilities of Employment

Transition Probabilities[1]	Presently not Working (%) (*employed* = 0)	Presently Working (%) (*employed* = 1)	Total (%)
Presently not Working (%) (*employed* = 0)	51	49	100
Presently Working (%) (*employed* = 1)	9	91	100
Total (%)	13	87	100

Source: Author's calculations using CHNS data.

[1] The author uses a Stata *xttrans* to calculate the wave-to-wave transitions in whether individuals were presently working. It should be noted that in a case of unbalanced panel data, *xttrans* does not count transitions from non missing to missing or from missing to non missing values.

Table 4.7b Transition Probabilities of Self-reported Health Status

Transition Probabilities[1]	*excellent* (%)	*good* (%)	*fair* (%)	*poor* (%)	Total (%)
***excellent* (%)**	26	57	15	2	100
***good* (%)**	14	62	22	2	100
***fair* (%)**	10	52	32	6	100
***poor* (%)**	6	38	33	23	100
Total (%)	15	59	23	3	100

Source: Author's calculations using CHNS data.

[1] The author uses a Stata *xttrans* to calculate the wave-to-wave transitions in whether individuals had excellent, good, fair or poor health status. It should be noted that in a case of unbalanced panel data, *xttrans* does not count transitions from non missing to missing or from missing to non missing values.

Table 4.8 The Impact of Having Poor Health Status on Employment[1]: Bivariate Probit

	Bivariate Probit						
Variables	**Total**	**Men**	**Women**	**Urban**	**Rural**	**Young (aged 16-45)**	**Older (aged 46-59)**
The Impact of Having Poor Health Status on Employment:							
poor	-0.0292	-0.0795*	0.2192***	0.0820	-0.0397	-0.0430	-0.0949
	(0.0479)	(0.0410)	(0.0671)	(0.2984)	(0.0283)	(0.0648)	(0.0690)
The Impact of Instruments on Having Poor Health Status:							
famine	0.0063**	-0.0002	0.0116***	0.0038	0.0080**	0.0096***	0.0046
	(0.0032)	(0.0051)	(0.0039)	(0.0060)	(0.0038)	(0.0028)	(0.0145)
ivsrhs_t4	-0.0628***	-0.0658***	-0.0599***	-0.0509***	-0.0680***	-0.0560***	-0.0911***
	(0.0060)	(0.0082)	(0.0085)	(0.0110)	(0.0072)	(0.0058)	(0.0176)
Wald test $\rho = 0$ test (p-value)	1.36 (0.24)	0.28 (0.60)	8.58 (0.00)	0.37(0.54)	1.26 (0.26)	0.24 (0.62)	0.03 (0.85)
Observations	40,723	22,411	18,312	11,407	29,316	31,801	8,922

Note: Only estimation results for marginal effects are reported. For dummy variables, the reported coefficient is the effect of changing from dummy from 0 to 1. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

[1] The following explanatory variables are also included in both the employment models and the health models but are not reported: SRHS, gender (not in both models for men and for women), age group (not in both models for the young and for the older sample), educational attainment, urban/rural (not in both models for the urban and for the rural sample), region of residence, marital status, and year. Explanatory variables indicating the presence of children in the household are only included in the health models.

Table 4.9 The Impact of BMI on Employment[1]: Pooled Probit, RE Probit and FE Logit - The Total Sample

	The Total Sample		
Variables	**Pooled Probit**	**RE Probit**	**FE Logit**
Estimates Based on WHO Classifications:			
uweight_who	-0.0160**	-0.0139**	-0.0510*
	(0.0074)	(0.0066)	(0.0284)
oweight_who	-0.0238***	-0.0178***	0.0070
	(0.0050)	(0.0044)	(0.0190)
obese_who	-0.0268**	-0.0251**	-0.0508
	(0.0136)	(0.0125)	(0.0543)
excellent	0.0937***	0.0768***	0.0640**
	(0.0130)	(0.0117)	(0.0312)
good	0.0961***	0.0798***	0.0766***
	(0.0125)	(0.0112)	(0.0288)
fair	0.0920***	0.0791***	0.0906***
	(0.0126)	(0.0113)	(0.0290)
Pseudo-R^2	0.1750		
Likelihood ratio test (*p*-value)		917.48 (0.00)	
Observations	35,791	35,791	8,037
Estimates Based on C-WGOC Classifications:			
uweight_cwgoc	-0.0179**	-0.0153**	-0.0504*
	(0.0073)	(0.0065)	(0.0285)
oweight_cwgoc	-0.0252***	-0.0210***	-0.0271
	(0.0046)	(0.0040)	(0.0177)
obese_cwgoc	-0.0281***	-0.0257***	-0.0408
	(0.0087)	(0.0079)	(0.0367)
excellent	0.0951***	0.0779***	0.0634**
	(0.0131)	(0.0117)	(0.0312)
good	0.0974***	0.0809***	0.0768***
	(0.0126)	(0.0113)	(0.0289)
fair	0.0929***	0.0799***	0.0906***
	(0.0126)	(0.0114)	(0.0291)
Pseudo-R^2	0.1754		
Likelihood ratio test (*p*-value)		918.34 (0.00)	
Observations	35,791	35,791	8,037

Note: Only estimation results for marginal effects are reported. Under the pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. Under the RE probit and the FE logit analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, and year.

Table 4.10 The Impact of BMI on Employment[1]: Pooled Probit, RE Probit and FE Logit – Men vs. Women

	Men			Women		
Variables	**Pooled Probit**	**RE Probit**	**FE Logit**	**Pooled Probit**	**RE Probit**	**FE Logit**
Estimates Based on WHO Classifications:						
uweight_who	-0.0307***	-0.0265***	-0.1063***	0.0008	0.0002	0.0088
	(0.0096)	(0.0086)	(0.0373)	(0.0109)	(0.0104)	(0.0419)
oweight_who	-0.0118**	-0.0081	0.0190	-0.0351***	-0.0292***	-0.0052
	(0.0059)	(0.0049)	(0.0256)	(0.0083)	(0.0078)	(0.0264)
obese_who	-0.0073	-0.0099	-0.0840	-0.0440**	-0.0428**	-0.0454
	(0.0161)	(0.0139)	(0.0839)	(0.0221)	(0.0216)	(0.0669)
excellent	0.1257***	0.1068***	0.1103***	0.0474**	0.0354**	0.0086
	(0.0170)	(0.0157)	(0.0407)	(0.0193)	(0.0173)	(0.0451)
good	0.1274***	0.1086***	0.1213***	0.0520***	0.0404**	0.0211
	(0.0165)	(0.0152)	(0.0372)	(0.0183)	(0.0164)	(0.0426)
fair	0.1184***	0.1040***	0.1343***	0.0525***	0.0440***	0.0357
	(0.0165)	(0.0153)	(0.0375)	(0.0184)	(0.0166)	(0.0432)
Pseudo-R^2	0.2017			0.1591		
Likelihood ratio test (*p*-value)		377.79 (0.00)			479.78 (0.00)	
Observations	19,248	19,248	3,763	16,543	16,543	4,274
Estimates Based on C-WGOC Classifications:						
uweight_cwgoc	-0.0316***	-0.0272***	-0.1051***	-0.0018	-0.0019	0.0091
	(0.0096)	(0.0085)	(0.0376)	(0.0109)	(0.0103)	(0.0421)
oweight_cwgoc	-0.0136**	-0.0112**	-0.0238	-0.0353***	-0.0311***	-0.0274
	(0.0055)	(0.0046)	(0.0241)	(0.0075)	(0.0071)	(0.0242)
obese_cwgoc	-0.0095	-0.0110	-0.0552	-0.0461***	-0.0432***	-0.0333
	(0.0101)	(0.0086)	(0.0496)	(0.0145)	(0.0142)	(0.0503)
excellent	0.1266***	0.1075***	0.1094***	0.0488**	0.0366**	0.0078
	(0.0171)	(0.0157)	(0.0409)	(0.0193)	(0.0174)	(0.0449)
good	0.1282***	0.1093***	0.1231***	0.0534***	0.0417**	0.0203
	(0.0165)	(0.0153)	(0.0373)	(0.0183)	(0.0165)	(0.0425)
fair	0.1191***	0.1045***	0.1355***	0.0532***	0.0447***	0.0345
	(0.0166)	(0.0153)	(0.0376)	(0.0185)	(0.0167)	(0.0431)
Pseudo-R^2	0.2020			0.1598		
Likelihood ratio test (*p*-value)		378.78 (0.00)			478.71 (0.00)	
Observations	19,248	19,248	3,763	16,543	16,543	4,274

Note: Only estimation results for marginal effects are reported. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. Under the RE probit and the FE logit analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, and year.

Table 4.11 The Impact of BMI on Employment[1]: Pooled Probit, RE Probit and FE Logit - Urban vs. Rural

		Urban			Rural	
Variables	Pooled Probit	RE Probit	FE Logit	Pooled Probit	RE Probit	FE Logit
Estimates Based on WHO Classifications:						
uweight_who	-0.0240	-0.0172	0.0022	-0.0110	-0.0115*	-0.0804**
	(0.0160)	(0.0155)	(0.0474)	(0.0081)	(0.0069)	(0.0346)
oweight_who	-0.0129	-0.0106	0.0065	-0.0232***	-0.0164***	0.0046
	(0.0098)	(0.0097)	(0.0326)	(0.0057)	(0.0048)	(0.0229)
obese_who	-0.0328	-0.0330	-0.0618	-0.0180	-0.0171	-0.0510
	(0.0294)	(0.0289)	(0.0918)	(0.0144)	(0.0126)	(0.0651)
excellent	0.1247***	0.1088***	0.0541	0.0808***	0.0660***	0.0662*
	(0.0277)	(0.0274)	(0.0543)	(0.0142)	(0.0122)	(0.0371)
good	0.1444***	0.1258***	0.0703	0.0759***	0.0634***	0.0750**
	(0.0265)	(0.0262)	(0.0512)	(0.0137)	(0.0117)	(0.0342)
fair	0.1294***	0.1177***	0.0846	0.0769***	0.0658***	0.0919***
	(0.0267)	(0.0264)	(0.0517)	(0.0137)	(0.0118)	(0.0345)
Pseudo-R^2	0.1760			0.1811		
Likelihood ratio test (*p*-value)		275.24 (0.00)			545.33 (0.00)	
Observations	10,114	10,114	2,782	25,677	25,677	5,255
Estimates Based on C-WGOC Classifications:						
uweight_cwgoc	-0.0241	-0.0172	0.0018	-0.0134*	-0.0132*	-0.0800**
	(0.0161)	(0.0155)	(0.0472)	(0.0081)	(0.0068)	(0.0348)
oweight_cwgoc	-0.0107	-0.0083	0.0089	-0.0272***	-0.0223***	-0.0477**
	(0.0092)	(0.0091)	(0.0291)	(0.0052)	(0.0044)	(0.0220)
obese_cwgoc	-0.0117	-0.0108	0.0024	-0.0268***	-0.0243***	-0.0711
	(0.0166)	(0.0165)	(0.0575)	(0.0100)	(0.0087)	(0.0477)
excellent	0.1251***	0.1090***	0.0531	0.0825***	0.0675***	0.0655*
	(0.0277)	(0.0274)	(0.0543)	(0.0143)	(0.0123)	(0.0373)
good	0.1449***	0.1261***	0.0693	0.0775***	0.0648***	0.0752**
	(0.0265)	(0.0262)	(0.0511)	(0.0137)	(0.0118)	(0.0344)
fair	0.1297***	0.1178***	0.0838	0.0780***	0.0669***	0.0915***
	(0.0267)	(0.0264)	(0.0517)	(0.0138)	(0.0119)	(0.0347)
Pseudo-R^2	0.1759			0.1822		
Likelihood ratio test (*p*-value)		275.52 (0.00)			546.23 (0.00)	
Observations	10,114	10,114	2,782	25,677	25,677	5,255

Note: Only estimation results for marginal effects are reported. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. Under the RE probit and the FE logit analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, and year.

Table 4.12 The Impact of BMI on Employment[1]: Pooled Probit, RE Probit and FE Logit - Young vs. Older

	Young (aged 16-45)			Older (aged 46-59)		
Variables	Pooled Probit	RE Probit	FE Logit	Pooled Probit	RE Probit	FE Logit
Estimates Based on WHO Classifications:						
uweight_who	-0.0217***	-0.0186***	-0.0639*	-0.0139	-0.0177	-0.0088
	(0.0080)	(0.0069)	(0.0350)	(0.0205)	(0.0202)	(0.0276)
oweight_who	-0.0086	-0.0053	0.0133	-0.0420***	-0.0374***	-0.0089
	(0.0056)	(0.0047)	(0.0265)	(0.0100)	(0.0099)	(0.0161)
obese_who	-0.0168	-0.0147	-0.0107	-0.0435	-0.0459	-0.0645**
	(0.0145)	(0.0129)	(0.0717)	(0.0298)	(0.0304)	(0.0303)
excellent	0.0745***	0.0571***	0.0417	0.1089***	0.1109***	0.0133
	(0.0160)	(0.0136)	(0.0466)	(0.0230)	(0.0237)	(0.0148)
good	0.0769***	0.0584***	0.0355	0.1282***	0.1319***	0.0325
	(0.0156)	(0.0132)	(0.0442)	(0.0207)	(0.0216)	(0.0248)
fair	0.0832***	0.0658***	0.0603	0.1051***	0.1151***	0.0331
	(0.0157)	(0.0133)	(0.0447)	(0.0209)	(0.0218)	(0.0269)
Pseudo-R^2	0.1659			0.1899		
Likelihood ratio test (*p*-value)		572.65 (0.00)			255.22 (0.00)	
Observations	27,584	27,584	4,941	8,207	8,207	1,315
Estimates Based on C-WGOC Classifications:						
uweight_cwgoc	-0.0230***	-0.0193***	-0.0633*	-0.0179	-0.0220	-0.0080
	(0.0080)	(0.0069)	(0.0351)	(0.0205)	(0.0200)	(0.0269)
oweight_cwgoc	-0.0127**	-0.0088**	-0.0105	-0.0420***	-0.0435***	-0.0227
	(0.0050)	(0.0042)	(0.0240)	(0.0096)	(0.0093)	(0.0194)
obese_cwgoc	-0.0152	-0.0128	-0.0212	-0.0394**	-0.0414**	-0.0398*
	(0.0097)	(0.0085)	(0.0531)	(0.0172)	(0.0169)	(0.0206)
excellent	0.0752***	0.0576***	0.0423	0.1105***	0.1121***	0.0102
	(0.0161)	(0.0136)	(0.0466)	(0.0230)	(0.0237)	(0.0142)
good	0.0775***	0.0589***	0.0365	0.1302***	0.1335***	0.0298**
	(0.0156)	(0.0132)	(0.0442)	(0.0207)	(0.0216)	(0.0152)
fair	0.0838***	0.0663***	0.0611	0.1057***	0.1155***	0.0306**
	(0.0157)	(0.0134)	(0.0448)	(0.0209)	(0.0218)	(0.0153)
Pseudo-R^2	0.1661			0.1902		
Likelihood ratio test (*p*-value)		570.79 (0.00)			260.71 (0.00)	
Observations	27,584	27,584	4,941	8,207	8,207	1,315

Note: Only estimation results for marginal effects are reported. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. Under the RE probit and the FE logit analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: gender, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, and year.

Table 4.13 Marginal Effects using Different Cutoff Points of BMI for Overweight[1]

Variables	Total	Men	Women	RE Probit Urban	Rural	Young (aged 16-45)	Older (aged 46-59)
Estimates Based on WHO Classifications:							
oweight_who2	-0.0179***	-0.0072	-0.0304***	-0.0114	-0.0160***	-0.0052	-0.0373***
	(0.0043)	(0.0049)	(0.0076)	(0.0096)	(0.0047)	(0.0046)	(0.0097)
excellent	0.0777***	0.1092***	0.0353**	0.1104***	0.0666***	0.0582***	0.1115***
	(0.0117)	(0.0158)	(0.0173)	(0.0274)	(0.0122)	(0.0136)	(0.0237)
good	0.0806***	0.1109***	0.0403**	0.1274***	0.0640***	0.0595***	0.1325***
	(0.0113)	(0.0154)	(0.0164)	(0.0263)	(0.0118)	(0.0133)	(0.0216)
fair	0.0797***	0.1058***	0.0438***	0.1190***	0.0662***	0.0666***	0.1155***
	(0.0114)	(0.0154)	(0.0166)	(0.0265)	(0.0119)	(0.0134)	(0.0218)
Likelihood ratio test (*p*-value)	918.02 (0.00)	379.31 (0.00)	479.58 (0.00)	276.67 (0.00)	544.71 (0.00)	573.68 (0.00)	254.85 (0.00)
Observations	35,791	19,248	16,543	10,114	25,677	27,584	8,207
Estimates Based on C-WGOC Classifications:							
oweight_cwgoc2	-0.0209***	-0.0099**	-0.0328***	-0.0077	-0.0220***	-0.0083**	-0.0421***
	(0.0039)	(0.0044)	(0.0068)	(0.0087)	(0.0042)	(0.0040)	(0.0088)
excellent	0.0789***	0.1099***	0.0367**	0.1105***	0.0681***	0.0587***	0.1129***
	(0.0118)	(0.0159)	(0.0174)	(0.0274)	(0.0123)	(0.0137)	(0.0238)
good	0.0818***	0.1116***	0.0418**	0.1276***	0.0654***	0.0600***	0.1342***
	(0.0113)	(0.0154)	(0.0165)	(0.0263)	(0.0119)	(0.0133)	(0.0217)
fair	0.0806***	0.1064***	0.0446***	0.1191***	0.0673***	0.0670***	0.1162***
	(0.0114)	(0.0155)	(0.0167)	(0.0265)	(0.0120)	(0.0135)	(0.0219)
Likelihood ratio test (*p*-value)	919.29 (0.00)	380.46 (0.00)	478.62 (0.00)	276.99 (0.00)	545.44 (0.00)	572.31 (0.00)	260.26 (0.00)
Observations	35,791	19,248	16,543	10,114	25,677	27,584	8,207
Estimates Based on Bao *et al.* (2008) Classifications:							
oweight_bao	-0.0192***	-0.0120**	-0.0276***	-0.0083	-0.0193***	-0.0055	-0.0428***
	(0.0041)	(0.0047)	(0.0073)	(0.0091)	(0.0045)	(0.0043)	(0.0093)
excellent	0.0783***	0.1104***	0.0351**	0.1104***	0.0673***	0.0582***	0.1132***
	(0.0117)	(0.0159)	(0.0173)	(0.0274)	(0.0123)	(0.0136)	(0.0237)
good	0.0812***	0.1120***	0.0404**	0.1275***	0.0647***	0.0596***	0.1341***
	(0.0113)	(0.0154)	(0.0164)	(0.0263)	(0.0118)	(0.0133)	(0.0216)

fair	0.0802***	0.1068***	0.0437***	0.1191***	0.0668***	0.0667***	0.1169***
	(0.0114)	(0.0155)	(0.0166)	(0.0265)	(0.0119)	(0.0134)	(0.0218)
Likelihood ratio test (*p*-value)	917.08 (0.00)	379.01 (0.00)	478.81 (0.00)	276.57 (0.00)	544.21 (0.00)	573.08 (0.00)	256.35 (0.00)
Observations	35,791	19,248	16,543	10,114	25,677	27,584	8,207
Estimates Based on Tuan *et al.* (2008) Classifications:							
oweight_tuan	-0.0164***	-0.0068*	-0.0298***	0.0016	-0.0184***	-0.0068*	-0.0296***
	(0.0035)	(0.0039)	(0.0064)	(0.0083)	(0.0037)	(0.0037)	(0.0081)
excellent	0.0790***	0.1101***	0.0367**	0.1095***	0.0687***	0.0588***	0.1129***
	(0.0118)	(0.0159)	(0.0174)	(0.0274)	(0.0123)	(0.0137)	(0.0238)
good	0.0817***	0.1115***	0.0415**	0.1266***	0.0657***	0.0601***	0.1338***
	(0.0113)	(0.0154)	(0.0164)	(0.0262)	(0.0119)	(0.0133)	(0.0217)
fair	0.0804***	0.1063***	0.0442***	0.1183***	0.0674***	0.0671***	0.1155***
	(0.0114)	(0.0155)	(0.0167)	(0.0265)	(0.0120)	(0.0135)	(0.0219)
Likelihood ratio test (*p*-value)	918.81 (0.00)	379.09 (0.00)	479.43 (0.00)	277.55 (0.00)	544.21 (0.00)	572.37 (0.00)	259.12 (0.00)
Observations	35,791	19,248	16,543	10,114	25,677	27,584	8,207

Note: Only estimation results for marginal effects after the RE probit model are reported. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: gender (not in the model for men and for women), age group (not in the model for the young and for the older sample), educational attainment, urban/rural (not in the model for the urban and for the rural sample), region of residence, marital status, presence of children in the household, and year.

Table 4.14a Transition Probabilities of BMI Categories (using WHO Classification)

Transition Probabilities[1]	*uweight_who* (%)	*nweight_who* (%)	*oweight_who* (%)	*obese_who* (%)	**Total (%)**
***uweight_who* (%)**	44.1	54.7	0.2	0.0	100
***nweight_who* (%)**	3.0	87.3	9.6	0.1	100
***oweight_who* (%)**	0.1	22.2	71.8	5.9	100
***obese_who* (%)**	0.0	2.3	26.8	70.9	100
Total (%)	5.2	74.2	18.7	1.9	100

Source: Author's calculations using CHNS data.

[1] The author uses a Stata *xttrans* to calculate the wave-to-wave transitions in whether individuals are underweight, normal weight, overweight or obese. It should be noted that in a case of unbalanced panel data, *xttrans* does not count transitions from nonmissing to missing or from missing to nonmissing values.

Table 4.14b Transition Probabilities of BMI Categories (using C-WGOC Classification)

Transition Probabilities[1]	*uweight_cwgoc* (%)	*nweight_cwgoc* (%)	*oweight_cwgoc* (%)	*obese_cwgoc* (%)	**Total (%)**
***uweight_cwgoc* (%)**	45.1	54.0	0.9	0.0	100
***nweight_cwgoc* (%)**	3.3	82.8	13.4	0.5	100
***oweight_cwgoc* (%)**	0.1	20.9	68.1	10.9	100
***obese_cwgoc* (%)**	0.0	2.5	25.8	71.7	100
Total (%)	5.2	65.3	24.2	5.3	100

Source: Author's calculations using CHNS data.

[1] See Note 1 in Table 5.14a. The author uses a Stata *xttrans* to calculate the wave-to-wave transitions in whether individuals are underweight, having healthy weight, overweight or obese. It should be noted that in a case of unbalanced panel data, *xttrans* does not count transitions from nonmissing to missing or from missing to nonmissing values.

Table 4.15 The Impact of BMI on Employment[1] - without vs. with Controlling for SRHS

				RE Probit			
Variables	**Total**	**Men**	**Women**	**Urban**	**Rural**	**Young (aged 16-45)**	**Older (aged 46-59)**
The Impact of BMI on Employment – without Controlling for SRHS							
uweight_who	-0.0156**	-0.0298***	-0.0003	-0.0211	-0.0124*	-0.0191***	-0.0217
	(0.0066)	(0.0088)	(0.0104)	(0.0157)	(0.0070)	(0.0069)	(0.0209)
oweight_who	-0.0171***	-0.0068	-0.0291***	-0.0100	-0.0157***	-0.0050	-0.0360***
	(0.0044)	(0.0049)	(0.0078)	(0.0098)	(0.0048)	(0.0046)	(0.0100)
obese_who	-0.0237*	-0.0086	-0.0416*	-0.0321	-0.0155	-0.0138	-0.0425
	(0.0124)	(0.0138)	(0.0215)	(0.0290)	(0.0125)	(0.0128)	(0.0304)
Likelihood ratio test (*p*-value)	939.66 (0.00)	397.71 (0.00)	482.96 (0.00)	290.04 (0.00)	553.72 (0.00)	582.70 (0.00)	257.26 (0.00)
Observations	35,791	19,248	16,543	10,114	25,677	27,584	8,207
The Impact of BMI on Employment – Controlling for SRHS							
uweight_who	-0.0139**	-0.0265***	0.0002	-0.0172	-0.0115*	-0.0186***	-0.0177
	(0.0066)	(0.0086)	(0.0104)	(0.0155)	(0.0069)	(0.0069)	(0.0202)
oweight_who	-0.0178***	-0.0081	-0.0292***	-0.0106	-0.0164***	-0.0053	-0.0374***
	(0.0044)	(0.0049)	(0.0078)	(0.0097)	(0.0048)	(0.0047)	(0.0099)
obese_who	-0.0251**	-0.0099	-0.0428**	-0.0330	-0.0171	-0.0147	-0.0459
	(0.0125)	(0.0139)	(0.0216)	(0.0289)	(0.0126)	(0.0129)	(0.0304)
Likelihood ratio test (*p*-value)	917.48 (0.00)	377.79 (0.00)	479.78 (0.00)	275.24 (0.00)	545.33 (0.00)	572.65 (0.00)	255.22 (0.00)
Observations	35,791	19,248	16,543	10,114	25,677	27,584	8,207

Note: Based on WHO Classifications, only estimation results for marginal effects after the RE probit model are reported. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: gender (not in the model for men and for women), age group (not in the model for the young and for the older sample), educational attainment, urban/rural (not in the model for the urban and for the rural sample), region of residence, marital status, presence of children in the household, and year.

Table 4.16 The Impact of Obesity on Employment[1]: Bivariate Probit

				Bivariate Probit			
Variables	**Total**	**Men**	**Women**	**Urban**	**Rural**	**Young (aged 16-45)**	**Older (aged 46-59)**
The Impact of Obesity on Employment:							
oweight_who2	-0.1216***	-0.0761***	-0.1420***	-0.2203***	-0.0878***	-0.0820***	0.0069
	(0.0201)	(0.0253)	(0.0294)	(0.0341)	(0.0206)	(0.0291)	(0.1044)
The Impact of Instruments on Obesity:							
famine	0.0309***	0.0164	0.0405***	0.0198	0.0336***	0.0446***	0.0298
	(0.0072)	(0.0104)	(0.0101)	(0.0149)	(0.0081)	(0.0074)	(0.0275)
hprotein	0.0177***	0.0169**	0.0225**	-0.0163	0.0366***	0.0185***	0.0207
	(0.0065)	(0.0085)	(0.0098)	(0.0112)	(0.0079)	(0.0071)	(0.0159)
Wald test $\rho = 0$ test (p-value)	24.79 (0.00)	7.766 (0.01)	13.07 (0.00)	28.36(0.00)	11.18 (0.00)	6.79 (0.62)	0.20 (0.65)
Observations	34,784	18,661	16,123	9,861	24,923	26,743	8,041

Note: Only estimation results for marginal effects are reported. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

[1] The following explanatory variables are also included in both the employment models and the health models but not reported: SRHS, gender (not in both models for men and for women), age group (not in both models for the young and for the older sample), educational attainment, urban/rural (not in both models for the urban and for the rural sample), region of residence, marital status, and year. Explanatory variables indicating the presence of children in the household are only included in the health models.

Appendix 4.1 The Impact of Instruments on Employment[1]

				Pooled Probit			
Variables	**Total**	**Men**	**Women**	**Urban**	**Rural**	**Young (aged 16-45)**	**Older (aged 46-59)**
The Impact of Instruments on Employment - Having Poor Health Status as Heath Indicator:							
famine	0.0057	-0.0282	0.0183	0.0496	-0.0254	-0.0027	0.2064**
	(0.0477)	(0.0681)	(0.0640)	(0.0779)	(0.0584)	(0.0518)	(0.0865)
ivsrhs_t4	-0.0347	0.1627*	-0.2261**	-0.2455**	0.1265	-0.0945	0.0655
	(0.0651)	(0.0954)	(0.0905)	(0.1032)	(0.0834)	(0.0768)	(0.1175)
Pseudo-R^2	0.1712	0.1983	0.1547	0.1704	0.1792	0.1617	0.1841
Observations	40,723	22,411	18,312	11,407	29,316	31,801	8,922
The Impact of Instruments on Employment - Obesity as Heath Indicator:							
famine	0.0002	-0.0398	0.0175	0.0706	-0.0416	-0.0030	0.1712**
	(0.0387)	(0.0643)	(0.0493)	(0.0670)	(0.0481)	(0.0437)	(0.0870)
hprotein	-0.0400	-0.0185	-0.0679	0.0335	-0.0962**	-0.0364	-0.0554
	(0.0309)	(0.0457)	(0.0426)	(0.0464)	(0.0420)	(0.0363)	(0.0595)
Pseudo-R^2	0.1712	0.1983	0.1547	0.1704	0.1792	0.1617	0.1841
Observations	34,784	18,661	16,123	9,861	24,923	26,743	8,041

Note: Only estimation results for coefficient effects after the pooled probit model are reported. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The following explanatory variables are also included in the models but not reported: gender (not in the model for men and for women), age group (not in the model for the young and for the older sample), educational attainment, urban/rural (not in the model for the urban and for the rural sample), region of residence, marital status, presence of children in the household, and year.

Chapter 5 Health and Wages

5.1 Introduction

As documented in Chapter 2, China has undergone a dramatic transition from central planning to a more market oriented economy since 1978. This process has been accompanied by a rapid income growth. In two decades, China has transferred from a world's most populous and extremely poor society to become the fourth biggest economy (World Bank, 2006). In the meantime, individuals' wage income has also increased dramatically. It is argued that health human capital played an important role in this rapid income growth (Liu, *et al.*, 2004). The previous chapter estimates the relationship between health and the probability of employment using the panel data from the CHNS, 1991-2006. When exploring the effect of health on labour market outcomes, another issue, which is therefore worth considering, is the link between health and labour productivity. This has received growing attention in developed countries (Currie and Madrian, 1999). Based on the same survey, this chapter examines the relationship between health and hourly wage income for the Chinese population. It has several key advantages.

The relationship between health and hourly wages is identified in the Grossman model in which better health can increase wages through the increase in market productivity. Healthy workers are able to do things more efficiently than unhealthy workers in a sense that health and the capacity to adequately perform job requirements are closely related. A large body of international literature (for example, Lee, 1982; Haveman *et al.*, 1994; Contoyannis and Rice, 2001; Cawley, 2004; Pelkowski and Berger, 2004; Gambin, 2005; Kline and Tobias, 2008; and Cai, 2009) has shown that health has a substantial effect on wages in western countries such as Australia, the UK and US. However, only a few Chinese studies provide evidence on this (Wei, 2004; Arcand and Labar, 2005; Fan *et al.*, 2008; and Gao and Smyth, 2009). Besides, a comparison of the effect of health on wages between China and other countries is rarely considered. Therefore, this chapter serves to provide more empirical evidence on the relationship between health and wages for the Chinese population. Moreover, compared to Arcand and Labar (2005), this chapter uses data from the most recent waves of the CHNS.

Labour productivity can be defined as the amount of goods and services that a worker produces in a given amount of time. Economic researchers often use the hourly wage rate as an indicator of individuals' productivity. Theoretically, more productive workers produce more output per hour worked, and are thus paid more by employers. Factors that affect individuals' productivity are thereby also likely to affect the wages that employers are prepared to offer them (Forbes *et al.*, 2010).[94] In examining the relationship between health and labour productivity, this chapter uses hourly wage income as an indicator of individuals' productivity. Sensitivity analysis is carried out based on two constructed variables of real hourly wage income, namely real hourly wage rate (*hrwage*), and real hourly wage plus bonuses (*hrinc*),

The main measures of the health used in empirical analysis have been discussed in Chapter 1.[95] This chapter employs the same measures of health in the previous chapter, which are the self-reported health status and BMI. As mentioned in the previous chapter, the prevalence of overweight and obesity in developed countries such the UK and US has been well documented. With the changes in diet habits and living standards improving substantially, the Chinese have seen their weight expanding rapidly over the two decades, which causes a number of problems to people's health and increase the burden on Chinese healthcare system. A journalist claimed that "As early as 2002, statistics showed that China had 200 million overweight people, 60 million of whom were obese. The Chinese Center for Disease Control and Prevention said that from 1992 to 2002, China's overweight rate went up by nearly 40 percent and its obesity rate almost doubled." (China.org.cn, Sept 14, 2012. Available: http://www.china.org.cn/china/2012-09/14/content_26521029.htm). Previous evidence from the Chinese literature conclude that poor health negatively reduces wages for the Chinese workers (Fan *et al.*, 2008), and BMI positively increases wages particularly for rural and for male workers (Zhang, 2003; and Arcand and Labar, 2005).

[94] Forbes *et al.* (2010) argue that using hourly wages as an indicator of productivity requires a number of assumptions such as competitive labour market, with perfect information, mobility of labour, no transaction costs and constant returns to scale, some of these might not fully hold in practice. Therefore, hourly wages are likely to under or overstate labour productivity, which places limitations to studies that use hourly wages to measure labour productivity.

[95] See Section 1.1 in Chapter 1.

This chapter uses the same waves of the CHNS as used in estimating the effect of health on employment, that is, wave 1991-2006. These data have a number of advantages in this context. As previously mentioned in Chapter 4, the use of the panel data facilitates control for individual unobserved factors that might bias the estimates. Further two other important issues should be carefully considered in order to obtain consistent estimates for the effect of health on wage income, which are endogeneity and sample selection bias. The endogeneity of health might arise from either the reverse effect of wages on health, or unobserved factors that affect health and/or wages. Sample selection bias might arise since wage income is only observed for employed workers. Many studies have found that it is important to correct for either one or both of these influences in estimating the effect of health on wages such as Haveman *et al.* (1994), Pelkowski and Berger (2004), Jackle and Himmler (2007), Cai (2009), and Forbes *et al.* (2010). This chapter employs a Two Stage Least Squares (2SLS) method to account for the potential endogeneity of health, and a Heckman (1979) selection correction method allowing for potential sample selection bias.

The analysis is carried out for different groups within the population, which enables to explore the differences across genders and area of residence (urban/rural) in China. In fact, several international studies have estimated the impact of health on wages between genders, although the results are mixed due to the different measures of health and the data from different countries.

The remainder of this chapter is structured as follows. The next section discusses both theoretical and econometric issues in exploring the relationship between health and wages. Section 5.3 reviews the results of previous empirical research. Section 5.4 outlines the empirical models and estimation strategies. A more detailed description of the data and variables used in the empirical analysis is then provided in Section 5.5. Section 5.6 discusses the empirical results. The last section provides the conclusions and suggestions for further research in this area.

5.2 The Link between Health and Wages

5.2.1 Human Capital Theory

The relationship between health and wages has been extensively examined by applied researchers in labour economics (see Section 5.3 for a detailed discussion). The effect of health on wages is identified in the Grossman model in which better health can increase wages through the increase in market productivity. In Grossman (1972b), health, as a component of human capital, is treated as a durable capital stock which depreciates over time and can be improved by investment. The demand for health is positively correlated to the return to health investment and hence to wages. Rationally speaking, healthy workers are able to do things more efficiently than unhealthy workers in a sense that health and the capacity to adequately perform job requirements are closely related. In other words, poor health not only limits the amount of healthy time but also affects the quality of the time available for work (Forbes *et al.*, 2010). The literature also offers two additional explanations on why poor health reduces workers' wages (Contoyannis and Rice, 2001; and Cai, 2009). One explanation is that employers might discriminate against worker with poor health irrespective of their productivity (Johnson and Lambrinos, 1985). The other is that workers with poor health are likely to accept lower wages since they could obtain other work-related benefits, such as health insurance or flexible working time (Mullahy and Sindelar, 1995).

5.2.2 Empirical Issues

According to the literature, there are three major econometric problems associated with accurately estimating the returns to health, namely endogeneity, measurement error, and sample selection bias.

Endogeneity

The endogeneity of health might arise from the reverse effect of wages on health. Some studies refer this endogeneity as '*simultaneity bias*' (Thomas and Strauss, 1997; Contoyannis and Rice, 2001; and Cai, 2009). While it is probable that health affect wages, wages might also affect health. The direction of the reverse effect of wages on health is ambiguous according to theory of health production (Cai, 2009). For example, Grossman (1972b) argues that health capital can be maintained and improved through investments which depend on both economic resources and time available. An increase in the return to health increases the economic input into health production, suggesting a positive effect of wages on health. However, Grossman and Benham (1974) argue that higher wages leads to individuals being involved in heavier market activity and less time input into health production, suggesting a negative effect of wages on health. Some researchers employ simultaneous equation models to control for the reverse effect of wages on health. For example, Lee (1982) proposes a three-stage simultaneous equations model with multiple discrete health indicators. The model is a generalized version of the dummy variables simultaneous system of Heckman (1978). Based on a cross-sectional sample of American male workers, Lee (1982) finds that health and wages are strongly interrelated, that is, wages positively affect health and vice versa. It is also found that treating health as exogenous overestimates its effect on wages. Haveman *et al.* (1994) use a Generalised Method of Moments (GMM) estimator to account for simultaneity. Using panel data on 613 white American males, the results show that there is a larger negative effect of ill-health on wages after accounting for endogeneity. However, the authors also find that wages do not have a statistically significant effect on prior health limitation, suggesting that endogeneity of health does not arise from the reverse effect of wages on health.

As well as the reverse effect of wages on health, the endogeneity of health might also occur due to the correlation of unobserved factors that affect both health and wages, such as time preference, individuals' genetic endowment, work ethic and attitudes towards risk (Contoyannis and Rice, 2001; Pelkowski and Berger, 2004; Forbes *et al.*, 2010). If the unobserved factors affect both health and wages in the same direction, there would be an upward bias in the estimated effect of health when treating health as exogenous. Some studies apply fixed effects panel model to correct for individual-specific effects such as Pelkowski and Berger (2004), and Gambin (2005). Pelkowski and Berger (2004) find that the negative effect of permanent health problem decreases for both American men and women after accounting for individual unobservable factors. Similar results are also found in Gambin (2005) for both genders in most of the 14 European countries. However, one limitation of such approach is that the influence of the time-invariant variables such as race cannot be identified. Contoyannis and Rice (2001) apply the FE estimator and a RE instrumental variable estimator which allows for the effect of time-invariant variables to be estimated, but requiring strong exogeneity assumptions to ensure consistency.[96] The authors suggest that both psychological and self-reported health measures are correlated with the time-invariant individual effect.

Measurement Error

The second type of bias is measurement error, which arises from the use of self-reported health measures. Since individuals' awareness of health conditions might vary with knowledge, employment, income and health insurance status, therefore different individuals with identical health condition may respond differently to the same health question (Bound, 1991). Lee's (1982) empirical strategy also takes into account the compatibility of the self-reported health indicator and the objective health indicator (heath limitation). Although there is a systematic measurement bias in self-reported health, the result shows that self-reported health indicator contains valuable information about the unobservable health capital. Much of the literature employs IV methods to account for both measurement error and endogeneity of self-reported health measures (Lee, 1982; Haveman *et al.*, 1994; and Contoyannis and Rice, 2001)

Sample Selection Bias

Since information on wages is only available for individuals in employment, the sample used to estimate the wage equation is restricted to the working population. This raises the possibility of sample selection bias because the employed people might have different characteristics from people who are unemployed or not in the labour force, that is, "*the unobserved determinants of wages and employment status are correlated*" (Cai, 2009, pp. 294-295). The Heckman's two-stage selection method (Heckman, 1979) is used to account for the potential sample selection bias in examining the impact of health on wages (Pelkowski and Berger, 2004; and Forbes *et al.*, 2010). Pelkowski and Berger (2004) find that the negative effect of permanent health problems on wages increases for men but decreases for women after

[96] For details see model specification in Contoyannis and Rice (2001), pp. 602-605.

accounting for sample selection bias. Forbes *et al.* (2010) conclude that sample selection bias is present for Australian men but not for women.

Overall, most studies usually account for at least one of the above econometric issues, and/or separately account for those issues using different methods. For example, some studies do not account for the reverse effect of wages on health (Pelkowski and Berger, 2004; Arcand and Labar, 2005; and Forbes *et al.*, 2010), while some do not control for sample selection bias (Lee, 1982; and Haveman *et al.*, 1994), while others do not control for either of the two issues (Contoyannis and Rice, 2001; and Gambin, 2005). Arcand and Labar (2005) consider the effect of height and BMI on wages by implementing an estimator that simultaneously corrects for correlated individual effects and selection bias (Wooldridge, 1995), as well as a pairwise differencing estimator that also allows one to account for reverse effect of wages on health (Kyriazidou, 1997). They use panel data from the CHNS 1991 and 1993 and find that once correlated individual effects are controlled for, height and BMI do not have a statistically significant impact on wages. Cai (2009) estimates a simultaneous equation system of wage and health using both two-stage and FIML methods. The author also estimates a model that accounts for the potential sample selection bias. The results suggest that endogeneity of health is found to arise from measurement error and unobserved factors, and controlling for endogeneity of health significantly increases the effect of health on wages. However, Cai (2009) argues that simultaneity bias is not an important issue since the wage variable is found to be statistically insignificant in the health equation. The author does not find evidence of sample selection bias since the error term in both employment and wage equation is not statistically significant.

To the author's knowledge, there is only one exception in this context that is able to control for endogeneity and sample selection in one framework. Jackle and Himmler (2007) utilize a method proposed by Semykina and Wooldridge (2006) and find evidence of sample selection bias for both the female and male German workers. In addition, endogeneity tests reject the null of no endogeneity for men but not for women. However, Forbes *et al.* (2010) argue that the methods to account for sample selection bias in panel models are still in an experimental stage.

5.3 Literature Review

5.3.1 Evidence from the International Literature

The Impact of Self-reported Health, Illness or Disability on Wages

Based on statistical analysis, Davis (1972) uses a sample of 1,583 middle aged (45-54) American males which is drawn from the first wave (1966) of the National Longitudinal Survey (NLS) to examine the relationship between health and earnings. The author defines the respondents' earnings as the total receipts from wages, salaries, commissions and tips in 1965. The health measure is based on a survey question which asks respondents whether or not their health or physical condition affects their work either in terms of work type or in terms of the amount of work. If the answer is affirmative, respondents are defined as disabled. Statistical results show that disabled males have lower earnings and work lower annual working hours comparing to healthy males. Davis (1972) also finds that healthy males report their hourly wage rates approximately 10 percent higher than their disabled counterparts.

Luft (1975) employs Ordinary Least Squares (OLS) to examine the impact of health on labour market outcomes for American adults using data from the Survey of Economic Opportunity (SEO), 1967. In order to explicitly explore the impact of health status on labour market behaviours, nine dependent variables are generated, where three are used to measure labour force participation, four are used to measure wages and two are used to measure the amount of time spent at work. Two of the four wage indicators are total earnings (the sum of wages and income from business and farm self-employment) in the last year and in the last week, respectively. The other two wage indicators are constructed on the basis of reported hours of work: a 'last year basis' weekly earnings and a 'last week basis' hourly wage rate. Based on five self-reported health limitation questions from the SEO, 'well' indicates individuals who report no health limitations at all, and 'sick' defines individuals who report any of the health limitations. The results confirm that labour force participation, wages and working time are all negatively affected by reporting 'sick' health status, and these negative effects are larger for women than for men.

By using the same survey as Davis (1972), Lee (1982) investigates the simultaneous effects of health and hourly wages for 2,876 (aged 45-59) males. Based on a minimum $-\chi^2$ principle, a three-step procedure is proposed for estimating a simultaneous equation system with multiple ordered polychotomous health indicators. One health indicator is based on the individual's self-evaluation of health, and another health indicator is related to health limitations. The results demonstrate that health and wages are strongly jointly dependent, and confirm the findings in the literature (for example, see Grossman and Benham, 1974) that health effect on wages would be understated if health is treated as an exogenous variable. Furthermore, the author also finds that individuals with different race or education levels have different perceptions about their health.

Haveman *et al.* (1994) examine the relationship between wages, working time, and health in a sample of 613 employed white men from the Michigan Panel Study of Income Dynamics (PSID) 1976-1993 through estimation of a 3-equation simultaneous model. This study adopts a Hansen's generalized method of moments method to capture the time dependent nature of these interrelationships. The health indicator in this study, i.e. a categorical self-reported health measure, is drawn from two questions in the PSID. The first question asks an individual whether or not he has a physical or nervous condition that limits the type of work or the amount of work he can do. If the individual responds yes, then he will be asked if the condition influences his work a lot, somewhat, a little, or does not influence at all. Although the findings, which confirm the prediction in Grossman (1972b)[97], show that age is associated positively and significantly with health limitations, working time (annual working hours) does not appear to be associated with health limitations. Haveman *et al.* (1994) conclude that prior health limitations negatively and significantly affect both hourly wages and working hours.

In studying the effect of self-assessed health and psychological health on hourly wages between genders, Contoyannis and Rice (2001) draw on panel data from the first six waves of the BHPS, involving a sample of 1670 individuals (859 males and 811 females). The BHPS was conducted annually and the first wave of the survey began at 1991. Self-assessed health is based on a question which asks individuals to assess their health using five health levels: very good, good, fair, poor, and very poor.

[97] The first prediction of the Grossman Model states that 'if the rate of depreciation increases with age, at least after some point in the life cycle, then the quantity of health capital demanded would decline over the life cycle……', see Grossman (1972b), p. 247.

The authors combine the last two responses into one level, therefore the original five health levels are reduced to four: very good, good, fair, poor/very poor. Another health indicator, psychological health, is derived from a General Health Questionnaire (GHQ) by applying Likert scale method (a score of zero indicates the best psychological status whereas a case of three indicates the worst outcome). In addition, Contoyannis and Rice (2001) exploit various econometric methods which involve OLS, General Least Squares (GLS), single equation fixed effects and random effects instrumental variable estimators to estimate a modified Mincer wage equation where the logarithm of an individual's hourly wages is function of a number of individual time-variant (e.g. age and health) and time-invariant (e.g. ethnicity and qualifications) regressors. They find that decreasing psychological health status leads to a fall in men's wages. However, for females there is a statistically significant and positive impact of excellent self-assessed health on their wages. In addition, the null hypotheses of exogeneity of health cannot be rejected for both genders.

Pelkowski and Berger (2004) distinguish between temporary and permanent illnesses to investigate the impact of health problems on labour market outcomes (including hourly wages and annual working hours) using panel data from the American Health and Retirement Survey (HRS). Estimation approaches include OLS, fixed effects and Heckman selection correction method, where the last two methods are used to control for unobserved heterogeneity and to account for sample selection bias, respectively. Furthermore, the measurement of health in this study includes two dummy variables indicating temporary and permanent health conditions, respectively. It refers to a survey question which asks individual whether the illness is temporary (lasts three months or less) or permanent. They find evidence that permanent health conditions have a statistically significant and negative effect on average hourly wages, whereas temporary health conditions have no statistically significant impact on hourly wages or working hours. Furthermore, male workers are shown to suffer a larger decrease in working hours, while female workers seem to bear a bigger fall in wages.

Gambin (2005) investigates the impact of health on hourly wages by using a sample of employed adults from up to eight waves of the ECHP and a Mincer-type empirical model. Various estimation techniques, which include pooled OLS estimation, the random effects and the fixed effects panel estimators, are applied to unbalanced panels from 14 European countries (1994-2001). This study involves two health indicators: four-levels of self-assessed health status and an indicator of chronic disease. The measurement of the first health indicator in this study is the same as that in Contoyannis and Rice (2001). The second health indicator refers to whether an individual has a chronic physical or mental health problem, illness or disability. Gambin (2005) concludes that individual wages are statistically affected by health in a number of European countries. The author uses i) the ratio of the male to female marginal effects of the health variables[98]; and ii) the difference between male and female marginal effects, to compare the health impact on wages between genders. The results show that the effect of self-assessed health on wages for males is larger than for females in most countries, whereas the effect of chronic disease or disability on wages is larger for females than for males. Overall, the greatest gender differences exist in France, Portugal, Spain and UK.

Based on the third wave of the HILDA survey, Cai (2009) studies the effect of health on hourly wages of Australian male adults by estimating the simultaneous equation model of health and wages. The estimation methods involve the two-stage method and the full information maximum likelihood (FIML) method in order to account for the endogeneity of health. Furthermore, the health indicator of this study is a three-level self-reported health status which is drawn from a health question in the self-completion questionnaire of the HILDA. The question asks individuals to describe their health using five health levels: excellent, very good, good, fair and poor. Because the number of observations in excellent, fair and poor group is insufficient, the author reduces the original five health levels to three health levels which are excellent/very good, good, and fair/poor. The results show that better health has a significant and positive effect on wages. The results also demonstrate that treating health as an exogenous factor underestimates its effect on wages, although the reverse effect of wages on health is insignificant. The author further finds that the endogeneity of health is mainly caused by unobserved factors which have an opposite effect on health and wages.

The Impact of Height, Weight and Nutrition on Wages

[98] Gambin (2005) only applies these ratios when the coefficient estimates for both males and females are statistically significant.

Thomas and Strauss (1997) study the influence of various indicators of health on wages across gender and employment position[99] in urban Brazil. These indicators include height, BMI, and two measures of nutrient intake (*per capita* calorie intake, and *per capita* protein intake).[100] The analysis applies instrumental variables to a sample of 16,169 working men and 17,925 working women aged 25-50, from the Estudo Nacional da Despesa Familiar (ENDEF), 1974-1975. In this paper, BMI and nutrient intake are treated as endogenous variables. Instruments are ten relative food prices, and non-labour income of an individual and other household members. Three major findings are as follows: firstly, height significantly and positively associated with wages for men (the effect in the self-employed sector is bigger than that in the market sector), and only for women in the market sector; secondly, BMI is a statistically significant and positive predictor of wages for males (again the effect on self-employed men is bigger than the effect on male market-workers), but not for females; and thirdly, the nutrient intake, that is, both *per capita* calorie intake and *per capita* protein intake, have a statistically significant and positive effect on wages for both male and female market-workers, but not for male or female self-employed workers. This paper also finds that after controlling for calorie intake, BMI and height, additional protein has a greater return at high level of intakes, reflecting a return to higher quality diets.

In studying the effect of weight on career payoffs (in the forms of personal income and occupational positions), Haskins and Ransford (1999) examine data from 306 female employees working in a large American aerospace organization in 1988. They adopt three measurements of weight namely 'objective weight', 'subjective weight', and 'entry weight'. The primary measure of weight is the 'objective weight' which is based on the respondent's self-reported weight, height, and frame size. This study uses Metropolitan Height-Weight Tables 1983 as the ideal weight standard and defines three weight categories which are overweight, slightly overweight, and ideal weight or underweight[101]. Both 'subjective weight' and 'entry weight' are considered

[99] The employment positions are distinguished between those who work in the market sector and those who are self-employed. The authors list two reasons of examine differences across sectors, see pp 177 in Thomas and Strauss (1997).
[100] This paper points out that utilizing *per capita* calorie intake and *per capita* protein intake of household members as measures of nutrient intakes could cause measurement error due to difficulty in measuring nutrient intakes at the individual level.
[101] Overweight indicates that respondent's is 10% or more over the upper limit of the ideal range in the Metropolitan Table. Slightly overweight indicates that respondent's is 1% and 9.9% over the

as the alternative measures of weight. The 'subjective weight' is drawn from a survey question in which individuals were asked to define their weight as overweight, slightly overweight, just right or underweight. To make sense of causal time order that weight is antecedent and affects occupational status, 'entry weight', is included in the analysis. It is based on a survey question in which individuals were asked to report their weight at the time they entered the organization. The regression results[102], which mainly focus on the effect of primary weight on career payoffs, confirm most of their hypotheses[103]. Firstly, weight negatively and significantly affects income among females, but only in entry-level professional and managerial position[104]. Secondly, weight has a negative and significant effect on occupational positions, especially so in traditionally male-dominated occupations, but not so in frontstage occupations (or jobs involving outside contact). The subjective weight and entry weight findings are very similar to the objective weight findings except that subjective weight has no effect on income. The results also confirm that weight based stratification occurs not only in the general society but also occurs in the inner workings of workplace organizations. Haskins and Ransford (1999) further point out that their study fails to provide comparisons across genders and different organizations.

Cawley (2004) explores how the impact of body weight on hourly wages differs by gender and race for the American population. Various estimation strategies are applied to a balanced panel data from the National Longitudinal Survey of Youth (NLSY) in order to take into account the potential endogeneity of weight, including pooled OLS, fixed-effects and instrumental variables. This study involves three weight indicators: continuous measures of BMI and weight in pounds (controlling for height in inches); and a category variable for clinical BMI classifications: underweight (a BMI below 18.5), overweight (a BMI of 25 to 30), obese (a BMI

upper limit of the ideal range in the Metropolitan Table. The last category indicates respondent's weight is within or below the ideal range in the Metropolitan Table.

[102] Unfortunately, the authors do not give much detail on their empirical strategies.

[103] Haskins and Ransford (1999) propose three hypotheses: a) incomes and occupational positions of females who are at or below their desirable weight will be higher than of those who are overweight, after controlling for other human capital factors; b) weight will be significantly related to the career payoffs for females whose job involves contact outside the organization; c) weight will be significantly related to the career payoffs for women who are in male-dominated occupations.

[104] The authors create four dummy variables for occupational position: upper-level professional (or managerial) position, entry-level professional (or managerial) position, clerical position and blue collar position.

above 30) and healthy weight (a BMI of 18.5 to 25), where the healthy weight is the reference category. The author uses the BMI of a sibling (controlling for age and gender of the sibling) as an instrument which is assumed to be strongly correlated with the respondent's BMI and uncorrelated with the wage residual of the respondent. The results show that weight affects wages negatively for white females. However, the negative relationship between weight and wages for other gender-race groups is a consequence of unobserved heterogeneity. In addition, the null hypothesis of the exogeneity of weight is rejected only for white women.

Brunello and D'Hombres (2007) estimate the effect of BMI on hourly wages across gender and geographical region in nine European countries, using a pooled sample of adults aged 18-65 from the ECHP 1998-2001. The authors use self-reported measures of height and weight to calculate the BMI. The BMI of a biological family member[105] is used as an instrument for the respondent's BMI and they find evidence that the negative impact of BMI on wages is larger for men than for women (a 10% increase in mean BMI decreases the real hourly wages by 3.27% for men and 1.86% for women, respectively). Moreover, comparing the difference between southern (Greece, Italy, Portugal and Spain) and northern Europe (Austria, Belgium, Denmark, Finland and Ireland), the effect is larger for southern Europe than for northern Europe.

Kline and Tobias (2008) examine the effect of BMI on hourly wages in a sample of 4343 native British people (2,561 men and 1,782 women) aged 29-30 from the 1970 British Cohort Study. They let the BMI variable to enter the log hourly wage equation nonparametrically and propose a Bayesian posterior simulator for fitting a standard two-equation model that allows for skew in the BMI distribution. Similar to Cawley (2004) and Brunello and D'Hombres (2007), this study applies the instrumental variables method and uses the BMI of a biological family member (parent BMI) as a source of exogenous variation. Three clinical classifications of the BMI in this study are normal BMI (18.5-25), overweight BMI (25-30) and obese BMI (30+). The presence of nonlinearities in relationship between BMI and wages is confirmed by this study. In addition, Kline and Tobias (2008) find evidence that

[105] A biological family member is defined as a parent, sibling or child. Brunello and D'Hombres (2007) average out all available BMIs in a case that individuals have more than one family member. For example, when an individual has two siblings and three children, the authors take the unweighted average of the average BMI of the siblings and the average BMI of the children.

British men receive smaller wage penalties for increases in BMI than British women receive when they are in the normal BMI range. However, overweight and obese males receive relatively large penalties for increases in BMI compared to the overweight and obese females, respectively.

Gregory and Ruhm (2009) use data from the 1986 and 1999-2005 Panel Study of Income Dynamics to examine the effect of BMI on wages of men and women aged 25-55 in United States. In this study, BMI ranges from 20 to 40. Two strategies are included: a semi-parametric strategy and an instrumental variable strategy. The semi-parametric estimates show that female earnings peak at levels considerably below the clinical classification of overweight or obesity. The findings also illustrate a moderately flat BMI-wage relationship for males where wages are maximized in the early overweight category. The results of the instrumental variables estimates for both genders are consistent with that of the semi-parametric estimates for females.

5.3.2 Evidence from the Chinese Literature

Wei (2004) uses data from the CHNS 1993 to examine the effect of health on non-agricultural employment and wages in rural areas. Based on six questions in the 1993 physical examination questionnaire[106], the author applies the Two-factor Varimax Rotation Method to obtain two health indicators: *Physiological Factor* and *Psychological Factor*. Furthermore, this study adopts the same Heckman selection correction method as utilized in Pelkowski and Berger (2004) to control for the sample selection bias. Wei (2004) concludes that having better health has a statistically significant and positive effect on non-agricultural labour force participation in rural area. However, health does not have a statistically significant effect on wages.

Using the self-reported health indicator, Fan *et al.* (2008) investigate the effect of poor health on wages of 2,519 Chinese adults (aged 20-64). The data is drawn from the Chinese Health and Family Life Survey (CHFLS). The self-reported health indicator is a dummy variable which takes value 1 for poor health status and 0 for healthy individuals. This study exploits a propensity score matching method to address the endogeneity of health and sample selection bias in estimating an average

[106] China Health and Nutrition Survey Questionnaire 1993. Available: http://www.cpc.unc.edu/projects/china/data/questionnaires.

treatment effect model[107] of poor health on wage losses. The authors find evidence that poor health status causes an 11% reduction in wages.

On the basis of household rather than individual level data, Zhang (2003) uses a sample of 460 households from the China Credit and Poverty Survey, to investigate the impact of health on farming productivity. This study exploits various health indicators, including SRHS, ADLs, chronic disease, the number of working days (in a month) loss due to illness, height, BMI, calorie intake and Desirable Dietary Pattern (DDP) score[108]. By employing an IV method on a family production function, Zhang (2003) finds evidence that for the rural low-income family, both BMI and calorie intake have significant and positive effect on farming incomes, while illness negatively affect farming incomes.

By using the same panel dimensions of CHNS as Benjamin *et al.* (2003), Liu *et al.* (2008) examine how household income is affected by the health of individual household members. The health indicator used in this study is the same self-reported general health measure as used in Benjamin *et al.* (2003). The household income is calculated as the sum of all market and non-market earnings across the household.[109] By employing pooled OLS estimation and a dummy variable variant of individual-level fixed-effects approach, this study concludes three major findings. First, better health has a significant positive effect on household income for the whole sample, although this relationship is less pronounced in the individual-level fixed-effects estimation. Second, there is no obvious evidence of gender differences on the health-household income relationship. Third, the statistically significant and positive relationship between health and household income is far more remarkable in rural areas, especially in the individual-level fixed-effects estimation. The authors also point out the shortcomings of this study. Their estimates provide a lower bound for the impact of health on individual income, since household members can jointly engage in various compensatory reallocation of labour supply in response to

[107] Both the model and methodology are proposed by Rosenbaum and Rubin (1983).

[108] In these eight health indicators, six are individual level indicators except for the last two. The author uses the weighted average method to convert those six individual level indicators into household level indicators.

[109] Liu *et al.* (2008) divide the total annual household income into market and non-market groups. The market group includes: wage income, home gardening income, collective and farming income, raising livestock income, fishing income and business income. The non-market group includes: food coupons, housing subsidy, childcare subsidy and other income subsidies.

individual health shocks. In addition, the results, to some extent, are affected by the subjective nature of the self-reported health indicator.

Arcand and Labar (2005) explore the impact of height and BMI on hourly wages for Chinese working men using panel data from the CHNS, 1991 and 1993. They adopt a method which is proposed by Wooldridge (1995) to control for both unobserved individual heterogeneity and sample selection bias. Furthermore, they also employ an IV pairwise differencing estimator proposed by Kyriazidou (1999) to account for endogeneity of health. The results show that neither of the two health measures has a statistically significant effect on wages for the Chinese working men aged below 27, whereas for those aged above 27, both BMI and height have a statistically significant and positive effect on wages.

Gao and Smyth (2009) study the impact of height (measured in centimetres) on hourly wages using a sample of 6,546 male adults and 4,966 female adults from the China Urban Labour Survey, 2005, through the estimation of a Mincer earnings function. Two estimation techniques which involve OLS and two-stage least squares (TSLS), are applied to data on individuals from twelve Chinese cities. Overall, height is positively and significantly correlated with wages for both men and women. Both the OLS estimates and the TSLS estimates imply that there are considerable returns to investment in health capital. Their findings also confirm previous work that the effect of height on wages using OLS method is much larger than that using TSLS method.

5.4 Model and Methodology

In examining the impact of SRHS on wages, this study implements pooled OLS, random effects (RE) and fixed effects (FE) techniques to estimate a modified Mincerian Wage Model. In addition, the analysis adopts a Two Stage Least Squares (2SLS) method to account for the potential endogeneity of health, and a Heckman (1979) selection correction method allowing for the potential sample selection bias discussed in the above section. The specification on these methods draw heavily on Heckman (1979), Contoyannis and Rice (2001), Gambin (2005), Cameron and Trivedi (2009), and Greene (2012).

Pooled OLS, RE and FE Estimators

The modified Mincerian Wage Model which is derived from Gambin (2005) is as follows:

$$w_{it} = \gamma_0 + \gamma_1 h_{it} + \gamma_2 e_{1it} + v_{1it}, \; i = 1, \dots, n, \mathrm{t} = 1, \dots, T \tag{5.1}$$

where, i = individuals,

t = time periods, that is, CHNS 1991, 1993, 1997, 2000, 2004 and 2006, and for each i, $t = 1, \dots, T$ (T<=6),

w_{it} = the logarithm of real hourly wage income,

h_{it} = a discrete measure of individuals' health (SRHS and BMI, respectively),

e_{1it} = a vector of explanatory variables which are assumed to be exogenous,

γ_0 = a constant term,

γ_1 and γ_2 = coefficients, and γ_1 is a measure of the effect of health on wages to be estimated,

v_{1it} = a disturbance term, assume $v_{1it} \sim N\left(0, \sigma_{v_1}^2\right)$.

Like cross-sectional OLS, pooled OLS also relies on a between comparison as all observations on all individuals are pooled across time. Therefore, in the pooled OLS model, it is assumed that the disturbance term v_{1it} is uncorrelated with the explanatory variables h_{it} and e_{1it}. Similar to the previous chapter, cluster-robust standard errors are used to correct for error correlation over time for a given individual.

Given the panel nature of the data, equation (5.1) can be written as follows:

$$w_{it} = \gamma_0 + \gamma_1 h_{it} + \gamma_2 e_{1it} + \alpha_i + \eta_{it}, \tag{5.2}$$

where, α_i = random individual-specific effect, i.e. a time-invariant disturbance term,

η_{it} = a time-varying (or idiosyncratic) disturbance term.

In the RE model, it is assumed that both the individual effects α_i and the time-varying error η_{it} are uncorrelated with the explanatory variables h_{it} and e_{it}. The strict exogeneity assumption can be stated as $\mathrm{E}(\eta_{it}|h_i, e_{1i}, \alpha_i) = 0$, and the orthogonality assumption (the orthogonality between the individual effects and the explanatory variables) can be stated as $\mathrm{E}(\alpha_i|h_i, e_{1i}) = \mathrm{E}(\alpha_i) = 0$. Estimation is then performed by a feasible generalized least-squares (GLS) estimator. Efficiency of RE requires that conditional variances are constant and the conditional covariances are zero, that is, $\mathrm{E}(\eta_{it}^2|h_i, e_{1i}) = \sigma_\eta^2$ and $\mathrm{E}(\eta_{it}\alpha_i|h_i, e_{1i}) = 0$. In addition, homoskedasticity assumption on the unobserved effect α_i, i.e.$E(\alpha_i^2|h_i, e_{1i}) = \sigma_\alpha^2$, is required. The RE model yields estimates of all coefficients even those of the time-invariant explanatory variables and it is more efficient than the pooled OLS model. However, a major limitation of the RE estimation is that the random disturbance term associated with each cross-sectional unit is assumed to be uncorrelated with any of the other explanatory variables. As argued by Gambin (2005), this assumption is somewhat unrealistic and failure of this assumption will lead to biased estimates.

In the FE model, the time-varying disturbance term η_{it} is assumed to be uncorrelated with the explanatory variables h_{it} and e_{1it}, while the time-invariant disturbance term α_i is allowed to be correlated with h_{it} and e_{1it}. Such assumption allows a limited form of endogeneity. In other words, the FE model helps to control for omitted variable bias only when the omitted variables are time-invariant. It also implies that RE estimates will be inconsistent if the FE model is appropriate. Therefore, it is possible to consistently estimate the coefficients for the time-varying regressors by eliminating the individual effects α_i.

According to the equation (5.2), the average w_{it} of overtime yields:

$$\bar{w}_i = \gamma_0 + \gamma_1\bar{h}_i + \gamma_2\bar{e}_{1i} + \alpha_i + \bar{\eta}_i, \tag{5.3}$$

Then subtracting this from w_{it} in equation (5.2) yields the within model:

$$(w_{it} - \bar{w}_i) = \gamma_1\left(h_{it} - \bar{h}_i\right) + \gamma_2(e_{1it} - \bar{e}_{1i}) + (\eta_{it} - \bar{\eta}_i), \tag{5.4}$$

This is the third method adopted in this study. The FE estimator is the OLS estimator of the within model which yields unbiased and consistent estimates as N and/or $T \to \infty$ even under the presence of the limited form of endogeneity, because the fixed effect α_i has been eliminated. However, the drawback of the within model is that the coefficients of time-invariant explanatory variables cannot be identified. Hence, in this study, the effect of gender and region of residence cannot be estimated by the FE estimator.

This chapter employs a standard Hausman test (Hausman, 1978) to compare the FE and the RE estimators. Under a strong assumption that RE estimator is fully efficient, the null hypothesis is that the individual effects are uncorrelated with the explanatory variables, i.e. the RE model is consistent. Pooled OLS estimates will be consistent if the RE model is appropriate but will be inconsistent if the FE model is appropriate.

The Hausman test compares only the coefficient estimates of the time-varying explanatory variables using the FE and the RE estimators. The test statistic is given as:

$$H = (\hat{\gamma}_{RE} - \hat{\gamma}_{FE})'(cov\, \hat{\gamma}_{FE} - cov\, \hat{\gamma}_{RE})^{-1}(\hat{\gamma}_{RE} - \hat{\gamma}_{FE}), \qquad (5.5)$$

where, $\hat{\gamma}_{RE}$ = a vector of RE estimates without the coefficients of the time-invariant explanatory variable,

$\hat{\gamma}_{FE}$ = a vector of FE estimates.

It has a chi-squared asymptotic distribution with K degrees of freedom, $\chi^2(K)$, where K is the number of coefficients.

The 2SLS Estimator

The fundamental assumption for consistency of pooled OLS is that the disturbance term v_{1it} is uncorrelated with the explanatory variables (both h_{it} and e_{1it}). When h_{it} is endogenous, it will be correlated with the disturbance term v_{1it}, hence leading the pooled OLS estimates to be biased. To account for potential endogeneity of health in a wage determination equation, this study employs the 2SLS estimator by adding a health determination equation. The model is as follows:

$$w_{it} = \gamma_3 + \gamma_4 h_{it}^{2SLS} + \gamma_5 e_{1it} + v_{2it} \qquad (5.6)$$

$$h_{it}^{2SLS} = \gamma_6 + \gamma_7 u_{it} + \gamma_8 e_{1it} + v_{3it} \qquad (5.7)$$

where, for simplicity, the analysis now consider health following Cai (2009). Therefore, h_{it}^{2SLS} = a continuous measure of individuals' health,

u_{it} = a vector instruments that are correlated with h_{it}^{2SLS} but not with v_{2it},

γ_3 and γ_6 = a constant term,

$\gamma_4, \gamma_5, \gamma_7,$ and γ_8 = coefficients, and the coefficients of interest is γ_4,

v_{2it} and v_{3it} = two disturbance terms.

To obtain a consistent estimator, it is assumed that Cov$[u_{it}, h_{it}^{2SLS}] \neq 0$, Cov$[u_{it}, v_{2it}]$ = 0, and Cov$[u_{it}, v_{3it}] = 0$. In other words, the 2SLS method requires at least one instrument that determines health (belongs in the health equation), but that does not influence wages (does not belong in the wage equation). If the number of instruments exceeds the number of endogenous variables, the model is over identified.

The first stage of the 2SLS method is to estimate Equation (5.7) using pooled OLS and obtain consistent estimates for the reduced form coefficients, indicated as $\hat{\gamma}_6$, $\hat{\gamma}_7$, and $\hat{\gamma}_8$. With these estimates, predicted values of health can be obtained as,

$$\hat{h}_{it}^{2SLS} = \hat{\gamma}_6 + \hat{\gamma}_7 u_{it} + \hat{\gamma}_8 e_{1it} \qquad (5.8)$$

A F test is used to test the joint significance of the coefficient of instruments. If the F statistic is significant, the instruments have significant explanatory power for health after controlling for the other explanatory variables.

In the second stage, endogenous health h_{it}^{2SLS} is replaced by predicted health $\hat{h}_{it}^{2SLS}$ to estimate Equation (5.6) using pooled OLS. The formula for the 2SLS estimator is given by $\hat{\gamma}_{2SLS} = \frac{\mathrm{Cov}\left(\hat{h}_{it}^{2SLS}, w_{it}\right)}{\mathrm{Cov}\left(\hat{h}_{it}^{2SLS}, h_{it}\right)}$.

A Durbin-Wu-Hausman (DWH) test is used to test for endogeneity of health by comparing the pooled OLS and the 2SLS estimates. Rewrite the Equation (5.6) with an additional variable, v_{3it},

$$w_{it} = \gamma_3 + \gamma_4 h_{it}^{2SLS} + \gamma_5 e_{1it} + \gamma_9 v_{3it} + v_{2it} \qquad (6.9)$$

Recall that v_{3it} is the disturbance term from Equation (5.7). The null hypothesis of exogeneity of health, $\mathrm{H}_0: \gamma_9 = 0$ indicates $E\left(v_{3it} v_{2it} \middle| h_{it}^{2SLS}, e_{1it}\right) = 0$. Since v_{3it} is not directly observed, the residuals $\hat{v}_{3it}$ is instead substituted from Equation (5.7). Then,

$$w_{it} = \gamma_3 + \gamma_4 h_{it}^{2SLS} + \gamma_5 e_{1it} + \gamma_9 \hat{v}_{3it} + v_{2it} \qquad (6.10)$$

If health is endogenous, the 2SLS estimator is consistent whereas the pooled OLS estimator is inconsistent. If there is no endogeneity, both estimators are consistent but the 2SLS estimator would be less efficient than the pooled OLS estimator.

It is possible to test the validity of instruments in an over identified model by using a Sargan-Hansen test. The joint null hypothesis is that the instruments are valid instruments (i.e. uncorrelated with the error term), and that the excluded instruments are correctly excluded from the wage equation. For the 2SLS estimator, where the cluster-robust standard errors are used to correct for error correlation over time for a given individual, the test statistic is Hansen J statistic. The Hansen J statistic is calculated as N*R-squared from a regression of the 2SLS residual $\hat{v}_{2it}$ on the full set of instruments. Under the null hypothesis that all instruments are valid, the Hansen J statistic has an asymptotic chi-squared distribution with degrees of freedom equal to the number of over identifying restrictions, where the number of over identifying restrictions is the number of instrumental variables minus the number of endogenous variables.

The Heckman Sample Selection Estimator

The Heckman sample selection model combines a selection equation (employment equation) and an outcome equation (wage equation), which are similar to Equation (4.1) in Chapter 4 and Equation (5.1) in this chapter, respectively. Simplifying the two equations gives:

$$y_{it} = \begin{cases} 1 \text{ if } y_{it}^{*} = \beta_0^s h_{it} + \beta_1^s x_{it}^s + \varepsilon_{0it}^s > 0, \\ 0 \text{ otherwise} \end{cases} \tag{5.11}$$

and

$$w_{it} = \gamma_0^s + \gamma_1^s h_{it} + \gamma_2^s e_{1it} + v_{1it}^s \tag{5.12}$$

where x_{it}^s indicates the other explanatory variables affect employment. Besides the variables which has been used in Chapter 4 (x_{it}), x_{it}^s also includes an exclusion restriction. The estimation with exclusion restrictions requires variables in the selection equation that are not in the outcome equation and vice versa. In other words, excluded variables should have impact on the probability of being employed but does not directly affect wages and vice versa. More detail regarding on the exclusion restrictions is discussed in the next section.

The sample selection bias arises because the disturbance term v_{1it}^s in the wage equation is correlated with the disturbance term ε_{0it}^s in the employment equation, which leads to inconsistent estimates. Heckman (1979) states that sample selection bias can be approached as an omitted variable bias. Assume $\varepsilon_{0it}^s \sim N(0,1)$, $v_{1it}^s \sim N(0, \sigma_v^2)$, and $\mathrm{Cov}(\varepsilon_{0it}^s, v_{1it}^s) = \rho_2$, the expected wage condition upon the individual to be employed is given by:

$$\mathrm{E}(w_{it}|y_{it}^{*} > 0) = \mathrm{E}(\gamma_1^s h_{it} + \gamma_2^s e_{1it} + v_{1it}^s | \beta_0^s h_{it} + \beta_1^s x_{it}^s + \varepsilon_{0it}^s > 0) \tag{4.13}$$

$$= \gamma_1^s h_{it} + \gamma_2^s e_{1it} + \mathrm{E}\left(v_{1it}^s \middle| \beta_0^s h_{it} + \beta_1^s x_{it}^s + \varepsilon_{0it}^s > 0\right)$$

$$= \gamma_1^s h_{it} + \gamma_2^s e_{1it} + \mathrm{E}(v_{1it}^s | \varepsilon_{0it}^s > -\beta_0^s h_{it} - \beta_1^s x_{it}^s)$$

If the disturbance terms ε_{0it}^{s} and v_{1it}^{s} are independent, then $E(v_{1it}^{s}) = 0$ and OLS will give consistent estimates. However, when they are correlated, a correction for selection bias needs to be included. Following Heckman (1979) and Greene (2012),

$$E(v_{1it}^{s} | \varepsilon_{0it}^{s} > -\beta_0^s h_{it} - \beta_1^s x_{it}^s) = \rho_2 \sigma_{v_1} \lambda_{it}(-\beta_0^s h_{it} - \beta_1^s x_{it}^s) \tag{5.14}$$

where $\lambda_{it}(-\beta_0^s h_{it} - \beta_1^s x_{it}^s)$ is the inverse Mill's ratio (IMR), $\lambda_{it}(-\beta_0^s h_{it} - \beta_1^s x_{it}^s) = \frac{\phi\left(\frac{-\beta_0^s h_{it} - \beta_1^s x_{it}^s}{\sigma_{\varepsilon_0^s}}\right)}{1-\Phi\left(\frac{-\beta_0^s h_{it} - \beta_1^s x_{it}^s}{\sigma_{\varepsilon_0^s}}\right)} = \frac{\phi\left(\frac{\beta_0^s h_{it} + \beta_1^s x_{it}^s}{\sigma_{\varepsilon_0^s}}\right)}{\Phi\left(\frac{\beta_0^s h_{it} + \beta_1^s x_{it}^s}{\sigma_{\varepsilon_0^s}}\right)}$.

Following Pelkowski and Berger (2004), the selection model is estimated by maximum likelihood (ML). Cluster-robust standard errors are used to correct for error correlation over time for a given individual. A Wald test of the significance of ρ_2 (H_0: $\rho_2 = 0$) compares the pooled OLS and Heckman sample selection method. If $\rho_2 = 0$, the pooled OLS parameter estimates would be appropriate. If $\rho_2 \neq 0$, there is evidence of sample selection bias, and the Heckman sample selection method should be used.

5.5 Data and Variables

5.5.1 Sample Construction

The data used in examining the relationship between health and wages comes from the same waves of the CHNS (1991-2006) that are used in the previous Chapter. The total working-age sample used in Chapter 4 contains 40,723 observations from 15,275 individuals pooled across the six waves, where 88% are currently in work.

According to the CHNS there are seven sources of individual income: non-retirement wages, collective and household gardening, collective and household farming, raising livestock or poultry, collective and household fishing, small handicraft and small commercial household business, and retirement wages.[110] However, it is important to note that only 38% of those in work report non-retirement wages. This analysis retains individuals who have positive wages income, leaving 13,712 observations from 7,298 individuals across the six waves.

[110] For details, see: http://www.cpc.unc.edu/projects/china/data/datasets/Individual%20Income%20Variable%20Construction.pdf. [cited 15th Aug 2012].

As discussed in Chapter 3, the low response rate of the wage income is mainly due to the low response rate from people who engage in agricultural work and/or who are self-employed. In terms of the distribution of the occupational status, half of the employed sample are agricultural workers, whereas only 4% of these agricultural workers report their source of income is from (non-retirement) wages. Instead of receiving non-retirement wages, more than 85% of the Chinese agricultural workers report their individual income is from one or more than one of the other sources, including gardening, farming, raising livestock, fishing and household business. In the case of the self-employed workers (which account for 59% of the employed sample) instead of receiving non-retirement wages (7%), 82% report their income is from one or more than one of the other sources. Moreover, recall from Chapter 4 that there is a close link between the agricultural work and self-employment.[111] Wang and Cai (2008) also argue that the wage structure of the self-employed workers might be different from that hired workers. For people who are currently in work but are not agricultural workers or self-employed, 85% report their pay. Therefore, this chapter further excludes observations who are either agricultural workers or self-employed who have positive wage income (which account for 19% of the employed sample who have positive wage income). The total sample used in the wage analysis involves 11,115 observations from 5,997 individuals across the six waves. Separate analysis is undertaken for 3,410 males and 2,587 females (6,514 and 4,601 observations, respectively), and for 3,188 urban and 2,809 rural residents (6,102 and 5,013 observations, respectively). The sample without excluding the agricultural workers and/or the self-employed is also estimated for comparison.

5.5.2 Dependent Variable

The dependent variable W_{it} denotes the logarithm of the real hourly wage income. As previously stated in Section 3.2.1 in Chapter 3, there is no direct measure of hourly wage income, thus this study defines it as the ratio between the monthly wage income and the monthly working hours. The monthly wage income is calculated by dividing the annual wage income by the monthly working months[112]. The monthly

[111] For the agricultural workers, 98% of them are self-employed. For the non-agricultural workers, however, less than 20% of them are self-employed. Likewise, 83% of the self-employed are agricultural workers, whereas only 2% of people who are not self-employed engage in agricultural work.

[112] The CHNS asks respondents "*Last year, how many months did you work at this occupation?*".

working hours are calculated by multiplying 4.3 weeks per month by weekly working hours.[113] This constructed nominal hourly wage income is then inflated to 2006 Chinese yuan.[114]

According to the CHNS definition, the annual wage income is the sum of the annualized non-retirement wage (monthly wage times months worked[115]), bonuses, and other cash or in-kind income received during the last year.[116] The monthly wage is surveyed under the section on income from wages in the Household and Individual Survey questionnaire for all the six waves. In waves 1991-1997, the respondents who had a job last year were asked "*How much money do you receive, on the average, for a month's work at this occupation, excluding subsidies and bonuses? (yuan)*". For waves 2000-2006, the monthly non-retirement wage is based on a revised question "*On the average, what was your monthly wage/salary last year, excluding subsidies and bonuses? (yuan)*", which does not change the meaning of the original question.

For the first three waves, information on total annual bonuses is surveyed under the section on welfare subsidies and rations in the Household and Individual Survey questionnaire and is based upon two questions. The first question is "*In the past year, did you receive any cash bonuses (including festival and any other bonuses)?*". If the answer is yes, then the respondent will be asked "*What was the total value of all bonuses last year?*". For waves 2000-2006, the bonus information is surveyed under the same section as monthly wage, and it is also based upon two revised questions (again without changing the meaning of the questions) which are "*Did you receive a bonus last year (including monthly bonus, quarterly bonus, year-end bonus, holiday bonus, and other bonus)?*" and "*Last year, what was the total value of all bonuses for the entire year?*", respectively.[117] Furthermore, the CHNS Work Manual (1993) also states that the annual bonuses include regular monthly bonus, holiday bonus,

---

[113] The CHNS asks respondents "*How many days in a week, on the average, did you work at this occupation?*", and "*How many hours in a day, on the average, did you work at this occupation?*", respectively.

[114] The inflation index sets the cost in urban areas in Liaoning province in 2006 equal to 1 and makes all other costs relative to it. It has been constructed and provided by the CHNS.

[115] Note that 99% of the employed sample work twelve months, and 100% the sub sample used in this analysis work twelve months.

[116] More detail regarding the CHNS definition of wage income, see the web link above in footnote 116.

[117] The response rates of these two questions are 99% and 95%, respectively. Calculation is based on the total sample in this chapter.

dividend bonus, bonus for increase in production, and any other bonuses from the work unit in the last year.

Other cash or in-kind income is based on two set of questions, which are surveyed under the section on other sources of income from work activities in the same questionnaire as for wage and bonuses. The first set of questions includes "*Did you have any other cash income last year?*", and "*How much money was it?*". The second set of questions includes "*Did you have any non-cash income (e.g. clothes, foods, etc) last year?*", and "*How much was it if you bought them from market?*".[118] However, this information on other cash/in-kind income from work is only surveyed in the 2006 wave.

According to the summary statistics in Table 5.1a, the average real annual wages of the Chinese workers increase nearly 4.4 times from 1991 to 2006. The total annual bonuses (Mean[a] in Table 5.1a) also have a three-fold increase. Nevertheless, the proportion of bonuses to total wage income decreases from 18% in 1991 to 8% in 2006. Beyond that, the share of people who have bonuses also decreases from 67% to 39% during this period. The average other cash/in-kind income only makes up a very small proportion of total wage income, and is only 3% for the 2006 wave. Nevertheless, it is a significant component for some subgroups. For example, for those 182 workers who have other cash/in-kind income, the proportion of other cash/in-kind income to total wage income is 20%.

As described in Chapter 2, the second wage reform (early 1980s-mid 1990s) linked both wages and bonuses to actual productivity, although it was still partly relying on a central planned system. After this the recent wage reform gave the enterprises full autonomy on determining their own (and market-oriented) reward standards and wage standards. According to Hu *et al.* (1988), the nature of the bonus remains questionable on whether it is truly reflect the labour productivity or merely a wage supplement. Accordingly, this chapter further pursues a sensitivity analysis of the health effect on wage income by using two measures of the real hourly wage income. The first measure is based on the non-retirement wage only, *hrwage*; and the second measure is based on the sum of the non-retirement wage and bonuses, *hrinc*. The

[118] The response rates of all these questions are 100% except the response rate on whether a respondent receive any in-kind (99%).

logarithms of the two measures (*lnhrwage, lnhrinc*) are then used in the empirical analysis. Figure 5.1 presents the two measures of the average real hourly wage income across six waves. The *hrwage* increases from 1.24 (yuan) in 1991 to 5.82 (yuan) in 2006. The *hrinc* curve almost parallels to the *hrwage* curve, and it increases from 1.50 (yuan) in 1991 to 6.41 (yuan) in 2006. Furthermore, a measure of wage income using the CHNS definition - sum of non-retirement wage plus bonuses and other cash/in-kind income (*hrinc'*), is also draw in Figure 5.1. It shows that adding other cash/in-kind income component only slightly increases in real hourly wage income. However, this measure cannot be utilized further since information on other cash/in-kind income is only surveyed in one year.

According to summary statistics in Table 5.1b, the mean of the two measures of the real hourly wage income for the total sample is 3.22 and 3.62 Chinese yuan, respectively, which are, for example, close to the Liaoning's 2006 minimum wage rate (range 3.00-5.00 Chinese yuan)[119]. On average, male and rural workers are better paid than female and urban workers, respectively.[120] Figures 5.2a and 5.2b present the average real hourly wages (*hrwage*) by gender and wave, and by urban/rural area and wave, respectively. The increasing trend in average hourly wages is evident. Overall, men have higher hourly wages than women for all waves. The gender wage gap increases from 1991 to 2006. However, according to Figure 5.2b, the rural-urban wage gap first increases from 1991 to 2000 and then starts to decrease after 2000. Indeed the average the hourly wage for both urban and rural workers are close to that for the total sample in 2006.[121]

5.5.3 Explanatory Variables

The first measure of health is the SRHS which is defined in Chapter 4 and includes four dummy variables *excellent*, *good*, *fair* and *poor* (the reference category), which take value of one if an individual has excellent health, has good health, has fair

[119] The minimum wage rates in China are varied by provinces (and urban/rural areas in each provinces) and municipalities.

[120] Recall from Section 3.2.3 in Chapter 3, *urban* takes value of 1 if people are residents in urban neighbourhood or county town areas (and zero if people are residents in suburban villages or rural villages), which is not defined based on the urbanization status. This might be the reason why rural workers are higher paid than urban workers. It is also worth to mention that the sample used in this chapter does not include agricultural workers or self-employed workers since they normally receive income from other sources.

[121] A similar trend is also found by applying the sample without excluding agricultural workers and/or self-employed workers.

health or has poor health, respectively. As shown in Table 5.1b, individuals with fair or with excellent health status have the highest hourly wages for all the sample groups. While for all the sample groups with good health status, the hourly wages are unexpectedly lower than those with poor health status. Possible factors driving this statistical result can be the other indirect effects, for example, differences in age, qualifications and work experience. As mentioned in the earlier section, following Cai (2009), the SRHS is measured continuously in the 2SLS analysis of this study. Therefore, h_{it}^{2SLS} is a continuous variable, namely *health2*, where a value of 1 to 4 indicating individual self-reported health status from poor to excellent.

The second measure of health is the BMI based on the WHO classification criteria, which is also used in the previous chapter. It includes four dummy variables: *uweight_who* (BMI<18.5), *nweight_who* (BMI 18.5-25), *oweight_who* (BMI 25-30), and *obese_who* (BMI>=30), where the dummy variable for normal weight, *nweight_who*, is the reference category. Similar to the above, a continuous measure of BMI (*bmi*) is used in the 2SLS analysis. As shown in Table 5.1c, individuals' hourly wages increase with BMI for all the sample groups, except for female and rural workers. For female workers, underweight and overweight have the highest and the lowest hourly wage income (*hrinc*), respectively. For rural workers, both of the two measures of hourly wage income are lowest for those with normal weight.

The other explanatory variables which have been found to affect wages (Haveman, *et al.*, 1994; Contoyannis and Rice, 2001; Gambin, 2005; and Jackle *et al.*, 2007)[122] include gender, age, educational attainment, urban/rural area, region of residence, marital status, presence of children in household, occupational status[123], employment position[124], type of work unit[125], number of employees in work unit[126], and whether

---

[122] However, this chapter fails to include two explanatory variables in the wage model: one is the the work experience; and the other one is ethnicity since it is not surveyed for waves 1991, 2004 and 2006.
[123] Recall from Chapter 3 the 8 dummy variables for the occupation category are: *job_1*, senior professional (or technical) personnel; *job_2*, junior professional (or technical) personnel; *job_3*, administrator/executive/manager; *job_4*, office staff; *job_6*, technical, skilled worker; *job_7*, non-technical, non-skilled worker; *job_8*, driver and service worker; *job_9*, engagement in small commercial household business, handicraft and other, where *job_1* is the reference category.
[124] Recall from Chapter 3 that the 2 dummy variables for the job position category are: *employee*, worker who works for another person or enterprise, and paid family worker; and *otherposi*, unpaid family worker, temporary worker and other, where the latter is the reference category.
[125] Recall from Chapter 3 that the 3 dummy variables for the type of work unit are: *privatesec*, individual or private enterprise, three source invested enterprise, household business, and family contract; *government*, state enterprise or institute/governmental unit; and *typeother*, small collective enterprise, three-capital enterprise and other, where *privatesec* is the reference category.

or not an individual has a second occupation. In addition, dummy variables for the survey year are also included in the wage model.

Appendix 5.1 shows average age and educational attainment by self-reported health status. As expected, the better the health status, the lower the average age, and the higher the educational attainment. Although the average level of educational attainment for workers with good health status is higher than that for workers with poor health status, the average age for the former is younger than the latter. Therefore, workers with good health status are likely to have less years of experience than workers with poor health status, which may reduce average hourly wages among this group. Unfortunately, this chapter is unable to obtain information on work experience since information on the age the respondent left school is not available from the CHNS. The average age and educational attainment by BMI categories are presented in Appendix 5.2. It can be seen that the BMI positively correlate with average age, which may possibly explain the positive wage relationship, but there is a less clear relationship between educational attainment and BMI Therefore, workers who are overweight or obese are like to have higher hourly wages. In addition, workers of normal weight have the lowest level of educational attainment among the four BMI categories.

Summary statistics of the other explanatory variables are provided in Table 5.1d. On average, urban workers are better educated than rural workers. The share of urban workers who have technical or vocational degree qualification (*techvoc*) and who have university or higher degree qualifications (*university*) account for 16% and 13%, respectively, which are more than double the corresponding shares for rural workers. The distribution of educational attainment for both male and female workers is similar to that of the total sample. Compared with female workers, male workers have higher share of employment in senior professional/technical personnel (*job_1*), administrator/manager (*job_3*) and technical worker (*job_6*), which are 9%, 15% and 21%, respectively. For all the sample groups except for the rural sample, more than 90% are employees. In terms of the type of work unit, 20% and 37% of rural residents working in private sector and in state enterprise or government units,

[126] Recall from Chapter 3 that the 4 dummy variables for the size of the work unit are: *firm020*, number of employees between 0 and 20; *firm20100*, number of employees between 20 and 100; *firmmt100*, number of employees more than 100; and *firmunk*, unknown or no response, where *firm020* is the reference category.

respectively. For urban residents, working in state enterprise or government units has a very large share of the three types of work unit, which account for 70%, whereas working in private sector only accounts for 7%. Moreover, urban residents have lower share in small size work units (9%) but higher share in large size work unit (54%) compared with rural residents.

5.5.4 Instruments – 2SLS Estimation

In estimating the effect of self-reported health on wages, most of the existing literature use detailed specific health conditions or diseases to instrument self-reported health to account for endogeneity and measurement error of self-reported health (Bound, 1991; and Cai, 2009). Some studies also use other instruments such as family assets (Lee, 1982), non labour income, socio-economic status of respondent's family members, and area measure of environment and health facilities (Haveman *et al.*, 1994; and Jackle and Himmler, 2007). This chapter follows the strategy of using specific health conditions as instruments of SRHS and uses information on medically diagnosed diseases and injury which are reported by individual.

The first instrument used in the IV estimation is a dummy variable indicating whether or not an individual has high blood pressure (*hypertension*), which is based on a survey question "*Has a doctor ever told you that you suffer from high blood pressure?*" The second instrument used to account for the potential endogeneity and measurement error of SRHS is based on three questions. The CHNS first asks respondents "*During the past 4 weeks, have you been sick or injured? Have you suffered from a chronic or acute disease?*" If the answer is yes, then respondents are asked "*What did you do when you felt ill?*", with four possible answers "*1 = self care; 2 = saw the local health worker; 3 = saw a doctor (clinic, hospital); 4 = did not pay any attention*".[127] Those who answered either saw the local health worker or a doctor are then asked "*What was the doctor's diagnosis of your illness or injury?*". The answer contains a list of types of illness, including: infectious/parasitic disease, heart disease, tumour, respiratory disease, injury, alcohol poisoning, endocrine disorder, haematological disease, mental/psychiatric disorder, mental retardation, neurological disorder, eye/ear/nose/throat/teeth disease, digestive disease, urinary disease, sexual

[127] In CHNS 1991, the question asks respondents "*Did you go to a hospital for the illness or injury?*".

dysfunction, obstetrical/gynaecological disease, neonatal disease, dermatological disease, muscular/rheumatologic disease, genetic disease, old age/mid-life syndrome, and other. Therefore, a dummy variable (*ill*) is created and takes on the value 1 if individuals answered any of the above diseases and zero otherwise. The two dummy variables are used as instruments for self-reported health status.

As shown in Table 5.1d, the number of people who have high blood pressure and ever have any of the above lists of diseases based on doctor' diagnosis during the last month accounts for approximately 3% and 19%, respectively. Moreover, a robustness test is also performed by using two alternative instruments for SRHS, which are two continuous variables indicating the number of years having high blood pressure (*hypyear*) and number of days being sick or injured during the last month (*illday*), respectively. On average, the number of years having high blood pressure is 0.16 for the Chinese workers. The number of years having high blood pressure for male workers is about 3 times larger than female workers. For urban workers it is twice as much as that for rural workers. The average days of being sick or injured during the last month for both genders are 0.22 which are the same to that for the total sample. In addition, urban workers have more average sick days than their rural counterparts.

In estimating the effect of BMI on wages, most of the existing literature use BMI of the biological family members (parents, children, and siblings) to instrument BMI (Cawley, 2004; Brunello and D'Hombres, 2007; and Kline and Tobias, 2008). Some studies also use other instruments such as relative food prices, non labour income of the worker and non labour income of all other household members (Thomas and Strauss, 1997). Following Thomas and Strauss (1977), Arcand and Labar (2005) use dummy variables to identify the person in the household preparing meals, the same for the person planning meals, smoking, the proportion of food purchases made in state store, and on the free market as instruments for BMI. This chapter, however, is unable to follow the above studies in terms of using BMI of the biological family members, food prices, or the identity of the person in the household preparing (or planning) meals since these information are not surveyed or fully surveyed in CHNS.[128]

In order to account for the endogeneity of BMI, this chapter uses two instruments. The first instrument of BMI is a dummy variable indicating the presence of one or more obese working-age adults[129] (BMI>=30) in respondent's household (*hhobese*). The second instrument, *energy*, is a continuous variable indicating the individual daily average energy intake (in kilocalories). A robustness test is also performed by using two alternative instruments for BMI. The first one is a continuous variable indicating the share of working-age adults who are obese in the area in which the respondent lives (*obese_t4*). The second one is also a continuous variable indicating the daily average energy intake of other working-age adults in respondent's household (*energy_hh*). As shown in Table 5.1d, 2% of the Chinese workers have at least one obese working-age adults in their household. The individual daily average energy intake for the Chinese workers is approximately 24 kilocalories. On average, men and rural residents have higher energy intake than women and urban residents, respectively. Additionally, the average share of obese working-age adults in respondent's living area (*oweight_t4*) and the average energy intake of other household members (*energy_hh*) are 2% and 20 kilocalories, respectively.

5.5.5 Exclusion Restrictions – Heckman Sample Selection Estimation

Similar to Cai (2009), variables for occupational status, employment position, type of work unit, number of employees in work unit, and whether or not an individual has a second occupation enter the wage equation but not the selection equation, which are served as exclusion restrictions for employment. Cai (2009) uses the presence of children in the household as exclusion restriction for employment. In analysing American female labour supply, Mroz (1987) assumes that both children and other source of income (family income) have impact on labour supply but do not have a direct impact on wages. Since the presence of children is assumed to have impact on wages for Chinese working-age adults, this analysis uses a dummy variable

[128] Information about the identity of the person in the household preparing meals, the same for the person planning meals, and the proportion of food purchases made in state store and on the free market is only surveyed in CHNS 1991 and 1993.

[129] The analysis assumes that the lifestyle of older people is different from that of the working-age people. Moreover, the BMI classification for children is different from that for adults.

indicating whether or not a household (where the respondent lives) receives non-wage income (*hhnwinc*) during the past year as the exclusion restriction for wages.[130] The household non-wage income refers to any income from household businesses, farming, fishing, gardening, livestock, rentals of household assets, boarders, welfare funds, relatives and friends during the past year. In addition, a robustness test is performed using an alternative exclusion restriction for wages, which is a dummy variable the presence of other household members who have individual non-wage income during the past year in the household in which the respondent lives (*oindnwinc_hh*). The individual non-wage income refers to individual income from household businesses, farming, fishing, gardening, and livestock. The total sample used in analysis which addresses the potential sample section bias includes 16,086 observations from 8,631 individuals. It includes all the Chinese working-age adults except for those who are either agricultural workers or self-employed. As shown in Table 5.1d, 52% of the total sample received household non-wage income and this does not vary by gender. For the urban selection sample, 30% received household non-wage income. While for the rural selection sample, the share is approximately 2.5 times larger than their urban counterparts. In addition, for the total selection sample and for both genders, 11% have other household members who receive individual non-wage income. The share of rural selection sample received individual non-wage income (18%) is about 4 times larger than their urban counterparts (4%).

5.6 Empirical Results

5.6.1 The Impact of SRHS on Wage Income

The estimation results of the impact of SRHS on hourly wage using the pooled OLS, RE and FE methods are shown in Tables 5.2-5.4, where the Hausman tests of comparing RE and FE are also presented at the bottom of these tables. The dependent variable used in Tables 5.2-5.4 is the first measure of individual wage income, *lnhrwage*. Sensitivity analyses of the effect of SRHS on wage income by excluding and including the agricultural and self-employed sample and using two alternative hourly wage income measures are presented in Table 5.5 and Table 5.6, respectively.

[130] The continuous measure of household non-wage income is not utilized since it is found to have statistically significant effect on wages whereas the dummy variable for having household non-wage income does not.

The Impact of SRHS on Wages - Total Sample

Table 5.2 shows the estimation results of the SRHS impact on wages for the total sample using the pooled OLS, RE and FE methods. According to the second and the third column of Table 5.2, the pooled OLS estimation results are similar to the RE estimation results. For example, individuals who have excellent health status are predicted to earn 11.1% and 12.6% more than those who have poor health status under the pooled OLS and the RE estimates, respectively. Comparing the estimated coefficients on the SRHS variables using the FE and the RE methods, the former are larger than the latter in terms of magnitude. The Hausman test strongly rejects the null hypothesis that RE provides consistent estimates, suggesting that FE model is preferable. Compared to the pooled OLS model, controlling for individual unobserved factors increases the effect of SRHS on wages for the Chinese workers. The FE estimation result in Table 5.2 suggests that there is a positive effect of better health status relative to the poor health status. However, in contrast to the effect on employment (found in Chapter 4), the estimated coefficient increases as the level of health increases. For example, the coefficient of good health status (17.0%) is higher than that of fair health status (13.3%).

The estimation results for non-health variables are also reported in Table 5.2. According to the preferable FE estimation, the number of statistically significant coefficients is much lower than those using the pooled OLS. For example, in the pooled OLS estimates, individuals with technical/vocational degree qualifications and with university or higher degree qualifications are predicted to earn 7.5% and 21.9% more than individuals with no qualifications, respectively. However, based on the FE estimates, neither of holding technical/vocational degree qualifications nor holding university or higher degree qualifications has a statistically significant effect on hourly wages. Additionally, the pooled OLS result shows that the hourly wages for non-skilled workers (*job_7*) and for service workers (*job_8*) are 16.9% and 16.0% lower, respectively, comparing to the hourly wages for senior professional workers (*job_1*). While neither of them has a statistically significant effect on wages when using the FE method. These results may reflect that individual educational attainment and occupational status are almost time-invariant, given the fact that only within variation of the data is used in the FE estimation.[131]

The Impact of SRHS on Wages - Men versus Women

The estimation results for men and women are separately shown in Table 5.3. According to the Hausman statistic, for both genders, $\chi^2(35)$ has p-value = 0.00, leading to a strong rejection of the null hypothesis which implies that the FE estimates are preferable. For men, the estimated effects of the SRHS variables are the same in sign and increase in terms of magnitude between the pooled OLS and the FE methods except of the good health status. For example, men who good health status are predicted to earn 9.8% and 14.1% more than those who have poor health status in the pooled OLS estimates and the FE estimates, respectively. However, for women, there is no significant effect of health using the pooled OLS. For men, both having excellent and having good health status tend to increase the hourly wages by about 14.3%, and having fair health status tends to increase the hourly wages by 10.4%. For women, both having excellent and having good health status are predicted to earn about 22.5% more than their poor health counterparts, and those with fair health status tend to increase the hourly wages by 17.9%. The difference in genders is consistent with that in Liu *et al.* (2008) on examining the effect of SRHS on *per capita* individual income using panel data from the CHNS (1991, 1993 and 1997), and that in Contoyannis and Rice (2001) for British (the coefficients for *excellent* and *good* are 0.9% and 0.4%, respectively) men and (the coefficients for *excellent* and *good* are 2.7% and 1.6%, respectively) women. The estimated coefficients are not directly comparable due to the use of different measure of productivity (Liu *et al.*, 2008) and health (Contoyannis and Rice, 2001). Liu *et al.* (2008) use a real average household income from wage, household gardening, farming, livestock, fishing, business, subsidies and benefits per working household member to measure income productivity. The reference group of self-reported health status used in Contoyannis and Rice (2001) refers to people who have fair health or worse. The finding is, however, different to the FE estimates in Gambin (2005) who finds that in a total of 14 western countries, there is a statistically significant

[131] By using data from the same survey, Arcand and Labar (2007) argue that the variation of education over time for a given individual might introduce measurement error since the most working adults would have finished their study.

relationship between SRHS and wages for men, and almost no statistically significant effects for women.

The Impact of SRHS on Wages - Urban versus Rural

The Hausman statistic in Table 5.4 rejects the null hypothesis that RE provides consistent estimates is also found for both urban and rural workers, suggesting that the FE model is preferable. For urban workers, all the SRHS variables are significantly different from zero at the 1% level. Both having excellent and having good health status tend to increase the hourly wages by about 21.0%, and having fair health status tends to increase the hourly wages by 16.5%. For rural workers, having excellent tends to increase the hourly wages by about 13.9% but this is only significant different from zero at the 10% level. Futher, there is no statistically significant of having good or fair health status. Overall, the effect of SRHS on wages is larger for urban than for rural workers, which is consistent with the findings relating to employment. However, this result is contrary to that in Liu *et al.* (2004). The different finding is probably because Liu *et al.* (2004) use the household income *per capita* to measure individual return to health between urban and rural population in China, where the household income *per capita* is defined as the real average household income from wage, household gardening, farming, livestock, fishing, business, subsidies and benefits, food coupons, housing subsidy, child care subsidy and other subsidies per working household member.

Sensitivity Analysis

Table 5.5 presents the estimation results of the SRHS impact on wages for all the sample groups with and without excluding the agricultural and the self-employed workers who have positive wages based on the preferred FE method.[132] It shows that including the agricultural and the self-employed sample reduces the coefficients of all the SRHS variables for the total sample and for both female and urban workers. For both male and rural sample, there is no statistically significant effect of SRHS on wages after including the agricultural and the self-employed sample. As mentioned in the last section, the proportion of agricultural and self-employed workers reporting wage income is small since most of agricultural and self-employed workers earn

[132] For comparison, the FE estimation results which are presented in the top part in Tables 5.5 are the same to those in Tables 5.2-5.4.

their income from alternative sources such as farming and home business, thus adding this sub sample might introduce measurement error.

Table 5.6 presents the sensitivity analysis based on the two hourly wage income measures. The top and bottom part of Table 5.6 show the FE estimates using *lnhrwage* and *lnhrinc* as the dependent variable, respectively.[133] The results suggest that adding component of bonuses to the wage income increase the positive effect of better health status on wage income for almost all the sample groups. Therefore, using both wages and bonus to measure labour productivity yields slightly more pronounced results than using wages on its own.

Addressing the Potential Endogeneity of SRHS

The previous analysis assumes that the discretely measured of self-reported health status is exogenous. In accounting for its potential endogeneity this chapter applies the 2SLS estimator as described in Section 5.3 and the estimation results from both the first stage and second stage regression are shown in the top part of Table 5.7, where self-reported health status is treated as a continuous variable. In addition, the over identification test statistic (the Hansen J statistic), the Durbin-Wu-Hausman test statistic for the endogeneity of SRHS, and the F statistic for the joint significance of the coefficients on the two instruments are also presented. The middle and bottom part of Table 5.7 present the estimation results based on the pooled OLS and the FE, respectively, without considering the potential endogeneity and discrete nature of self-reported health status.

The insignificance of the Hansen J statistics in Table 5.7 suggests that the instruments are correctly excluded from the wage equation. The F statistics are significant at the 1% level for all the sample groups, indicating that the two instruments have significant explanatory power for SRHS after controlling for the other explanatory variables. The lower part of the 2SLS estimation in Table 5.7 shows that both having hypertension and having other medically diagnosed diseases or injury have statistically significant and negative effect on health, which are consistent with expectations. However, for all the sample groups, the null

[133] Again, for comparison, the FE estimation results which are presented in the top part in Tables 5.6 are the same to those in Tables 5.2-5.4.

hypothesis of exogeneity of SRHS cannot be rejected since the Durbin-Wu-Hausman test statistics are not statistically significant, suggesting that the pooled OLS estimation is more appropriate than the 2SLS estimation. The sensitivity analysis of using alternative instruments which is presented in Appendix 5.3 shows similar results to that in Table 5.7. Compared to the FE estimates in the bottom part of Table 5.7, the effect of SRHS for all the sample groups becomes statistically insignificant when after controlling for the potential endogeneity of SRHS. The FE estimation result in Table 5.7 shows that better self-reported health status positively affects wages for Chinese workers.

Addressing the Potential Sample Selection Bias

The top part of Table 5.8 shows the estimation results of the effect of SRHS on wages using the Heckman selection method along with the estimated ρ_2 and the Wald test statistic (H_0: $\rho_2 = 0$). The middle and bottom part of Table 5.8 presents the estimation results based on the pooled OLS and the FE method[134], respectively, without considering the sample selection bias.

It can be seen for Table 5.8 that the Wald test strongly rejects the null hypothesis that pooled OLS provides consistent estimates for all sample groups within the population, thus the Heckman selection model is preferable.[135] For example, the correlation ρ_2 for the total sample is -0.27, which implies that the unobserved determinants of wages and employment are negatively correlated. After controlling for sample selection bias, none of the coefficients of SRHS is statistically significant except for the coefficient of excellent health for the total and the urban sample, which are significantly different from zero at the 10% level. For the total and the urban sample, having excellent health are predicted to increase hourly wages by 6.4% and 8.9% compared to their poor health counterparts, respectively. The lower part of the Heckman estimates in Table 5.8 shows the effects of SRHS and exclusion restriction in the employment equation. Overall, the statistically significant and positive relationship between having better health status and employment is consistent with that in the previous chapter. In addition, having household non-wage

[134] The pooled OLS and FE estimation results are the same to those in Tables 5.2-5.4.

[135] The effect of SRHS on wages based on a two-step Heckman method is also estimated in this analysis, but not reported. The estimation results are similar to those using the maximum likelihood method.

income negatively affects the individual employment probability. Sensitivity analysis of using alternative exclusion restriction which is presented in Appendix 5.4 shows the similar result compared to the Heckman estimates in Table 5.8.

The presence of sample selection bias for both Chinese men and women is found to be different from the international literature. Forbes et al. (2010) find that there is sample selection bias present for Australian men but not for women. In Pelkowski and Berger (2004), the negative effect of permanent health problems on wages is found to be decreased for American women but increased for men after account for selection bias. According to the Heckman estimates, the statistically significant and positive relationship between having better health and wages disappears after correcting for the sample selection bias. There is no statistically significant relationship between better health and wages for the sub sample groups except for the urban worker, where having excellent health has a statistically and positive effect on wages (8.9%). Based on a sample of rural workers from the CHNS 1993, Wei (2004) uses Heckman selection model and concludes that there is no statistically significant relationship between health (both the physical and the psychological measure of health) and wages.

5.6.2 The Impact of BMI on Wage Income

The estimation results of the impact of BMI on wage income by controlling the effect of SRHS based on the pooled OLS, RE and FE methods are shown in Tables 5.9-5.11. Hausman tests which compare RE and FE are also presented at the bottom of these tables. The dependent variable used in Tables 5.9-5.11 is the preferable measure of individual wage income, *lnhrwage*. As before the Hausman test rejects the null hypothesis that RE provides consistent estimates for all the sample groups, suggesting that FE model is preferable.

As shown in Table 5.9, none of the coefficients on BMI has a statistically significant effect on hourly wages, which is different to the descriptive statistics. Sensitivity analysis which examines the effect of other explanatory variable is shown in Table 5.12: Model (1) estimates for the effect of BMI only; Model (2) estimates for the effect of BMI by controlling for other explanatory variables but not for SRHS; and Model (3) controls for the effect of the other explanatory variables including SRHS, which is the same as the pooled OLS estimates in Table 5.9. According to Model (1), being underweight is negatively correlated with wages, whereas being overweight or obese is positively correlated with wages. However, all the statistically significant effects of BMI disappear once controlling for the other explanatory variables. In other words, the effects of BMI are accounted for by including the other explanatory variables. In addition, including controls for the effect of SRHS does not change the effect of BMI on wages. Recall the discussion from the previous chapter, the effect of body weight on labour market outcomes can work through the health channel and/or due to the presence of the weight-based discrimination. Therefore, assuming SRHS completely captures the impact of health on productivity, there is no evidence of weight-based wage discrimination for the Chinese workers.

The estimation results for men and women are shown separately in Table 5.10. The FE estimates suggest that there is no statistically relationship between BMI and hourly wages for men. For women, being obese is predicted to reduce wages by 22.4% compared to their normal weight counterparts (which is only significantly different from zero at the 10% level), whereas neither being underweight nor being overweight has a statistically significant effect on hourly wages. This result is consistent with that in Garcia and Quitana-Domeque (2006) for men and women in Denmark, Finland and Portugal. Table 6.11 shows the impact of BMI on hourly wages for urban and rural workers. For urban workers, there is no statistically relationship between BMI and hourly wages. For rural workers, only being overweight have a statistically significant and negative effect on hourly wages (but is only significantly different from zero at the 10% level), with a value of -6.9%. The sensitivity analysis which examines the effect of other explanatory variables for the sub sample groups are shown in Appendices 5.5 and 5.6, where the results are similar to that in Table 5.12 for the total sample.

Addressing the Endogeneity of BMI

The above analysis assumed that BMI is exogenous. In accounting for its endogeneity of BMI this chapter applies the 2SLS estimator which is similar to that for SRHS and the results from both the first stage and second stage regression are shown in the top part of Table 5.13, using a continuous measure of BMI. As before the over identification test statistic (the Hansen J statistic), the Durbin-Wu-Hausman test statistic for the endogeneity of SRHS, and the F statistic for the joint significance of the coefficients on the two instruments are also presented. The bottom part of Table 5.13 shows the estimation results based on the pooled OLS and the preferable method (FE) without considering the endogeneity of BMI.

The insignificance of the Hansen J statistics in Table 5.13 suggests that the instruments are correctly excluded from the wage equation. The F statistics are significant at the 1% level for all the sample groups, indicating that the two instruments have significant explanatory power for BMI after controlling for the other explanatory variables. The results of the second stage 2SLS regression in Table 5.13 show that both the presence of other obese working-age adults in respondent's household and having higher individual daily average energy intake is positively correlated with individual BMI. However, for all the sample groups, the null hypothesis of exogeneity of BMI cannot be rejected since the Durbin-Wu-Hausman test statistics are not statistically significant, suggesting that the pooled OLS estimation is more appropriate than the 2SLS estimation. This finding is consistent with that in Cawley (2004) for the population in the US. Although the exogeneity of BMI in Thomas and Strauss (1997) is rejected for both men and women in urban Brazil, their instruments fail to pass the test of over identification. The sensitivity analysis of using alternative instruments which is presented in Appendix 5.7 shows the similar result compared to that in Table 5.13.

The Hausman test statistics shown in the bottom part of Table 5.13 reject the null hypothesis that RE provides consistent estimates, suggesting that FE model is preferable. In terms of the previous literature, the impact of BMI on wages for population in western countries is mixed. Thomas and Strauss (1997) find a statistically positive effect of BMI on hourly wages for men but not for women in Brazil. For population in the US, Cawley (2004) finds a positive effect of BMI on hourly wages for black men and a negative effect for white women, but no statistically effect for white men or black women. Based on a pooled sample of working-age adults from nine European countries, Brunello and D'Hombres (2007) find a statistically significant and negative effect for both genders and the effect are larger for men than for women. For the Chinese workers, there is no statistically significant relationship between BMI and hourly wages for all the sample groups. In addition, the pooled OLS and the FE estimates for men are consistent with those in Arcand and Labar (2005) who use the same survey to examine the effect of BMI on hourly wages for the Chinese men.

Addressing the Selection Bias

The top part of Table 5.14 shows the estimation results of the effect of BMI on hourly wages using the Heckman selection method along with the estimated ρ_2 and the Wald test statistic (H_0: $\rho_2 = 0$). The middle and bottom part of Table 5.14 presents the estimation results based on the pooled OLS and the FE method[136], respectively, without considering the sample selection bias.

The Wald test strongly rejects the null hypothesis that pooled OLS provides consistent estimates for all the sample groups, suggesting the Heckman selection model is preferable. For all the sample groups within the population, ρ_2 is negative which implies that the unobserved determinants of wages and employment are negatively correlated. This finding is the same as that in estimating the effect of SRHS on hourly wages. Nevertheless, there is only a limited effect of sample selection bias since the Heckman selection estimation results for all the sample groups are similar to the pooled OLS estimates. For example, male workers who are overweight are predicted earn by 4.0 % and 4.2 % more compared to those with normal weight based on the pooled OLS and the Heckman method, respectively. For urban workers, being overweight are predicted to increase hourly wages by 3.3% and 3.5% compared to those with normal weight based on the pooled OLS and the Heckman method, respectively.

The lower part of the Heckman estimates in Table 5.14 shows that the effect of BMI in the employment equation is similar to that in the previous chapter. As before having household non-wage income negatively affects the individual employment probability. Sensitivity analysis of using alternative exclusion restrictions is presented in Appendix 5.8 and the results are consistent with Heckman estimation results in Table 5.14.

[136] The pooled OLS and FE estimation results are the same to those in Tables 5.9-5.11.

The FE estimates in the bottom part of Table 5.14, however, show that after controlling for individual specific unobserved factors, the statistically significant and positive relationship between being overweight and hourly wages both for men and for urban residents disappears. Recall obese female workers are predicted to reduce hourly wages by 22.40% compared to their normal weight counterparts. The result are larger than that in Cawley (2004) for American white women (-8.7%) based on the same method. For American white men, Cawley (2004) finds that being overweight increases hourly wages by 2.2%. Kline and Tobias (2008) find obesity is negatively correlated with hourly wages for both genders in the UK. However, the result is not directly comparable since the authors use different econometric strategies. Assuming SRHS captures the impact of health on productivity, there is evidence of weight-based wage discrimination for women and for the rural workers.

5.7 Conclusions

This chapter contributes to the literature by examining the relationship between health and labour productivity for the Chinese workers. The sensitivity analysis based on two constructed variables of the real hourly wage income shows that using the real hourly wage plus bonuses to measure labour productivity yields more pronounced results than the use of the real hourly wage rate. Instead of using a single health measure, the same measures of health in the previous chapter are utilized: SRHS and BMI which is based on the international cutoff points for overweight and obesity.

There is a statistically and positive effect of having excellent health on wages for the Chinese workers. As for the sub sample groups, having excellent health is positively correlated with wages for the urban sample, whereas the SRHS does not have a statistically significant effect on wages for the male, the female and the rural sample. Based on the international BMI classification, neither of three BMI categories has s statistically significant effect on wages for the total, the male or the urban sample. For female workers, being obese is predicted to reduce hourly wages by 22.40% compared to their normal weight counterparts. This result is larger than that of the international literature. In addition, being overweight is found to have a statistically significant and negative effect on wages for the Chinese rural workers. Therefore, further efforts in improving health in China could focus on specific groups of the population such as obese women and rural workers who are overweight. The local government could make some reasonable policies and implement these policies effectively such as adding more tax on consuming fast food or restricting the number of fast food restaurants should be given greater concern. Moreover, the health authorities could make an effort to promote a healthy diet and lifestyle such as opening a health TV channel.

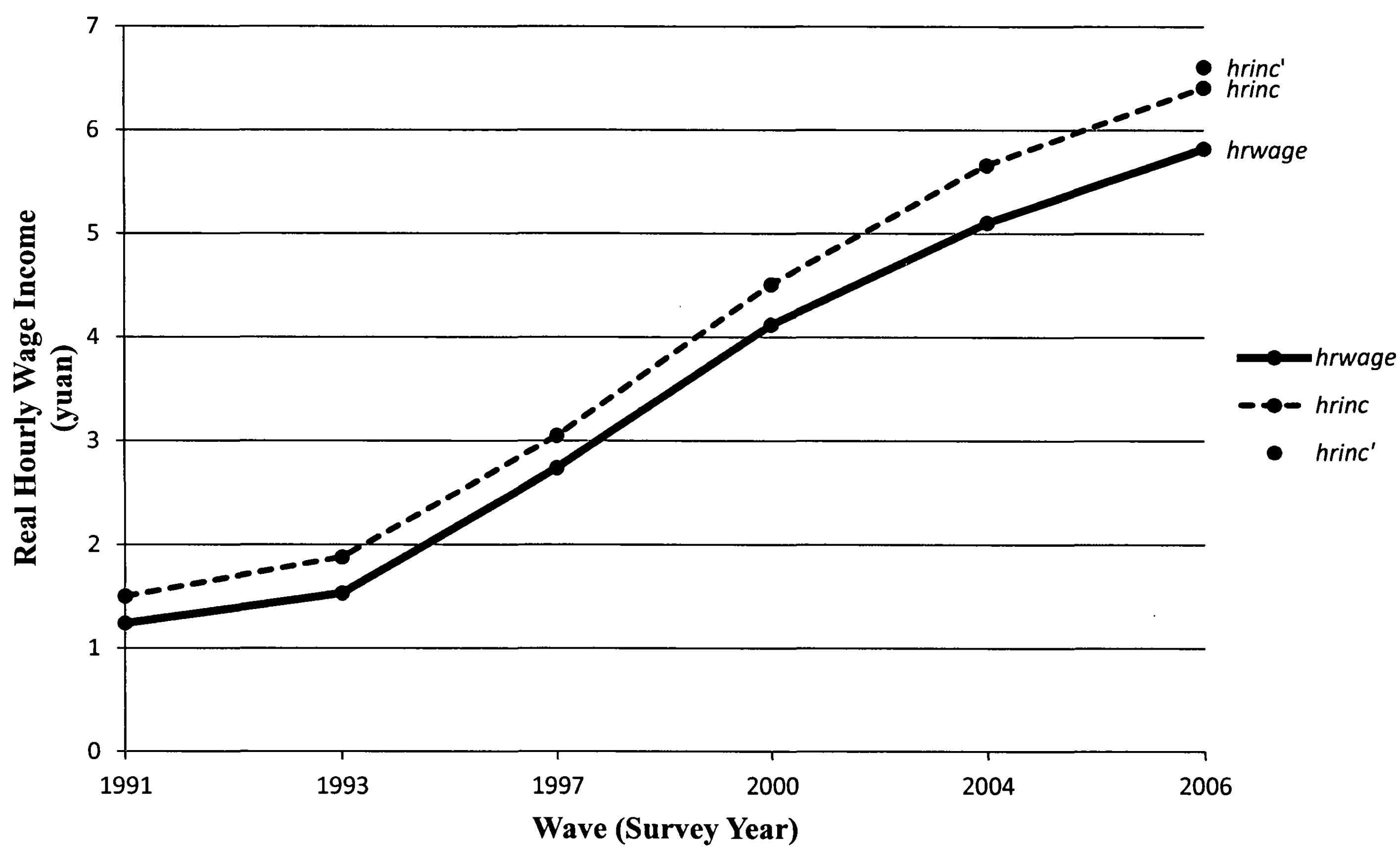
This chapter allows for the panel nature of the CHNS data and employs the FE model to control for individual unobserved factors. Compared to the pooled OLS method, controlling for individual unobserved factors using the FE method almost doubles the positive effect of better health on wages for the Chinese workers. In other words, without controlling for the individual unobserved factors underestimates the positive effect of the SRHS on wages. Furthermore, this chapter applies a 2SLS and a Heckman selection model to account for potential endogeneity of health and sample selection bias, respectively. Based on the 2SLS method, the null hypothesis of exogeneity of health cannot be rejected for all sample groups within the Chinese population, implying that the pooled OLS are preferred to the 2SLS. However, based on the Heckman method, the null hypothesis that pooled OLS provides consistent estimates is rejected, suggesting the presence of sample bias in estimating the effect of health on hourly wages for the Chinese workers. The statistically significant and positive relationship between better health and wage largely disappears after controlling for the sample selection bias.

There are some limitations to the analysis in this chapter. In estimating the effect of health on wages when allowing for its endogeneity, this chapter does not consider the discrete nature of self-reported health status. Lee (1982) argues that failure to take into account the discrete nature of the indicators might introduce econometric problems[137]. The sample used in the analysis does not include the agricultural workers and self-employed because most of them reporting their income from other sources. Therefore, future analysis could focus on this group of the population by using a net profit earned from agricultural work and self-employment to measure labour productivity. In addition, this study applies different methods to separately address different econometric issues (such as endogeneity and sample selection bias).

[137] For details, see Nerlove and Press (1973) and Domencich and McFadden (1975).

The current methods to account for sample selection bias in panel models are still in an experimental stage and remain questionable. Accordingly, further efforts are expected to extend the analysis to consider the above problems and a more satisfactory method to address those econometric issues is required.

Figure 5.1 Real Hourly Wage Income by Wave



Source: Author's calculations using CHNS data.

Figure 5.2a Average Real Hourly Wages by Gender and Wave

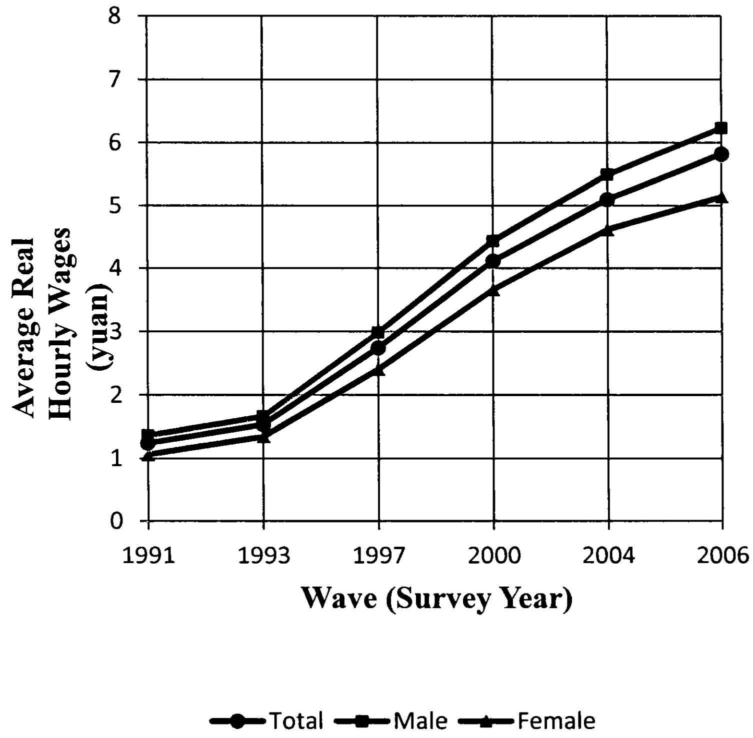
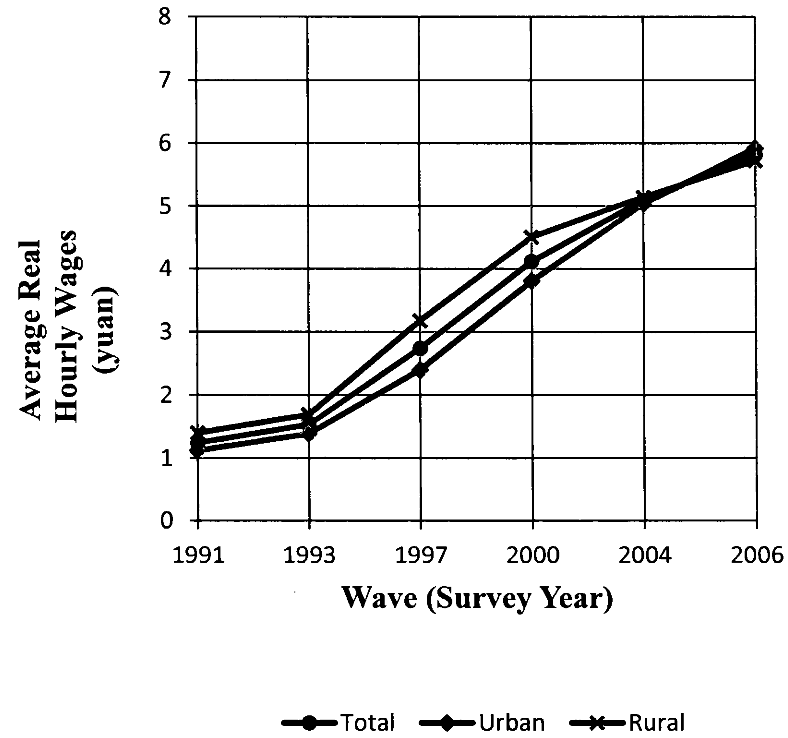


Figure 5.2b Average Real Hourly Wages by Urban/Rural and Wave



Source: Author's calculations using CHNS data.

Table 5.1a Summary Statistics - Annual Wages, Bonuses, and Other Cash/In-kind Income

Year	Real Annual Wages (Chinese yuan)		Real Annual Bonuses (Chinese yuan)			Real Annual Other Cash/In-kind Income (Chinese yuan)			Total	
	Mean (SD) (%)	N	Mean[a] (SD)	Mean[b] (SD) (%)	N (%)[1]	Mean[c] (SD)	Mean[d] (SD) (%)	N (%)[2]	Mean (SD) (100%)	N
1991	2,966.87 (1,555.01) (82%)	2,202	1,002.47 (1,196.56)	647.83 (1,054.54) (18%)	1,481 (67%)	-	-	-	3,641.11 (1,960.02) (100%)	2,202
1993	3,668.04 (3,195.97) (81%)	1,846	1,434.90 (1,617.75)	853.35 (1,509.88) (19%)	1,097 (59%)	-	-	-	4,520.74 (3,595.35) (100%)	1,846
1997	5,847.27 (3,779.63) (90%)	2,239	1,423.05 (1,472.39)	636.86 (1,214.97) (10%)	999 (45%)	-	-	-	6,482.21 (4,068.91) (100%)	2,239
2000	8,606.95 (7,945.06) (92%)	1,618	1,808.45 (1,795.02)	799.63 (1,487.66) (8%)	714 (44%)	-	-	-	9,405.00 (8,164.37) (100%)	1,618
2004	11,119.65 (6,620.36) (90%)	1,562	2,889.22 (3,911.90)	1182.01 (2,857.82) (10%)	640 (41%)	-	-	-	12,303.45 (7,684.84) (100%)	1,562
2006	12,980.94 (7,956.75) (89%)	1,648	3,142.83 (4,094.08)	1237.53 (2,993.44) (8%)	647 (39%)	3,812.93 (6,468.58)	421.09 (2,455.10) (3%)	182 (11%)	14,635.90 (9,519.15) (100%)	1,648
Total	6,851.14 (6,569.61) (88%)	11,115	1,702.09 (2,391.36)	868.57 (1,932.55) (11%)	5,578 (50%)	3,812.93 (6,468.58)	62.43 (916.77) (1%)	182 (2%)	8,046.05 (7,301.95) (100%)	11,115

Source: Author's calculations using CHNS data.
Mean[a] denotes average real annual bonuses only for people who have bonuses.
Mean[b] denotes average real annual bonuses for people who have wage income.
Mean[c] denotes average real annual other cash/in-kind income only for people who have other cash/in-kind income.
Mean[d] denotes average real annual other cash/in-kind income for people who have wage income.
[1] The share of people who have bonuses.
[2] The share of people who have other cash/in-kind income.

Table 5.1b Average Real Hourly Wage Income by Self-reported Health Status (pooled waves, 1991-2006)

	Average Real Hourly Wage Income (Chinese yuan): *hrwage* *hrinc*				
Self-reported Health Status	**Total (N=11,115)**	**Men (N=6,514)**	**Women (N=4,601)**	**Urban (N=6,102)**	**Rural (N=5,013)**
excellent	3.46	3.79	2.95	3.10	3.93
	3.82	4.19	3.25	3.47	4.27
(Sample Size: %)	(18%)	(19%)	(17%)	(19%)	(18%)
good	3.07	3.40	2.60	2.82	3.35
	3.46	3.84	2.93	3.25	3.71
(Sample Size: %)	(61%)	(60%)	(61%)	(59%)	(62%)
fair	3.49	3.72	3.16	3.29	3.75
	3.95	4.23	3.56	3.76	4.20
(Sample Size: %)	(19%)	(19%)	(20%)	(20%)	(18%)
poor	3.18	3.31	2.98	3.06	3.36
	3.51	3.65	3.29	3.45	3.59
(Sample Size: %)	(2%)	(2%)	(2%)	(2%)	(2%)
Total	3.22	3.53	2.78	2.97	3.53
	3.62	3.98	3.11	3.40	3.90
(Sample Size: 100%)	(100%)	(100%)	(100%)	(100%)	(100%)

Source: Author's calculations using CHNS data.

Table 5.1c Average Real Hourly Wage Income by BMI Categories (pooled waves, 1991-2006)

BMI Categories	Average Real Hourly Wage Income (Chinese yuan): *hrwage* *hrinc*				
	Total (N=10,016)	**Men (N=5,083)**	**Women (N=4,213)**	**Urban (N=5,572)**	**Rural (N=4,444)**
WHO Classification[1]:					
uweight_who	2.76	2.80	2.73	2.21	3.57
	3.19	3.15	3.23	2.62	4.02
(Sample Size: %)	(6%)	(5%)	(8%)	(6%)	(6%)
nweight_who	3.04	3.25	2.75	2.81	3.31
	3.44	3.69	3.11	3.24	3.68
(Sample Size: %)	(72%)	(70%)	(74%)	(71%)	(73%)
oweight_who	3.79	4.32	2.77	3.53	4.14
	4.25	4.88	3.03	4.00	4.58
(Sample Size: %)	(20%)	(23%)	(16%)	(21%)	(19%)
obese_who	4.25	4.85	2.94	3.79	4.90
	4.65	5.32	3.17	4.18	5.31
(Sample Size: %)	(2%)	(2%)	(2%)	(2%)	(2%)
Total	3.19	3.51	2.76	2.94	3.51
	3.61	3.97	3.11	3.37	3.90
(Sample Size: 100%)	(100%)	(100%)	(100%)	(100%)	(100%)

Source: Author's calculations using CHNS data.

[1] Recall from Chapter 4 the WHO criteria of BMI cutoffs are: underweight (BMI<18.5); healthy weight (BMI 18.5-25); overweight (BMI 25-30); and obese (BMI>=30).

Table 5.1d Summary Statistics - Other explanatory Variables and Instruments (pooled waves, 1991-2006)

Variables	Mean (Standard Deviation)				
	Total	Male	Female	Urban	Rural
Other Explanatory Variables[1]					
male (%)	0.60 (0.49)	-	-	0.57 (0.50)	0.63 (0.48)
age (in years)	36.42 (10.19)	38.39 (10.70)	33.52 (8.59)	37.24 (9.74)	35.43 (10.61)
age16_25 (%)	0.19 (0.39)	0.15 (0.36)	0.24 (0.43)	0.15 (0.36)	0.24 (0.42)
age26_35 (%)	0.30 (0.46)	0.28 (0.45)	0.34 (0.47)	0.31 (0.46)	0.29 (0.46)
age36_45 (%)	0.32 (0.46)	0.30 (0.46)	0.34 (0.47)	0.34 (0.48)	0.28 (0.45)
age46_59 (%)	0.19 (0.40)	0.27 (0.44)	0.08 (0.27)	0.20 (0.40)	0.19 (0.39)
noedu (%)	0.04 (0.21)	0.04 (0.20)	0.05 (0.22)	0.03 (0.17)	0.06 (0.24)
primary (%)	0.12 (0.32)	0.13 (0.33)	0.10 (0.30)	0.07 (0.26)	0.17 (0.38)
lowermid (%)	0.37 (0.48)	0.36 (0.48)	0.38 (0.49)	0.32 (0.46)	0.44 (0.50)
uppermid (%)	0.25 (0.43)	0.25 (0.43)	0.25 (0.43)	0.29 (0.45)	0.20 (0.40)
techvoc (%)	0.12 (0.33)	0.12 (0.32)	0.14 (0.34)	0.16 (0.37)	0.08 (0.27)
university (%)	0.10 (0.29)	0.10 (0.31)	0.08 (0.27)	0.13 (0.34)	0.05 (0.22)
urban (%)	0.54 (0.50)	0.52 (0.50)	0.58 (0.49)	-	-
married (%)	0.80 (0.40)	0.82 (0.38)	0.76 (0.42)	0.82 (0.38)	0.77 (0.42)
child0 (%)	0.43 (0.50)	0.45 (0.50)	0.39 (0.49)	0.42 (0.49)	0.43 (0.49)
child1 (%)	0.45 (0.50)	0.42 (0.49)	0.50 (0.50)	0.49 (0.50)	0.41 (0.49)
child2m (%)	0.12 (0.32)	0.13 (0.33)	0.11 (0.31)	0.09 (0.28)	0.16 (0.37)
Occupation[2]					
job_1 (%)	0.08 (0.26)	0.09 (0.28)	0.06 (0.23)	0.10 (0.30)	0.04 (0.20)
job_2 (%)	0.09 (0.29)	0.07 (0.25)	0.12 (0.33)	0.10 (0.30)	0.08 (0.27)
job_3 (%)	0.11 (0.32)	0.15 (0.36)	0.06 (0.23)	0.13 (0.34)	0.10 (0.29)
job_4 (%)	0.11 (0.32)	0.09 (0.29)	0.14 (0.34)	0.14 (0.35)	0.07 (0.26)
job_6 (%)	0.19 (0.39)	0.21 (0.41)	0.15 (0.36)	0.19 (0.39)	0.19 (0.39)
job_7 (%)	0.26 (0.44)	0.24 (0.43)	0.29 (0.45)	0.18 (0.38)	0.35 (0.48)
job_8 (%)	0.13 (0.34)	0.11 (0.31)	0.16 (0.37)	0.13 (0.33)	0.13 (0.34)
job_9 (%)	0.03 (0.17)	0.04 (0.18)	0.02 (0.157)	0.03 (0.16)	0.04 (0.19)
Employment Position[3]					
employee (%)	0.91 (0.28)	0.92 (0.28)	0.91 (0.28)	0.96 (0.20)	0.87 (0.34)
otherposi (%)	0.09 (0.28)	0.08 (0.28)	0.09 (0.28)	0.04 (0.20)	0.13 (0.34)
Type of Work Unit[4]					
privatesec (%)	0.13 (0.34)	0.13 (0.34)	0.13 (0.34)	0.07 (0.26)	0.20 (0.40)
government (%)	0.55 (0.50)	0.57 (0.50)	0.52 (0.50)	0.70 (0.46)	0.37 (0.48)
typeother (%)	0.32 (0.47)	0.30 (0.46)	0.35 (0.48)	0.23 (0.42)	0.43 (0.49)
Number of Employees in Work Unit[5]					
firm020 (%)	0.14 (0.35)	0.14 (0.35)	0.13 (0.33)	0.09 (0.29)	0.19 (0.39)
firm20100 (%)	0.27 (0.45)	0.27 (0.45)	0.28 (0.45)	0.29 (0.45)	0.26 (0.44)

firmmt100 (%)	0.49 (0.50)	0.49 (0.50)	0.50 (0.50)	0.54 (0.50)	0.44 (0.50)
firmunk (%)	0.10 (0.29)	0.10 (0.39)	0.09 (0.29)	0.08 (0.27)	0.11 (0.32)
secjob (%)	0.07 (0.25)	0.08 (0.27)	0.05 (0.22)	0.02 (0.14)	0.13 (0.34)
Instrumets for SRHS					
hypertension (%)	0.03 (0.18)	0.04 (0.21)	0.02 (0.12)	0.04 (0.19)	0.03 (0.16)
ill (%)	0.05 (0.21)	0.05 (0.21)	0.05 (0.21)	0.05 (0.22)	0.04 (0.20)
hypyear (in years)	0.16 (1.30)	0.21 (1.48)	0.08 (0.98)	0.20 (1.48)	0.11 (1.03)
illday (in days)	0.22 (2.02)	0.22 (2.05)	0.22 (1.97)	0.25 (2.17)	0.19 (1.83)
No. of observations	11,115	6,514	4,601	6,102	5,013
Instrumets for BMI					
hhobese (%)	0.02 (0.16)	0.02 (0.15)	0.03 (0.17)	0.02 (0.16)	0.03 (0.16)
energy (kilocalories)	23.80 (6.78)	25.52 (6.92)	21.43 (5.82)	23.32 (6.45)	24.40 (7.14)
obese_t4 (×100%)	2.07 (2.22)	2.05 (2.21)	2.10 (2.22)	2.08 (2.38)	2.06 (2.00)
energy_hh (kilocalories)	19.58 (10.43)	18.65 (9.99)	20.86 (10.88)	19.04 (10.42)	20.25 (10.41)
No. of observations	10,016	5,803	4,213	5,572	4,444
Exclusion restriction for employment					
hhnwinc (%)	0.52 (0.50)	0.52 (0.50)	0.51 (0.50)	0.30 (0.46)	0.74 (0.44)
oindnwinc_hh (%)	0.11 (0.31)	0.11 (0.31)	0.11 (0.31)	0.04 (0.20)	0.18 (0.39)
No. of observations	16,086	8,671	7,415	8,115	7,971

Source: Author's calculations using CHNS data.

[1] Dummy variables for the survey year are included in the wage analysis but not reported.

[2] The 8 dummy variables for the occupation: *job_1*, senior professional/technical personnel; *job_2*, junior professional/technical personnel; *job_3*, administrator/executive/manager; *job_4*, office staff; *job_6*, technical, skilled worker; *job_7*, non-technical, non-skilled worker; *job_8*, driver and service worker; *job_9*, engagement in small commercial household business, handicraft and other.

[3] The 2 dummy variables for the job position category: *employee*, worker who works for another person or enterprise, and paid family worker; *otherposi*, unpaid family worker, temporary worker and other.

[4] The 3 dummy variables for the type of work unit: *privatesec*, individual or private enterprise, three source invested enterprise, household business, and family contract; *government*, state enterprise or institute/governmental unit; *typeother*, small collective enterprise, three-capital enterprise and other.

[5] The 4 dummy variables for the size of the work unit: *firm020*, number of employees between 0 and 20; *firm20100*, number of employees between 20 and 100; *firmmt100*, number of employees more than 100; *firmunk*, unknown or no response.

Notes: For full details of the other explanatory variables, see Table 4.5 in Chapter 4.

Table 5.2 The Impact of SRHS on Wages: Pooled OLS, RE and FE - The Total Sample

	The Total Sample		
Variables	**Pooled OLS**	**RE**	**FE**
SRHS Variables			
excellent	0.1112***	0.1263***	0.1785***
	(0.0379)	(0.0368)	(0.0491)
good	0.0870**	0.1042***	0.1706***
	(0.0361)	(0.0354)	(0.0469)
fair	0.0647*	0.0800**	0.1329***
	(0.0369)	(0.0362)	(0.0474)
Non-SRHS Variables			
male	0.1423***	0.1515***	
	(0.0116)	(0.0119)	
age26_35	0.1037***	0.0900***	0.0106
	(0.0192)	(0.0183)	(0.0329)
age36_45	0.1979***	0.1836***	0.0379
	(0.0209)	(0.0199)	(0.0452)
age46_59	0.2035***	0.1946***	0.0129
	(0.0236)	(0.0227)	(0.0594)
primary	-0.0343	-0.0243	-0.0085
	(0.0311)	(0.0296)	(0.0717)
lowermid	-0.0172	-0.0168	-0.1290
	(0.0288)	(0.0277)	(0.0836)
uppermid	0.0295	0.0345	-0.1182
	(0.0303)	(0.0292)	(0.0897)
techvoc	0.0749**	0.0841**	-0.0412
	(0.0330)	(0.0327)	(0.0949)
university	0.2188***	0.2288***	0.0915
	(0.0354)	(0.0344)	(0.0974)
urban	-0.2360***	-0.2226***	
	(0.0130)	(0.0130)	
liaoning	-0.0169	-0.0107	
	(0.0228)	(0.0254)	
heilongjiang	-0.0322	-0.0301	
	(0.0260)	(0.0283)	
jiangsu	0.1580***	0.1494***	

	(0.0224)	(0.0238)	
shandong	0.0351	0.0307	
	(0.0240)	(0.0253)	
henan	0.0721***	0.0657**	
	(0.0255)	(0.0276)	
hubei	-0.0120	-0.0141	
	(0.0238)	(0.0257)	
hunan	0.1446***	0.1605***	
	(0.0229)	(0.0266)	
guangxi	-0.0278	-0.0310	
	(0.0229)	(0.0259)	
married	0.0421**	0.0420**	0.0206
	(0.0193)	(0.0176)	(0.0309)
children1	-0.0110	-0.0100	0.0090
	(0.0128)	(0.0121)	(0.0177)
children2m	-0.0227	-0.0162	0.0713**
	(0.0196)	(0.0189)	(0.0302)
job_2	-0.0369	-0.0390	-0.0300
	(0.0238)	(0.0245)	(0.0358)
job_3	-0.0049	-0.0023	0.0471
	(0.0243)	(0.0245)	(0.0388)
job_4	-0.0339	-0.0262	0.0147
	(0.0236)	(0.0244)	(0.0380)
job_6	-0.0875***	-0.0718***	0.0088
	(0.0241)	(0.0247)	(0.0411)
job_7	-0.1694***	-0.1390***	-0.0132
	(0.0246)	(0.0249)	(0.0416)
job_8	-0.1598***	-0.1348***	-0.0623
	(0.0271)	(0.0262)	(0.0441)
job_9	-0.0009	0.0160	0.0710
	(0.0390)	(0.0346)	(0.0533)
employee	0.1807***	0.1689***	0.1319***
	(0.0224)	(0.0187)	(0.0297)
government	-0.0083	0.0012	0.0160
	(0.0213)	(0.0185)	(0.0306)
typeother	-0.0225	-0.0053	0.0507*
	(0.0214)	(0.0180)	(0.0288)
firm20100	0.0919***	0.0853***	0.0526**
	(0.0184)	(0.0163)	(0.0249)
firmmt100	0.1158***	0.1118***	0.0958***
	(0.0185)	(0.0162)	(0.0263)
firmunk	0.1063***	0.1059***	0.1015***
	(0.0247)	(0.0222)	(0.0337)

secjob	-0.1005***	-0.0721***	0.0186
	(0.0237)	(0.0198)	(0.0301)
1993	0.1323***	0.1353***	0.1579***
	(0.0136)	(0.0145)	(0.0172)
1997	0.7343***	0.7294***	0.7570***
	(0.0146)	(0.0149)	(0.0220)
2000	1.0636***	1.0653***	1.1488***
	(0.0175)	(0.0166)	(0.0274)
2004	1.3160***	1.3255***	1.4723***
	(0.0177)	(0.0181)	(0.0348)
2006	1.4475***	1.4548***	1.6238***
	(0.0183)	(0.0188)	(0.0384)
Observations	11,115	11,115	11,115
R-squared	0.6043		0.5779
Number of id		5,997	5,997
Standard Hausman Test:			
$\chi^2(35)$ (*p*-value)		128.75 (0.00)	

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.
*** p<0.01, ** p<0.05, * p<0.1.

Table 5.3 The Impact of SRHS on Wages[1]: Pooled OLS, RE and FE - Men vs. Women

	Men			Women		
Variables	**Pooled OLS**	**RE**	**FE**	**Pooled OLS**	**RE**	**FE**
SRHS Variables						
excellent	0.1249***	0.1306***	0.1446**	0.1023	0.1288**	0.2350***
	(0.0425)	(0.0483)	(0.0630)	(0.0722)	(0.0569)	(0.0789)
good	0.0979**	0.1079**	0.1405**	0.0822	0.1070*	0.2207***
	(0.0401)	(0.0464)	(0.0602)	(0.0694)	(0.0547)	(0.0752)
fair	0.0696*	0.0773	0.1041*	0.0656	0.0892	0.1792**
	(0.0413)	(0.0474)	(0.0607)	(0.0705)	(0.0560)	(0.0764)
Observations	6,514	6,514	6,514	4,601	4,601	4,601
R-squared	0.5872		0.5898	0.6155		0.5637
Number of id		3,410	3,410		2,587	2,587
Standard Hausman Test:						
$\chi^2(33)$ (p-value)		72.79 (0.00)			85.49 (0.00)	

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The following explanatory variables are also included in the models but not reported: age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.4 The Impact of SRHS on Wages[1]: Pooled OLS, RE and FE - Urban vs. Rural

	Urban			Rural		
Variables	**Pooled OLS**	**RE**	**FE**	**Pooled OLS**	**RE**	**FE**
SRHS Variables						
excellent	0.1167**	0.1327***	0.2093***	0.1117*	0.1158*	0.1388*
	(0.0497)	(0.0456)	(0.0606)	(0.0590)	(0.0599)	(0.0822)
good	0.0994**	0.1193***	0.2097***	0.0805	0.0805	0.1168
	(0.0473)	(0.0437)	(0.0575)	(0.0563)	(0.0578)	(0.0789)
fair	0.0746	0.0936**	0.1652***	0.0589	0.0547	0.0802
	(0.0484)	(0.0448)	(0.0584)	(0.0573)	(0.0589)	(0.0793)
Observations	6,102	6,102	6,102	5,013	5,013	5,013
R-squared	0.6501		0.6014	0.5593		0.5673
Number of id		3,188	3,188		2,809	2,809
Standard Hausman Test:						
$\chi^2(33)$ (*p*-value)		93.70 (0.00)			67.77 (0.00)	

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.5 The Impact of SRHS on Wages[1] - With vs. Without Excluding Agricultural Workers and Self-employed

			FE[1]		
Variables	Total	Men	Women	Urban	Rural
Without the Agricultural and the Self-employed Sample					
excellent	0.1785***	0.1446**	0.2350***	0.2093***	0.1388*
	(0.0491)	(0.0630)	(0.0789)	(0.0606)	(0.0822)
good	0.1706***	0.1405**	0.2207***	0.2097***	0.1168
	(0.0469)	(0.0602)	(0.0752)	(0.0575)	(0.0789)
fair	0.1329***	0.1041*	0.1792**	0.1652***	0.0802
	(0.0474)	(0.0607)	(0.0764)	(0.0584)	(0.0793)
Observations	11,115	6,514	4,601	6,102	5,013
R-squared	0.5779	0.5898	0.5637	0.6014	0.5673
Number of id	5,997	3,410	2,587	3,188	2,809
Standard Hausman Test:					
$\chi^2(35)$ (*p*-value)	128.75 (0.00)	72.79 (0.00)	85.49 (0.00)	93.70 (0.00)	67.77 (0.00)
With the Agricultural and the Self-employed Sample					
excellent	0.1280***	0.0965	0.1779**	0.1557***	0.1113
	(0.0481)	(0.0621)	(0.0755)	(0.0584)	(0.0802)
good	0.1303***	0.0972	0.1836**	0.1554***	0.1082
	(0.0461)	(0.0596)	(0.0723)	(0.0558)	(0.0771)
fair	0.0832*	0.0421	0.1464**	0.1005*	0.0651
	(0.0468)	(0.0604)	(0.0733)	(0.0567)	(0.0781)
Observations	13,712	8,326	5,386	6,966	6,746
R-squared	0.5308	0.5333	0.5364	0.5761	0.5009
Number of id	7,298	4,268	3,030	3,548	3,750
Standard Hausman Test:					
$\chi^2(35)$ (*p*-value)	123.63 (0.00)	72.40 (0.00)	88.39 (0.00)	96.98 (0.00)	87.90 (0.00)

Notes: Standard errors in parentheses. The FE estimation results which are presented in the top part in this table are the same to those in Tables 6.2-6.4.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.6 Sensitivity Analysis Based on Measures of Wage Income

	FE[1]				
Variables	Total	Men	Women	Urban	Rural
Dependent Variable: *lnhrwage*					
excellent	0.1785***	0.1446**	0.2350***	0.2093***	0.1388*
	(0.0491)	(0.0630)	(0.0789)	(0.0606)	(0.0822)
good	0.1706***	0.1405**	0.2207***	0.2097***	0.1168
	(0.0469)	(0.0602)	(0.0752)	(0.0575)	(0.0789)
fair	0.1329***	0.1041*	0.1792**	0.1652***	0.0802
	(0.0474)	(0.0607)	(0.0764)	(0.0584)	(0.0793)
Observations	11,115	6,514	4,601	6,102	5,013
R-squared	0.5779	0.5898	0.5637	0.6014	0.5673
Number of id	5,997	3,410	2,587	3,188	2,809
Standard Hausman Test:					
$\chi^2(35)$ (p-value)	128.75 (0.00)	72.79 (0.00)	85.49 (0.00)	93.70 (0.00)	67.77 (0.00)
Dependent Variable: *lnhrinc*					
excellent	0.1850***	0.1632***	0.2221***	0.1954***	0.1724**
	(0.0492)	(0.0628)	(0.0798)	(0.0610)	(0.0820)
good	0.1948***	0.1677***	0.2406***	0.2137***	0.1669**
	(0.0470)	(0.0599)	(0.0760)	(0.0579)	(0.0787)
fair	0.1505***	0.1256**	0.1920**	0.1669***	0.1199
	(0.0475)	(0.0605)	(0.0772)	(0.0588)	(0.0791)
Observations	11,115	6,514	4,601	6,102	5,013
R-squared	0.5322	0.5490	0.5107	0.5557	0.5219
Number of id	5,997	3,410	2,587	3,188	2,809
Standard Hausman Test:					
$\chi^2(35)$ (p-value)	153.05 (0.00)	85.16 (0.00)	96.63 (0.00)	104.99 (0.00)	73.80 (0.00)

Notes: Standard errors in parentheses. The FE estimation results which are presented in the top part in this table are the same to those in Tables 6.2-6.4.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.7 The Impact of SRHS on Wages: 2SLS

Variables	Total	Men	Women	Urban	Rural
			2SLS[1]		
SRHS in the Wage Equation					
health2	0.0253	0.0232	0.0230	0.0549	-0.0139
	(0.0380)	(0.0461)	(0.0676)	(0.0426)	(0.0743)
R-squared	0.6042	0.5871	0.6154	0.6494	0.5581
Hansen J statistic:					
χ^2(d.f.) (*p*-value)	0.84 (0.36)	0.91 (0.34)	0.11 (0.74)	0.01 (0.91)	1.34 (0.25)
DHW test (*p*-value)	0.00 (0.98)	0.03 (0.87)	0.00 (0.99)	0.49 (0.49)	0.33 (0.57)
Instruments in the Health Equation					
hypertension	-0.2979***	-0.2901***	-0.3476***	-0.3219***	-0.2402***
	(0.0392)	(0.0441)	(0.0906)	(0.0496)	(0.0642)
ill	-0.5016***	-0.5320***	-0.4535***	-0.5515***	-0.4347***
	(0.0346)	(0.0451)	(0.0539)	(0.0462)	(0.0511)
R-squared	0.1061	0.1164	0.0996	0.1229	0.079
F statistic (*p*-value)	141.38 (0.00)	94.86 (0.00)	46.54 (0.00)	97.97 (0.00)	45.62 (0.00)
Number of id	5,997	3,410	2,587	3,188	2,809
Observations	11,115	6,514	4,601	6,102	5,013
			Pooled OLS[2]		
SRHS Variables					
health2	0.0265***	0.0310***	0.0219*	0.0258**	0.0285**
	(0.0078)	(0.0099)	(0.0126)	(0.0101)	(0.0120)
R-squared	0.6042	0.5872	0.6154	0.6500	0.5593
Observations	11,115	6,514	4,601	6,102	5,013
			FE[3]		
SRHS Variables					
health2	0.0328***	0.0278**	0.0417***	0.0367***	0.0333**
	(0.0101)	(0.0131)	(0.0158)	(0.0129)	(0.0160)
R-squared	0.5773	0.5894	0.5627	0.6003	0.5672
Standard Hausman Test: χ^2(d.f.) (*p*-value)	122.31 (0.00)	70.56 (0.00)	81.19 (0.00)	88.77 (0.00)	66.39 (0.00)
Number of id	5,997	3,410	2,587	3,188	2,809
Observations	11,115	6,514	4,601	6,102	5,013

Notes: Standard errors in parentheses. Under 2SLS and pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.
[1-3] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year. *** p<0.01, ** p<0.05, * p<0.1.

Table 5.8 The Impact of SRHS on Wages: Heckman Selection Model

Variables	Total	Men	Women	Urban	Rural
			Heckman[1]		
SRHS in the Wage Equation					
excellent	0.0638*	0.0604	0.0823	0.0893*	0.0256
	(0.0381)	(0.0432)	(0.0720)	(0.0498)	(0.0605)
good	0.0389	0.0325	0.0619	0.0705	-0.0029
	(0.0364)	(0.0410)	(0.0693)	(0.0475)	(0.0581)
fair	0.0212	0.0117	0.0464	0.0490	-0.0198
	(0.0372)	(0.0420)	(0.0704)	(0.0485)	(0.0586)
ρ	-0.2682	-0.3079	-0.1603	-0.2141	-0.3577
	(0.0275)	(0.0352)	(0.0442)	(0.0283)	(0.0651)
Wald test $\rho = 0$ test (p-value)	85.99 (0.00)	66.91 (0.00)	12.70 (0.00)	52.01 (0.00)	25.12 (0.00)
Observations	11,115	6,514	4,601	6,102	5,013
SRHS and Exclusion Restriction in the Employment Equation					
excellent	0.6252***	0.8149***	0.3924***	0.5147***	0.7417***
	(0.0781)	(0.1027)	(0.1177)	(0.1092)	(0.1182)
good	0.6364***	0.8358***	0.4022***	0.5807***	0.6786***
	(0.0737)	(0.0959)	(0.1118)	(0.1037)	(0.1114)
fair	0.5651***	0.7029***	0.3762***	0.5164***	0.6142***
	(0.0754)	(0.0978)	(0.1147)	(0.1065)	(0.1125)
hhnwinc	-0.3872***	-0.2945***	-0.5204***	-0.4411***	-0.3168***
	(0.0293)	(0.0411)	(0.0424)	(0.0406)	(0.0434)
Observations	16,086	8,671	7,415	8,115	7,971
			Pooled OLS[2]		
SRHS Variables					
excellent	0.1112***	0.1249***	0.1023	0.1167**	0.1117*
	(0.0379)	(0.0425)	(0.0722)	(0.0497)	(0.0590)
good	0.0870**	0.0979**	0.0822	0.0994**	0.0805
	(0.0361)	(0.0401)	(0.0694)	(0.0473)	(0.0563)
fair	0.0647*	0.0696*	0.0656	0.0746	0.0589
	(0.0369)	(0.0413)	(0.0705)	(0.0484)	(0.0573)
Observations	11,115	6,514	4,601	6,102	5,013

R-squared	0.6043	0.5872	0.6155	0.6501	0.5593
			FE[3]		
SRHS Variables					
excellent	0.1785***	0.1446**	0.2350***	0.2093***	0.1388*
	(0.0491)	(0.0630)	(0.0789)	(0.0606)	(0.0822)
good	0.1706***	0.1405**	0.2207***	0.2097***	0.1168
	(0.0469)	(0.0602)	(0.0752)	(0.0575)	(0.0789)
fair	0.1329***	0.1041*	0.1792**	0.1652***	0.0802
	(0.0474)	(0.0607)	(0.0764)	(0.0584)	(0.0793)
Observations	11,115	6,514	4,601	6,102	5,013
R-squared	0.5779	0.5898	0.5637	0.6014	0.5673
Number of id	5,997	3,410	2,587	3,188	2,809
Standard Hausman Test:					
χ^2(d.f.) (*p*-value)	128.75 (0.00)	72.79 (0.00)	85.49 (0.00)	93.70 (0.00)	67.77 (0.00)

Notes: Standard errors in parentheses. Under Heckman and pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1-3] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.9 The Impact of BMI on Wages[1]: Pooled OLS, RE and FE - The Total Sample

		The Total Sample	
Variables	**Pooled OLS**	**RE**	**FE**
WHO BMI Classification			
uweight_who	0.0066	0.0003	-0.0111
	(0.0222)	(0.0217)	(0.0372)
oweight_who	0.0164	0.0167	-0.0215
	(0.0140)	(0.0138)	(0.0245)
obese_who	-0.0079	-0.0032	-0.0477
	(0.0382)	(0.0384)	(0.0678)
SRHS Variables			
excellent	0.1014**	0.1189***	0.1799***
	(0.0404)	(0.0384)	(0.0523)
good	0.0892**	0.1068***	0.1737***
	(0.0384)	(0.0369)	(0.0498)
fair	0.0730*	0.0899**	0.1465***
	(0.0393)	(0.0376)	(0.0503)
Observations	10,016	10,016	10,016
R-squared	0.6177		0.5843
Number of id		5,484	5,484
Standard Hausman Test:			
χ^2(d.f.) (*p*-value)		105.50 (0.00)	

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.10 The Impact of BMI on Wages[1]: Pooled OLS, RE and FE - Men vs. Women

	Men			Women		
Variables	**Pooled OLS**	**RE**	**FE**	**Pooled OLS**	**RE**	**FE**
WHO BMI Classification						
uweight_who	0.0041	0.0133	0.0593	0.0055	-0.0116	-0.0680
	(0.0315)	(0.0316)	(0.0544)	(0.0308)	(0.0295)	(0.0512)
oweight_who	0.0424**	0.0371**	-0.0131	-0.0090	-0.0052	-0.0326
	(0.0180)	(0.0178)	(0.0309)	(0.0231)	(0.0223)	(0.0411)
obese_who	0.0266	0.0333	0.0237	-0.0446	-0.0514	-0.2240*
	(0.0479)	(0.0475)	(0.0812)	(0.0572)	(0.0665)	(0.1275)
SRHS Variables						
excellent	0.1163***	0.1240**	0.1333**	0.0863	0.1153*	0.2530***
	(0.0448)	(0.0504)	(0.0672)	(0.0773)	(0.0591)	(0.0842)
good	0.0959**	0.1053**	0.1241*	0.0873	0.1133**	0.2556***
	(0.0420)	(0.0484)	(0.0639)	(0.0745)	(0.0568)	(0.0802)
fair	0.0704	0.0790	0.1011	0.0783	0.1047*	0.2217***
	(0.0434)	(0.0494)	(0.0644)	(0.0756)	(0.0581)	(0.0815)
Observations	5,803	5,803	5,803	4,213	4,213	4,213
R-squared	0.6009		0.6011	0.6287		0.5652
Number of id		3,117	3,117		2,367	2,367
Standard Hausman Test:						
χ^2(d.f.) (*p*-value)		70.47 (0.00)			74.25 (0.00)	

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The following explanatory variables are also included in the models but not reported: age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.11 The Impact of BMI on Wages[1]: Pooled OLS, RE and FE - Urban vs. Rural

	Urban			Rural		
Variables	**Pooled OLS**	**RE**	**FE**	**Pooled OLS**	**RE**	**FE**
WHO BMI Classification						
uweight_who	-0.0210	-0.0128	0.0346	0.0580	0.0269	-0.0765
	(0.0252)	(0.0265)	(0.0451)	(0.0382)	(0.0354)	(0.0635)
oweight_who	0.0330*	0.0399**	0.0258	0.0051	-0.0041	-0.0685*
	(0.0179)	(0.0174)	(0.0323)	(0.0218)	(0.0217)	(0.0374)
obese_who	-0.0443	-0.0369	-0.0425	0.0598	0.0555	-0.0372
	(0.0486)	(0.0480)	(0.0973)	(0.0581)	(0.0606)	(0.0958)
SRHS Variables						
excellent	0.1157**	0.1328***	0.2184***	0.0992	0.1099*	0.1265
	(0.0533)	(0.0472)	(0.0633)	(0.0630)	(0.0629)	(0.0901)
good	0.1114**	0.1316***	0.2209***	0.0742	0.0788	0.1025
	(0.0509)	(0.0453)	(0.0600)	(0.0601)	(0.0606)	(0.0861)
fair	0.0931*	0.1135**	0.1899***	0.0608	0.0612	0.0769
	(0.0520)	(0.0463)	(0.0610)	(0.0610)	(0.0616)	(0.0862)
Observations	5,572	5,572	5,572	4,444	4,444	4,444
R-squared	0.6626		0.6095	0.5731		0.5755
Number of id		2,998	2,998		2,486	2,486
Standard Hausman Test:						
χ^2(d.f.) (*p*-value)		79.63 (0.00)			70.91 (0.00)	

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.
[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.
*** p<0.01, ** p<0.05, * p<0.1.

Table 5.12 The Impact of BMI on Wages - Sensitivity Analysis (The Total Sample)

Variables	Model (1)	Pooled OLS Model (2)	Model (3)
WHO BMI Classification			
uweight_who	-0.1295***	0.0041	0.0066
	(0.0390)	(0.0222)	(0.0222)
oweight_who	0.2401***	0.0172	0.0164
	(0.0222)	(0.0140)	(0.0140)
obese_who	0.3786***	-0.0069	-0.0079
	(0.0648)	(0.0384)	(0.0382)
SRHS Variables			
excellent			0.1014**
			(0.0404)
good			0.0892**
			(0.0384)
fair			0.0730*
			(0.0393)
Other Explanatory Variables[1] Controlled: (Yes/No)	No	Yes	Yes
Observations	10,016	10,016	10,016
R-squared	0.0206	0.6174	0.6177

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The other explanatory variables are: gender, age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Table 5.13 The Impact of BMI on Wages: 2SLS

Variables	Total	Men	Women	Urban	Rural
			2SLS[1]		
BMI in the Wage Equation					
bmi	-0.0326	-0.0270	-0.0418	-0.0439*	0.0030
	(0.0199)	(0.0238)	(0.0372)	(0.0234)	(0.0322)
R-squared	0.6033	0.5870	0.6107	0.6347	0.5727
Hansen J statistic:					
χ^2(d.f.) (*p*-value)	0.02 (0.89)	1.86 (0.17)	2.66 (0.11)	0.01 (0.92)	0.01 (0.96)
DHW test (*p*-value)	0.03 (0.83)	2.07 (0.15)	1.09 (0.30)	0.04 (0.73)	0.01 (0.93)
Instruments in the Health Equation					
hhobese	0.9755***	1.0336***	0.9667***	0.7833***	1.0671***
	(0.2614)	(0.3365)	(0.3789)	(0.4104)	(0.3008)
energy	0.0341***	0.0363***	0.0252***	0.0429***	0.0274***
	(0.0049)	(0.0061)	(0.0077)	(0.0070)	(0.0067)
R-squared	0.1699	0.1934	0.1967	0.1755	0.1855
F statistic (*p*-value)	31.97 (0.00)	23.07 (0.00)	8.66 (0.00)	20.80 (0.00)	15.15 (0.00)
Number of id	5,484	3,117	2,367	2,998	2,486
Observations	10,016	5,803	4,213	5,572	4,444
			Pooled OLS[2]		
Health Variables					
bmi	0.0014	0.0063**	-0.0026	0.0026	0.0003
	(0.0019)	(0.0026)	(0.0030)	(0.0024)	(0.0031)
R-squared	0.6177	0.6010	0.6288	0.6623	0.5728
Observations	10,016	5,803	4,213	5,572	4,444
			FE[3]		
Health Variables					
bmi	-0.0067	-0.0087	-0.0032	-0.0065	-0.0035
	(0.0049)	(0.0065)	(0.0077)	(0.0063)	(0.0078)
R-squared	0.5844	0.6011	0.5640	0.6093	0.5744
Standard Hausman Test: χ^2(d.f.) (*p*-value)	104.92 (0.00)	69.87 (0.00)	69.86 (0.00)	80.72 (0.00)	62.68 (0.00)
Number of id	5,484	3,117	2,367	2,998	2,486
Observations	10,016	5,803	4,213	5,572	4,444

Notes: Standard errors in parentheses. Under 2SLS and pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1-3] The following explanatory variables are also included in the models but not reported: SRHS, gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year. *** p<0.01, ** p<0.05, * p<0.1.

Table 5.14 The Impact of BMI on Wages: Heckman Selection Model

Variables	Total	Men	Women	Urban	Rural
			Heckman[1]		
BMI in the Wage Equation					
uweight_who	0.0107	0.0172	0.0037	-0.0185	0.0618
	(0.0221)	(0.0314)	(0.0306)	(0.0250)	(0.0383)
oweight_who	0.0190	0.0401**	-0.0060	0.0346*	0.0073
	(0.0140)	(0.0181)	(0.0230)	(0.0179)	(0.0219)
obese_who	-0.0028	0.0246	-0.0376	-0.0408	0.0640
	(0.0386)	(0.0483)	(0.0577)	(0.0489)	(0.0589)
ρ	-0.2489	-0.2854	-0.1553	-0.2059	-0.3134
	(0.0276)	(0.0376)	(0.0433)	(0.0295)	(0.0597)
Wald test $\rho = 0$ test (p-value)	74.49 (0.00)	51.36 (0.00)	12.44 (0.00)	46.11 (0.00)	24.04 (0.00)
Observations	10,016	5,803	4,213	5,572	4,444
BMI and Exclusion Restriction in the Employment Equation					
uweight_who	-0.0682	-0.2382***	0.1043	-0.0599	-0.0445
	(0.0598)	(0.0847)	(0.0802)	(0.0836)	(0.0872)
oweight_who	-0.0697*	0.0229	-0.1066*	-0.0535	-0.0632
	(0.0364)	(0.0493)	(0.0549)	(0.0526)	(0.0512)
obese_who	-0.1685*	-0.0262	-0.2665*	-0.1249	-0.1822
	(0.0960)	(0.1402)	(0.1409)	(0.1473)	(0.1231)
hhnwinc	-0.4158***	-0.2957***	-0.5674***	-0.4455***	-0.3775***
	(0.0309)	(0.0434)	(0.0448)	(0.0428)	(0.0459)
Observations	14,391	7,674	6,717	7,373	7,018/
			Pooled OLS[2]		
WHO BMI Classification					
uweight_who	0.0066	0.0041	0.0055	-0.0210	0.0580
	(0.0222)	(0.0315)	(0.0308)	(0.0252)	(0.0382)
oweight_who	0.0164	0.0424**	-0.0090	0.0330*	0.0051
	(0.0140)	(0.0180)	(0.0231)	(0.0179)	(0.0218)
obese_who	-0.0079	0.0266	-0.0446	-0.0443	0.0598
	(0.0382)	(0.0479)	(0.0572)	(0.0486)	(0.0581)
Observations	10,016	5,803	4,213	5,572	4,444

R-squared	0.6177	0.6009	0.6287	0.6626	0.5731
			FE[3]		
WHO BMI Classification					
uweight_who	-0.0111	0.0593	-0.0680	0.0346	-0.0765
	(0.0372)	(0.0544)	(0.0512)	(0.0451)	(0.0635)
oweight_who	-0.0215	-0.0131	-0.0326	0.0258	-0.0685*
	(0.0245)	(0.0309)	(0.0411)	(0.0323)	(0.0374)
obese_who	-0.0477	0.0237	-0.2240*	-0.0425	-0.0372
	(0.0678)	(0.0812)	(0.1275)	(0.0973)	(0.0958)
Observations	10,016	5,803	4,213	5,572	4,444
R-squared	0.5843	0.6011	0.5652	0.6095	0.5755
Number of id	5,484	3,117	2,367	2,998	2,486
Standard Hausman Test:					
χ^2(d.f.) (*p*-value)	105.50 (0.00)	70.47 (0.00)	74.25 (0.00)	79.63 (0.00)	70.91 (0.00)

Notes: Standard errors in parentheses. Under pooled OLS and Heckman analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1-3] The following explanatory variables are also included in the models but not reported: SRHS, gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Appendix 5.1 The Average Age and Educational Attainment by Self-Reported Health Status

SRHS	Age (in Years) Mean (SD)	Educational Attainment[1] Mean (SD)
excellent	34.27 (9.74)	3.70 (1.21)
good	35.90 (10.11)	3.54 (1.23)
fair	39.69 (10.00)	3.58 (1.35)
poor	42.46 (9.11)	3.29 (1.14)

Source: Author's calculations using CHNS data.
[1] Education attainment: range 1 = no educational attainment - 6 = university or higher educational attainment.

Appendix 5.2 The Average Age and Edncational Attainment by BMI Categories

BMI Categories	Age (in Years) Mean (SD)	Educational Attainment[1] Mean (SD)
uweight_who	31.33 (10.13)	3.67 (1.23)
nweight_who	36.24 (9.90)	3.59 (1.26)
oweight_who	40.91 (8.97)	3.64 (1.28)
obese_who	41.01 (8.89)	3.70 (1.41)

Source: Author's calculations using CHNS data.
[1] See Note 1 in Appendix 6.1.

Appendix 5.3 The Impact of SRHS on Wages: 2SLS (Sensitivity Analysis Based on Alternative Instruments)

	2SLS[1]				
Variables	**Total**	**Men**	**Women**	**Urban**	**Rural**
SRHS in the Wage Equation					
health2	-0.0247	-0.0271	-0.0034	-0.0279	-0.0610
	(0.0705)	(0.0810)	(0.1336)	(0.0806)	(0.1476)
R-squared	0.6025	0.5849	0.6150	0.6481	0.5541
Hansen J statistic:					
χ^2(d.f.) (*p*-value)	1.18 (0.28)	2.59 (0.11)	0.01 (0.95)	0.07 (0.79)	2.60 (0.11)
DHW test (*p*-value)	0.55 (0.46)	0.55 (0.46)	0.04 (0.85)	0.47 (0.49)	0.39 (0.53)
Instruments in the Health Equation					
hypyear	-0.0295***	-0.0271***	-0.0376***	-0.0306***	-0.0252**
	(0.0066)	(0.0074)	(0.0148)	(0.0075)	(0.0137)
illday	-0.0329***	-0.0325***	-0.0335***	-0.0369***	-0.0261***
	(0.0041)	(0.0053)	(0.0066)	(0.0075)	(0.0065)
R-squared	0.0873	0.0934	0.0869	0.1011	0.0830
F statistic (*p*-value)	42.46 (0.00)	25.97 (0.00)	16.57 (0.00)	33.26 (0.00)	9.72 (0.00)
Number of id	5,997	3,410	2,587	3,188	2,809
Observations	11,115	6,514	4,601	6,102	5,013

Notes: Standard errors in parentheses. Under 2SLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix 5.4 The Impact of SRHS on Wages: Heckman (Sensitivity Analysis Based on Alternative Exclusion Restriction)

			Heckman[1]		
Variables	Total	Men	Women	Urban	Rural
SRHS in the Wage Equation					
excellent	0.0614	0.0606	0.0789	0.0877*	0.0213
	(0.0382)	(0.0432)	(0.0720)	(0.0498)	(0.0607)
good	0.0363	0.0327	0.0581	0.0685	-0.0068
	(0.0364)	(0.0411)	(0.0693)	(0.0475)	(0.0583)
fair	0.0187	0.0115	0.0430	0.0469	-0.0235
	(0.0373)	(0.0421)	(0.0704)	(0.0486)	(0.0588)
ρ	-0.2834	-0.3009	-0.2049	-0.2200	-0.3815
	(0.0268)	(0.0357)	(0.0402)	(0.0266)	(0.0644)
Wald test $\rho = 0$ test (p-value)	99.70 (0.00)	62.44 (0.00)	24.54 (0.00)	64.05 (0.00)	28.45 (0.00)
Observations	11,115	6,514	4,601	6,102	5,013
SRHS and Exclusion Restriction in the Employment Equation					
excellent	0.6508***	0.8387***	0.4202***	0.5503***	0.7546***
	(0.0768)	(0.1019)	(0.1145)	(0.1081)	(0.1160)
good	0.6686***	0.8597***	0.4439***	0.6197***	0.6987***
	(0.0724)	(0.0951)	(0.1085)	(0.1024)	(0.1092)
fair	0.5936***	0.7253***	0.4078***	0.5594***	0.6273***
	(0.0740)	(0.0969)	(0.1112)	(0.1051)	(0.1104)
oindnwinc_hh	-0.1912***	-0.2477***	-0.1952***	-0.3741***	-0.1334***
	(0.0368)	(0.0526)	(0.0538)	(0.0751)	(0.0424)
Observations	16,086	8,671	7,415	8,115	7,971

Notes: Standard errors in parentheses. Under Heckman analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The following explanatory variables are also included in the models but not reported: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year. *** p<0.01, ** p<0.05, * p<0.1.

Appendix 5.5 The Impact of BMI on Wages - Sensitivity Analysis (Men vs. Women)

	Men			Women		
Variables	Model (1)	Model (2)	Model (3)	Model (1)	Model (2)	Model (3)
WHO BMI Classification						
uweight_who	-0.1621***	0.0002	0.0041	-0.0706	0.0040	0.0055
	(0.0517)	(0.0315)	(0.0315)	(0.0580)	(0.0308)	(0.0308)
oweight_who	0.3078***	0.0441**	0.0424**	0.0591*	-0.0092	-0.0090
	(0.0267)	(0.0181)	(0.0180)	(0.0350)	(0.0231)	(0.0231)
obese_who	0.4556***	0.0284	0.0266	0.1388	-0.0441	-0.0446
	(0.0697)	(0.0482)	(0.0479)	(0.1163)	(0.0575)	(0.0572)
SRHS Variables						
excellent			0.1163***			0.0863
			(0.0448)			(0.0773)
good			0.0959**			0.0873
			(0.0420)			(0.0745)
fair			0.0704			0.0783
			(0.0434)			(0.0756)
Other Explanatory Variables[1] Controlled: (Yes/No)	No	Yes	Yes	No	Yes	Yes
Observations	5,803	5,803	5,803	4,213	4,213	4,213
R-squared	0.0357	0.6004	0.6009	0.0019	0.6285	0.6287

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The other explanatory variables are: age group, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Appendix 5.6 The Impact of BMI on Wages - Sensitivity Analysis (Urban vs. Rural)

		Urban			Rural	
Variables	**Model (1)**	**Model (2)**	**Model (3)**	**Model (1)**	**Model (2)**	**Model (3)**
WHO BMI Classification						
uweight_who	-0.1996***	-0.0248	-0.0210	-0.0032	0.0570	0.0580
	(0.0459)	(0.0251)	(0.0252)	(0.0614)	(0.0382)	(0.0382)
oweight_who	0.2483***	0.0332*	0.0330*	0.2411***	0.0063	0.0051
	(0.0293)	(0.0179)	(0.0179)	(0.0332)	(0.0219)	(0.0218)
obese_who	0.3264***	-0.0434	-0.0443	0.4711***	0.0610	0.0598
	(0.0927)	(0.0488)	(0.0486)	(0.0758)	(0.0584)	(0.0581)
SRHS Variables						
excellent			0.1157**			0.0992
			(0.0533)			(0.0630)
good			0.1114**			0.0742
			(0.0509)			(0.0601)
fair			0.0931*			0.0608
			(0.0520)			(0.0610)
Other Explanatory Variables[1] Controlled: (Yes/No)	No	Yes	Yes	No	Yes	Yes
Observations	5,572	5,572	5,572	4,444	4,444	4,444
R-squared	0.0244	0.6622	0.6626	0.0199	0.5728	0.5731

Notes: Standard errors in parentheses. Under pooled OLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.
[1] The other explanatory variables are: gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix 5.7 The Impact of BMI on Wages: 2SLS (Sensitivity Analysis Based on Alternative Instruments)

	2SLS[1]				
Variables	Total	Men	Women	Urban	Rural
SRHS in the Wage Equation					
bmi	-0.0021	-0.0109	0.0061	-0.0047	0.0080
	(0.0167)	(0.0271)	(0.0206)	(0.0184)	(0.0408)
R-squared	0.6175	0.5972	0.6279	0.6616	0.4962
Hansen J statistic:					
χ^2(d.f.) (*p*-value)	0.90 (0.34)	0.03 (0.81)	0.37 (0.54)	0.04 (0.87)	0.03 (0.86)
DHW test (*p*-value)	0.05 (0.83)	0.40 (0.53)	0.18 (0.67)	0.16 (0.69)	0.51 (0.32)
Instruments in the Health Equation					
obese_t4	0.1743***	0.1374***	0.2105***	0.1970***	0.1447***
	(0.0237)	(0.0306)	(0.0357)	(0.0299)	(0.0422)
energy_hh	0.0066**	0.0094**	0.0026*	0.0044*	0.0087**
	(0.0031)	(0.0041)	(0.0045)	(0.0042)	(0.0044)
R-squared	0.1746	0.1928	0.2091	0.1838	0.1846
F statistic (*p*-value)	29.77 (0.00)	12.56 (0.00)	17.95 (0.00)	22.59 (0.00)	7.89 (0.00)
Number of id	5,997	3,410	2,587	3,188	2,809
Observations	10,016	5,803	4,213	5,572	4,444

Notes: Standard errors in parentheses. Under 2SLS analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.

[1] The following explanatory variables are also included in the models but not reported: SRHS, gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

*** p<0.01, ** p<0.05, * p<0.1.

Appendix 5.8 The Impact of BMI on Wages: Heckman Selection Model (Sensitivity Analysis Based on Alternative Exclusion Restriction)

			Heckman[1]		
Variables	Total	Men	Women	Urban	Rural
BMI in the Wage Equation					
uweight_who	0.0115	0.0181	0.0038	-0.0179	0.0630
	(0.0221)	(0.0314)	(0.0307)	(0.0251)	(0.0384)
oweight_who	0.0192	0.0402**	-0.0052	0.0348*	0.0072
	(0.0140)	(0.0181)	(0.0230)	(0.0179)	(0.0219)
obese_who	-0.0012	0.0253	-0.0339	-0.0401	0.0669
	(0.0386)	(0.0483)	(0.0580)	(0.0489)	(0.0590)
ρ	-0.2616	-0.2798	-0.1877	-0.2143	-0.3380
	(0.0269)	(0.0380)	(0.0391)	(0.0276)	(0.0599)
Wald test $\rho = 0$ test (p-value)	86.28 (0.00)	48.49 (0.00)	21.95 (0.00)	56.78 (0.00)	27.10 (0.00)
Observations	10,016	5,803	4,213	5,572	4,444
BMI and Exclusion Restriction in the Employment Equation					
uweight_who	-0.0811	-0.2469***	0.0855	-0.0676	-0.0615
	(0.0598)	(0.0853)	(0.0801)	(0.0832)	(0.0876)
oweight_who	-0.0664*	0.0272	-0.1099**	-0.0628	-0.0500
	(0.0362)	(0.0492)	(0.0546)	(0.0524)	(0.0510)
obese_who	-0.1431	-0.0228	-0.2094	-0.1117	-0.1524
	(0.0976)	(0.1412)	(0.1443)	(0.1499)	(0.1259)
oindnwinc_hh	-0.2298***	-0.3029***	-0.2201***	-0.3970***	-0.1888***
	(0.0397)	(0.0567)	(0.0579)	(0.0811)	(0.0459)
Observations	14,391	7,674	6,717	7,373	7,018

Notes: Standard errors in parentheses. Under Heckman analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual.
[1] The following explanatory variables are also included in the models but not reported: SRHS, gender, age group, educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.
*** p<0.01, ** p<0.05, * p<0.1.

Chapter 6 Health Related Behaviour and Labour Market Outcomes in China

6.1 Introduction

Drinking alcohol has been traditionally accepted in China during major social events, such as the Chinese New Year, wedding ceremonies and birthday parties. Furthermore, the rapid growth in the Chinese economy has been accompanied by noticeable changes in the alcohol consumption of the Chinese population (Hao *et al.*, 2005). Alcohol is commonly used, particularly by men, to facilitate social interaction, maintain good relations between supervisors and employees and to promote friendship among colleagues, since these often eat out together after work or hold business meetings over dinner (Hao *et al.*, 1995a; Hao *et al.*, 1999; and Cochrane *et al.*, 2003).

Figure 6.1 presents the adult per capita alcohol consumption across countries between 1991 and 2006. Although *per capita* alcohol consumption in China was still low compared with developed countries, it is growing faster than other countries[138], which is accompanied by the fast economic development and increase in average income level (Tang *et al.*, 2013).

In the Chinese culture, women are discouraged from drinking alcohol and from smoking. As shown in Appendix 6.1, both the percentage of lifetime abstainers and of abstainers (who did not drink last year) in China are significantly higher than in other developed countries. Hao *et al.* (1995b) examine reasons for drinking alcohol among the Chinese population. They find that women were much more likely to report being sensitive to disapproval from family and peers, whereas men were more concerned with health and finances. As expected in many countries, reasons for drinking are: alcohol's positive social effects, relief of tension and worry, and relief of craving and withdrawal symptoms (Cochrane *et al.*, 2003). However, as in many other countries, excessive drinking in China has shown an association with both health-related problems such as alcohol related liver disease (Lu *et al.*, 2004), and social harm such as traffic accidents and work-related injuries (Zhang *et al.*, 2007; and Cherpitel *et al.*, 2012).

[138] Note that in Figure 6.1, the alcohol consumption data reported from 2000 onwards have been calculated using a different source. Therefore, the decline between 1999 and 2000 may not necessarily reflect a real decline in alcohol consumption. In addition, both the early period (1991-1999) and the late period (2000-2006) show an upward trend.

As argued by Cochrane *et al.* (2003) and Tang *et al.* (2013), China does not have comprehensive public health policy to deal with these problems. According to Appendix 6.2, many aspects of alcohol policy are weaker in China compared to selected countries. For example, China has no enforceable legal drinking age and does not regulate when or where alcoholic products are sold. The substantial economic and political reforms currently taking place in China which are aimed at achieve balance between economic development and public health and between short-term and long-term alcohol control strategies (Tang *et al.*, 2013). A better understanding of the relationship between drinking behaviour and labour market outcomes for the Chinese population would make a contribution to achieve these balances.

Similar to drinking alcohol, smoking is much more accepted for Chinese men than for women. According to Appendix 6.3, there is a very high prevalence of smoking among men (51%) and a very low prevalence among women (2%) in China, which is similar to that found by Levy (2006). Moreover, China has the highest male prevalence of smoking and the lowest female prevalence of smoking among the seven countries. However, recent studies show that there is a declining trend of prevalence of smoking for both genders. For example, using data from National Health Service Survey 1993, 1998 and 2003, Qian *et al.* (2010) find that the prevalence of smoking decrease from 60% to 49% for men and from 5% to 3% for women between 1993 and 2003. Pan and Hu (2008) using data from the CHNS also find a trend of decreasing prevalence of smoking between 1991 and 2004 for Chinese adults.

Compared to many other countries, smoking is a social custom in China and it is unique to Chinese culture. Giving cigarettes at any social interaction is a sign of respect and hospitality (Leo, 2006). According to Pan (2004), there is a special group of so-called 'social smokers', who smoke only when they are with friends in social settings where peer pressure is very high and smoking is important to reinforce relationships. Moreover, cigarette offering, along with alcohol offering are two typical social necessities and are highly important social function in building or maintain connections in China.

According to the Chinese Association on Tobacco Control (CATC)[139], China comprises 20% of the world's population but currently produces 42% of the world's tobacco products, and consumes 31% of the world's cigarettes (or 1.8 trillion cigarettes per year) Facing this increasing public health problem and pressure from the WHO, on October 11, 2005, China became the 78th country in the world to ratify the WHO Framework Convention on Tobacco Control (FCTC), an international treaty intended to reduce tobacco-related disease and death. Under the conditions of the FCTC, China is required to completely ban promotion and sponsorship on radio, television, print media and the internet within five years, as well as to prohibit tobacco companies from sponsoring international events or activities. China has also resolved to ban all tobacco vending machines, as well as smoking in indoor work places, public areas, and public transportation vehicles (Miao and Wang, 2010).

Based on the same survey, the CHNS, which is used in the previous two chapters, this chapter examines the effect of drinking and smoking behaviour on labour market outcomes (measured as employment and real hourly wages) for the Chinese working-age adults and it has several key advantages. Firstly, this chapter serves to provide more empirical evidence on the relationship between health-related behaviour and labour market outcomes for the Chinese population and adds to the international literature in this area. Most of the studies relating to the impact of drinking and/or smoking on labour market outcomes focuses on population in the US and in many other western countries such as the UK, Germany, Canada, and Australia. To the author's knowledge, there is no evidence for China except for one study, Lin and Chen (2010). However, instead of using a more representative sample, their analysis focuses on a special administrative region in China, Taiwan.

Secondly, this chapter contributes to the literature in estimating the simultaneous effect of drinking and smoking on employment and wages. Previous studies usually concentrate on the effect of drinking on employment (Mullahy and Sindelar, 1996) or wages (French and Zarkin, 1995; Hamilton and Hamilton, 1997; Zarkin *et al.*, 1998; MacDonald and Shields, 2001; Tekin, 2002; and Barrett, 2002) without controlling

[139] The Chinese Association on Tobacco Control (CATC) was established in 1990, composed of members of the voluntary sector who are willing to work on tobacco control activities, including academic, social and mass organizations. It is supervised and administrated by the Ministry of Health and Ministry of Civil Affairs in carrying out its professional activities. See: http://www.catcprc.org.cn/index.aspx?language=en. [cited 24th May 2013].

for smoking status. There are relatively few studies that examine the impact of smoking on wages such as Levine *et al.* (1995), Heineck and Schwarze (2003), and Lin and Chen (2010), where Lin and Chen (2010) is the only study that examines the effect of smoking on both employment and wages. However, drinking and smoking are highly correlated, so that failing to control for one when estimating the effect of the other might lead to biased results (Auld, 2005). Studies that investigate the simultaneous effects of drinking and smoking on wages include Van Ours (2002), Lye and Hirschberg (2001), and Auld (2005). To the author's knowledge, there is no evidence relating to simultaneous effects of drinking and smoking on employment.

Thirdly, the panel nature of the data is used. The RE models yield estimates of all coefficients even those of the time-invariant explanatory variables and it is more efficient than cross-sectional models. The FE models are advantageous to cross-sectional analyses as unobservable individual factors can be controlled for. In other words, if unobservable individual factors exist that are correlated both with health-related behaviours and labour market outcomes, the use of cross-sectional data leads to biased estimates.

A number of studies have employed an IV method to address the endogeneity of drinking and/or smoking, such as Mullahy and Sindelar (1996), Hamilton and Hamilton (1997), Macdonald and Shields (2001), Barrett (2002), Lye and Hirschberg (2004), Auld (2005), and Lin and Chen (2010). However, these studies do not typically perform either a test of endogeneity or a test of instrument validity. Other studies fail to employ the IV method usually due to lacks of valid instruments such as French and Zarkin (1995), Zarkin *et al.* (1998), and Tekin (2002). In accounting for the potential endogeneity of drinking and smoking, this chapter adopts a Newey's (1978) two-step and a 2SLS method to estimate the effect on employment and wages, respectively. The instruments used in this chapter include variables measuring the share of the respondents' working-age household members with one out of five types of drinking frequency, and two variables measuring the average amount of alcohol and cigarettes consumption of the same household members, respectively. In addition, both a test of endogeneity and instruments validity are performed based on the two estimators.

The fifth advantage of this chapter is that the analysis is carried out for sub samples within the Chinese population, and it is therefore be able to explore the differences across genders and age group. One of the focuses of this chapter is on the gender differences in the effect of drinking and smoking status on labour market outcomes. The existing literature mainly focuses on gender differences in estimating effect of drinking and/or smoking behaviour on the labour market outcomes, although the results are somewhat mixed. As discussed earlier, Chinese women are discouraged either from drinking or smoking. However, both of the two behaviours serve high import social function in building or maintain connections, thereby facilitating promotions and raises. Another focus in this chapter is to compare the effect of drinking and smoking on labour market outcomes for the young (aged 16-45) to that for the older (aged 46-59) working-age adults. Mullahy and Sindelar (1996) point out that both labour market and drinking/smoking behaviour are generally qualitatively different across individuals at different age groups, so that empirical analyses should consider age groups separately. In other words, the underlying question is whether the young and the older working-age adults would yield different results when examining the effect of drinking and smoking behaviour on labour market outcomes.

The structure of this chapter is as follows: Section 6.2 provides a theoretical background and reviews the results of previous empirical research. Section 6.3 presents the data and variables, along with the econometric methods used in this chapter. Results are discussed in Section 6.4 and are followed by conclusions in Section 6.5.

6.2 Theoretical Background and Empirical Evidence

6.2.1 Theoretical Background

There are several channels through which drinking and smoking behaviour might impact on labour market outcomes.

Alcohol consumption could influence labour market outcomes through a health channel. There is a body of medical literature documenting an inverse U-shaped relationship between alcohol consumption and health (Beaglehole and Jackson, 1992; Boffetta and Garfinkel, 1990; Doll *et al.*, 1994; Marmot and Brunner, 1991; and Shaper, 1993). For example, light or moderate drinkers have lower risk of cardiovascular disease than either non-drinkers or heavy drinkers. Moderate alcohol consumption may be actually beneficial for health by reducing the incidence of coronary heart disease, relieving stress or depression, and improving cognitive performance (Baum-Baicker, 1985a; and 1985b), thereby positively affecting labour productivity (Tekin, 2002). On the other hand, the harmful health consequences associated with alcoholism or excessive alcohol intake, which involve gastrointestinal diseases, heart disease, endocrine and metabolic disorder, cancers of the head and throat, and psychological and cognitive impairments (Last, 1998; Lye and Hirschberg, 2004), would negatively affect labour market outcomes (Rice et al., 1990; Kenkel and Ribar, 1994; and Mullahy and Sindelar 1993).

In addition to the health channel, there are unobserved factors which might be associated with both drinking and labour market outcomes (Tekin, 2002). For example, alcohol may serve as a bridge of communication with colleagues, employers and customers and thus helps workers to stay employed or increase promotion opportunities (MacDonalds and Shields, 2001). However, it is also argued that heaving drinking might in fact send a negative signal to employers about the individual's productivity or suitability for promotion (Tekin, 2002).

The adverse health effect of smoking, such as causing heart disease, stroke, lung disease, cancers and many other health problems, has been examined by medical researchers (Brown *et al.*, 1987; Doll, 1986; and Mattsom *et al.*, 1987). Combining this medical evidence related to smoking, there are four main mechanisms through which smoking has a negative effect on labour market outcomes (Levine *et al.*, 1995; Levine *et al.*, 1997; Lin and Chen, 2010; and Lye and Hirschberg, 2004). Firstly, smokers are more likely to be unhealthy and are therefore less productive due to higher absentee rates or inferior performance on certain tasks. Secondly, it may be more costly to hire smokers due to increased health and fire insurance, or higher cleaning costs (Swart, 1990; and Levine *et al.*, 1997). According to the theory of compensating wage differentials, employers would be willing to employ smokers at lower wages by offsetting higher insurance costs. Smokers may also be willing to accept jobs with lower wages that provide medical insurance. Thirdly, there are unobserved factors, for example, time preference, which might be correlated with both smoking and labour market outcomes. For example, if smokers are individuals who have high time preference rates (that is, strongly discount their future lifetime utility), they may be less likely to invest in productivity enhancing human capital, resulting in lower wages (Heineck and Schwarze, 2003). Finally, smokers may be subjected to discrimination from employers, colleagues or customers, thereby having adverse effect on labour market outcomes.

6.2.2 Empirical Evidence

Evidence on The Effect of Drinking on Labour Market Outcomes

Most research has documented that alcohol abuse or alcoholism has detrimental effects on an individual's labour market outcomes. French and Zarkin (1995) explore the relationship between alcohol consumption and wages in the US using both a full sample of randomly selected workers and a sub sample of prime-age workers (aged 30-59) from four worksites. Their analysis employs OLS, and uses four binary variables of alcohol consumption which are 'Never drinkers', 'Former drinkers', 'Daily drinkers', 'Heavy drinkers', and also one continuous variable of alcohol consumption. 'Never drinkers' is defined as never drink any alcohol. According to SAMHSA (1992, 1993)[140], 'Former drinkers' are defined as drinking in the past, but not during the past 12 months. 'Daily drinkers' are defined as drinking on 20 or more days in the 30 days, and 'Heavy drinkers' are defined as consuming 5 or more drinks per occasion on 5 or more days during the same period. The fifth variable (ALCUSE) is the number of drinks consumed during the past 12 months. Furthermore, French and Zarkin (1995) add a quadratic and a cubic ALCUSE in order to examine a potential nonlinear relationship between alcohol consumption and wages. The findings suggest an inverse-U-shaped relationship between drinking status and weekly wages with a turning point at approximately 1.69 to 2.40 drinks per day. They also find evidence that moderate (or 'Daily') drinkers have higher weekly wages than abstainers (including both 'Never' and 'Former' drinkers) and heavy drinkers at these four worksites. Moreover, the results from a sensitivity analysis based on including and excluding the human capital and family variables (which include education, health status, tenure and marital status) suggests that alcohol

[140]In French and Zarkin (1995), SAMHSA (1992) refers to 'Substance Abuse and Mental Health Services Administration (SAMHSA), 1992, National Household Survey on Drug Abuse: Main Findings 1991 (US Department of Health and Human Services, Rockville, MD)'; and SAMHSA (1993) refers to 'Substance Abuse and Mental Health Services Administration (SAMHSA), 1993, National Household Survey on Drug Abuse: Highlights 1991, HHS Publication No. SMA 93-1979 (U.S. Department of Health and Human Services, Rockville, MD)'.

consumption affects wages through human capital and family variables. However, the authors argue that there are four limitations in their study. First, they cannot estimate fixed effects model as the data is cross-sectional. Second, the self-reported alcohol consumption measures may induce measurement error. Third, the analysis fails to account for the possible endogeneity of drinking since their data lacks suitable instruments. Moreover, their study fails to account for the potential sample selection bias.

Hamilton and Hamilton (1997) study the relationship between drinking and wages using a sample of 1,741 prime-age (aged 25-59) Canadian male employees from the 1985 General Social Survey (GSS). Annual wages are used as the dependent variable since the information on weekly wages or hours worked per week are not provided in the survey. To address the potential endogeneity of drinking status and sample selection bias, the analysis applies a polychotomous choice model, where the drinking status equations are estimated using a multinomial logit method and the wages equations are estimated using a Heckman selection method. The analysis also estimates wage differentials for different types of drinkers by following the technique of Idson and Feaster (1990). The authors define three types of alcohol consumers by taking into account both frequency and intensity of use on each occasion: i) non-drinkers, are individuals who drink not at all or less often than once a month during the past year; ii) moderate drinkers, are individuals who drink once a month, once a week, or every day, but consume no more than eight drinks[141] on a single day during the past week; and iii) heavy drinkers, are individuals who consume eight or more drinks during the past week, and drink at least once a week during the past year. They use separate price measure for spirits, wine and beer as instruments for drinking status. However, no formal test of instrument validity is performed by the authors. The results show that moderate alcohol consumption leads to increased wages relative to abstention, whereas heavy alcohol consumption leads to reduced wages relative to moderate alcohol consumption. Hamilton and Hamilton (1997) also find that there is a positive effect of sample selection bias which almost offsets the negative wage differential due to heavy drinking in estimating the wage gap between moderate drinking and heavy drinking.

[141] The GSS defines 'one drink' as one and a half ounces of liquor, a small glass of wine, or a one-pint bottle of beer which being consumed by individuals in the past week.

Based on French and Zarkin (1995), Zarkin *et al.* (1998) assess whether the inverse U-shaped relationship between alcohol use and wages can be replicated on a combined sample of US prime-age employees (which include 5,948 males and 6,177 females, aged 30 to 54) from the 1991 and 1992 National Household Surveys on Drug Abuse (NHSDA). According to Zarkin *et al.* (1998), the alcohol consumption questions of the survey introduce self-administered answer sheets which increase the confidentiality and anonymity of the answers, thereby minimizing the measurement error. The authors utilize their calculated continuous drinking variable [142] to define four types of alcohol consumption which include eight drinking categories: one category for non-drinkers, two for light drinkers, three for moderate drinkers, and two for heavy drinkers. Then multivariate OLS models are estimated using three variations of these eight drinking categories.[143] The first variation involves all the drinking categories where non-drinkers are the reference category. The second variation comprises four types of alcohol consumption described above: non-drinkers (the reference category), light drinkers, moderate drinkers and heavy drinkers. The final variation is simply defined as a binary alcohol 'use' (that involves seven positive drinking indicators) / 'no use' indicator (no use is the reference category). Findings suggest that alcohol consumption has a positive and significant impact on hourly wages for males, but not for females. In other words, male workers who consume alcohol have approximately 7% higher wages than male workers who are non-drinkers. This paper also finds an inverse U-shaped relationship between alcohol consumption and wages for males at a lower drinking level than that in French and Zarkin (1995). In addition, the authors find no evidence that high level of alcohol consumption negatively affect wages. However, they argue that their analysis fails to account for the possible endogeneity of drinking due to lack of suitable instruments.

MacDonald and Shields (2001) explore the relationship between drinking and mean occupational hourly wages based on a pooled sample of 15,819 men (aged 25-65) and 18,430 women (aged 25-60) from the Health Survey for England between 1992 and 1996. The authors define three measures of drinking status according to either drinking intensity or drinking frequency during the last year. The first measure is a

[142] In Zarkin *et al.* (1998), the total number of drinks an individual consumed during the last 30 days, which ranges from 0 to 1,800 drinks for males and from 0 to 900 drinks for females, is calculated by multiplying the typical number of drinks an individual consumed per occasion by the number of times that individual drank during the same period.
[143] Based on the Heckman model, Zarkin *et al.* (1998) find no evidence of sample selection bias.

continuous variable indicating the average weekly consumption in units[144]. Based on the first measure of drinking status, the second measure includes seven dummy variables indicating seven types of alcohol consumers: abstainers, light drinkers, light to moderated drinkers, moderate drinkers, moderate to heavy drinkers, heavy drinkers, and very heavy drinkers. The last measure includes five dummy variables based on the number of episodes of drinking per week: non-drinkers, infrequent drinkers, occasional drinkers, frequent drinkers and daily drinkers.[145] In addition to estimating pooled OLS, MacDonald and Shields (2001) adopt 2SLS to control for endogeneity of drinking. Three sets of instruments are used: three dummy variables for long term non-acute illnesses (diabetes, stomach ulcers and asthma); a binary variable indicating whether or not a respondent's mother or father smoked regularly; and three dummy variables indicating whether the respondent feeling that he/she should reduce drinking, feeling guilty about one's own drinking, and having been annoyed by criticism from others about one's own drinking during the last three months, respectively. The results suggested a positive relationship between alcohol consumption and wages that appeared to have an inverted U-shape for both genders. However, MacDonald and Shields (2001) argue that the estimated results based on the IV model should be treated with considerable caution since these are found to be much larger than those based on the OLS. Furthermore, they find no evidence that drinking is related to wages through family structure and health status variables.

Tekin (2002) utilizes data from four waves of the Russia Longitudinal Monitoring Survey (RLMS) to examine the effects of alcohol consumption on employment and hourly wages for 974 male and 1,382 female prime-age workers (aged 24-58). Employment is measured by whether the individual is currently employed. The hourly wages is defined as the ratio between the total monthly earnings[146] and the total number of hours worked in the last month. The author utilizes three measures of drinking status based on the self-reported alcohol consumption questions in the RLMS. The first measure is a binary variable based on a survey question concerning whether respondents consumed any alcohol in the last month. Those who respond

[144] According to MacDonald and Shields (2001), 1 unit of alcohol is equivalent to 8 grams of ethanol, or approximately 1.5 pint of beer, a small glass of wine or a single measure of spirits.

[145] For detail information of the alcohol consumption definition, see page 434 in MacDonald and Shields (2001).

[146] The author defines total monthly earnings as the sum of salaries, wages, bonuses, grants, benefits, profits and non cash in-kind (see p.10 in Tekin, 2002).

yes to the above question are then asked to report the frequency of the alcohol use during the same period, including: every day, 4-6 times a week, 2-3 times a week, once a week, 2-3 times and once a month. Accordingly, the second measure which is similar to the first measure of alcohol use in Zarkin *et al.* (1998) is created. The third measure of alcohol consumption is a continuous variable indicating monthly ethanol intake (per litre). The author also includes linear and quadratic measures of ethanol intake to capturing the possible nonlinear relationship between alcohol consumption and labour market outcomes. In addition to estimating pooled OLS model, Tekin (2002) employs a fixed effects estimator to control for the individual unobserved factors. The results show that after controlling for individual unobserved factors: (i) the positive effect of alcohol consumption on employment disappears for men and decreases for women; (ii) alcohol consumption is positively related to wages for both genders and the effect is larger for men than for women; and (iii) for both genders, the inverse U-shaped relationship between alcohol consumption and labour market outcomes, which is observed in the pooled OLS analysis, disappears. Furthermore, a number of sensitivity tests are performed with the following results: (i) there is no evidence that alcohol consumption is related to labour market outcomes through human capital and family variables;[147] (ii) the results are robust to the way the continuous measure of alcohol use is defined; and (iii) based on a Heckman selection method, there is no evidence for the presence of sample selection bias for both genders. However, Tekin (2002) argues that the analysis fails to control for the possible endogeneity of alcohol consumption due to lack of suitable instruments.

Barrett (2002) estimates the impact of drinking on individual hourly earnings using a sample of 5,705 working males (aged 25-59) from the first wave of the Australian National Health Survey (NHS89). By applying the same definition as Hamilton and Hamilton (1997), three types of drinkers are defined (non-drinkers, moderate drinkers and heavy drinkers). The estimation procedure used in this analysis is similar to that in Hamilton and Hamilton (1997). To account for the potential endogeneity of drinking status, the author introduces two sets of variables as instruments for drinking behaviour which include a dummy variable indicating whether an individual drank at the age of 18, and a set of variables measuring the

[147] Although the human capital and family variables used in Tekin (2002) are the identical to those used in French and Zarkin (1995), the latter draws the opposite conclusion.

proportion of sample members in that individual's local region who are non-drinkers, who are moderate drinkers and who are heavy drinkers. A Durbin-Wu-Hausman (DWH) specification test approach is used and the result supports the validity of both instruments. Since the grouped regression results support endogenous selection for heavy drinkers (and exogenous selection for the other two types of drinkers), Barrett (2002) further exploits both unconditional and conditional wage decomposition method to isolate the effect of drinking status on wages. It is found that moderate drinking leads to a large wage premium relative to non-drinking (11%) and heaving drinking (19%). Moreover, a large wage penalty for heavy drinking relative to non-drinking (8%) is also found.

Evidence on The Effect of Smoking on Labour Market Outcomes

Levine *et al.* (1995) investigate the effect of smoking on weekly wages and annual hours worked for young American workers (aged 27-34) by using data from the National Longitudinal Survey of Youth (NLSY) 1984 and 1992. Additionally, the authors also examine the effect of smoking on wages for men and women separately. The study applies three methods. First, standard cross-sectional OLS is used to estimate the effect of smoking on labour market outcomes for each survey year. The second method is a FE method, which exploits the panel nature of the data to control for individual specific characteristics. The third method takes advantage of the household structure of the NLSY by looking at the differences in labour market outcomes and smoking across siblings, which controls for the unobservable characteristics that are constant within a family. The authors conclude the following: (i) similarity across the three estimates suggests that there is no evidence of endogeneity of smoking behaviour; (ii) smoking reduces wages by about 4-8%, but there is no statistically significant effect of smoking on hours worked; (iii) men who smoke are found to earn 10.5% less than non-smoking men, whereas smoking women are found to earn 3.3% more than non-smoking women. However, there is a potential disadvantage of defining smoking status in this study. Based on the NLSY 1984, smoking is defined as respondents who report smoking at least one cigarette per day, on average, over the last month. Using the NLSY 1992, smoker are defined as whether or not the respondents smoke daily. Levine *et al.* (1995) argue that using different definition of smoking may lead to some false transitions in smoking status during the survey periods.

Based on the German Socio-Economic Panel (GSOEP) 1998, 1999, and 2001, Heineck and Schwarze (2003) examine the effect of smoking behaviour on monthly earnings for both male and female employees aged 25-55. Both cross-sectional and longitudinal models are estimated. In the cross-sectional OLS analysis (GSOEP, 1998), three types of smoking measures are used. The first measure is defined as whether an individual is currently smoking. The second measure is a continuous variable indicating the daily tobacco consumption amount. The third measure includes five dummy variables indicating whether the individual smokes up to 10, 11-20, 21-30, 31-40, and 41 and more cigarettes or other tobacco products per day. For the pooled OLS, the RE and the FE analysis, only the first smoking measure is used since the amount of tobacco consumption is not available for the other two survey years. Results from the cross-sectional analysis suggest that smoking has a negative effect on earnings for men. However, the FE estimation result shows that young male smokers aged 25-35 earn 2.5% more than non-smoking counterparts. There is no statistically significant relationship between smoking and earnings for women. The Hausman test for the absence of unobservable individual effect is rejected for both genders, thereby supporting the FE estimates.

Lin and Chen (2010) examine the effect of smoking on the probabilities of being employed and monthly earnings using data from Taiwan's Panel Survey of Family Dynamics (PSFD) 1999, 2000 and 2004. Smoking is defined as whether or not respondents currently smoking. The study employs both a pooled OLS method and a IV method based on PSFD 2004. The authors use a combined index of individual knowledge regarding adverse effects of smoking on health as the instrumental variables to correct for potential endogeneity of smoking. The results find that (i) according to the Hausman test statistics in the IV estimation of the employment equation, the null hypothesis that smoking is exogenous cannot be rejected for both genders, whereas in the IV estimation of the earnings equation, the null hypothesis that smoking is exogenous is rejected for both genders, suggesting that the IV model is preferable; (ii) smoking has a statistically significant and negative effect on employment for both men (-4.26%) and women (-6.64%). In addition, men who smoke earn 3.7% less than those who do not smoke. There is no statistically significant relationship between smoking and earnings for women.

Evidence on The Simultaneous Effect of Drinking and Smoking on Labour Market Outcomes

Van Ours (2002) considers the impact alcohol and tobacco consumption of hourly wages for both men and women (aged 26-55) in the Netherlands. The author applies three sets of drinking and smoking measures. In the first set, both drinking and smoking are continuous variables. Drinking is defined as the average number of glasses of alcohol (beer, wine, genever) the individual consumes per month. Smoking is defined as the average number of cigarettes, cigars or pipes the individual smokes per day. In the second, eight dummy variables are used to indicate alcohol consumption group (0, 1-5, 6-16, 17-31, 32-62, 63-93, 94-124 and 125 or more drinks). Smoking in the second set involves five dummy variables indicating daily tobacco consumption (0, 1-2, 3-10, 11-20, more than 20). In the third set, both drinking and smoking are binary variables indicating whether an individual drinks and smokes, respectively. By applying both OLS and an analysis of starting rates for alcohol and tobacco to identify the presence of unobserved heterogeneity, the author concludes that for men drinking and smoking have a positive effect (9.8%) and a negative effect (9.0%) on wages, respectively. The OLS method overestimates the positive returns to drinking and underestimates the smoking penalty for men. However, women's wages are not affected by drinking or smoking.

By using a sample of 4,338 Australian working men from the second wave of the NHS (NHS95), Lye and Hirschberg (2004) examine the effect of both alcohol consumption and smoking on annual wage income. The sample of this study is divided into smokers (current smokers) and non-smokers (include ex-smokers) in order to model the interaction of smoking with drinking status. An estimated amount of weekly alcohol consumption is used to measure drinking status. The estimation procedure is similar to Barrett (2002) in terms of addressing the selectivity bias and endogeneity of drinking status through an ordered probit model and a simultaneous equation model. Lye and Hirschberg (2004) employ five sets of instrumental variables which are: (i) a set of four dummy variables that measure BMI; (ii) total amount of time spent on exercise; (iii) a variable that measures the number of doctor visits; (iv) a set of country of birth variables; (v) a set of state dummy variables. The results confirm the inverse U-shaped relationship between alcohol consumption and wages for non-smokers, which are consistent with the findings of Barrett (2002), but no significant relationship between alcohol consumption and wages for smokers.

Auld (2005) investigates the effect of both drinking and smoking on annual income for Canadian male (aged 25-59) workers using cross-sectional data from the GSS, 1985 and 1991. Smoking is defined using a binary variable indicating whether an individual is a daily smoker. Following Hamilton and Hamilton (1997), the author defines three types of alcohol consumers: abstainers, moderate drinkers and having drinkers. The estimation methods include the OLS, a maximum likelihood (ML), and a full information maximum simulated likelihood (FIMSL). Auld (2005) uses religious status, an indicator for Catholicism, and prices of alcohol and tobacco as instruments to address the endogeneity of drinking and smoking. The results conclude that: (i) the exogeneity of drinking and smoking status is rejected; (ii) moderate drinkers are estimated to earn 10% more than abstainers, but to earn 2% less than heavy drinkers, these findings are consistent across models; (iii) treating smoking as endogenous increase its negative effect on income (from -8% to -24%); (iv) failing to control for smoking when estimating the effect of drinking on wages leads to small downward bias in the estimated effect of heavy drinking on income and small upward bias in the effect of drinking abstention; and failing to control for drinking when estimating the effect of smoking underestimates of the negative smoking effect; (v) controlling for a measure of leisure-time physical exercise or excluding self-reported health status variables almost does not alter the estimated drinking and smoking effects, implying that the results are not driven by unobserved components of health status.

Based on panel data from the National Job Corps Study (NJCS), Neumann (2009) examines the effect of drinking and smoking on labour market outcomes for the American lower income young adults (aged 16-24). Drinking and smoking are simply defined as whether or not a respondent is a drinker and a smoker, respectively. The three labour market outcome variables involve average weekly earnings, a dummy variable indicating whether a respondent is working, and average weekly hours worked. The study employs both a pooled OLS and a FE using baseline treatment status, contemporaneous treatment status and lagged treatment. The results show that drinking has a statistically significant and negative effect (about 6-16 percentage points) on hours worked. However, there is no statistically significant effect of drinking on earnings or employment status, and no statistically significant relationship between smoking and labour market outcomes. Moreover, not controlling for the individual specific unobserved factors or the unobserved factors that change over time would overestimate the effect of drinking on labour market outcomes.

Overall, some studies have found that moderate and even heavy drinking raise personal earnings (French & Zarkin, 1995; Hamilton and Hamilton, 1997; Zarkin *et al.*, 1998; and Barrett, 2002), while others identify an inverted U-shaped relationship between alcohol consumption and wages (French and Zarkin, 1995; Heien 1996; Hamilton and Hamilton, 1997; Zarkin *et al.*, 1998; and Lye and Hirschberg, 2004). Meanwhile smoking has been shown to be negatively correlated with labour market outcomes (Levine *et al.*, 1995; Levine *et al.*, 1997; and Lin and Chen, 2010). Several studies use an IV approach to address the endogeneity of drinking/smoking behaviour but the results of these are somewhat mixed (Levine *et al.*, 1995; Levine *et al.*, 1997; Hamilton and Hamilton, 1997; Barrett, 2002; Lye and Hirschberg, 2004; Neumann, 2009; and Lin and Chen, 2010).

6.3 Data and Methodology

6.3.1 Data and Variables

The data used in examining the relationship between health related behaviour and labour market outcomes comes from the six waves of the CHNS (CHNS, 1991-2006) as described in Chapter 3. The first wave of the CHNS (CHNS, 1989) is not used because it did not survey drinking or smoking status. While the latest wave of the CHNS (CHNS, 2009) surveyed drinking and smoking status, it did not survey the self-reported health status (and neither did the first wave) which is used as explanatory variables for employment and wages. It is not possible to test the indirect effects of drinking and smoking status on labour market outcomes without having information on self-reported health status. Therefore, the 2009 CHNS is not used in this chapter.

Labour Market Outcomes

Two labour market outcomes are considered, namely employment status and real hourly wages, which have been introduced in Chapter 4 and Chapter 5, respectively. Recall from Chapter 4 employment status is defined as a dummy variable, *employed*, which equals to one if an individual is presently working (including self-employed), and equals to zero otherwise. Recall from Chapter 5 the average hourly wages is defined as the ratio between the monthly wages and the monthly working hours. Monthly wages are calculated by dividing annual wages by working months. Monthly working hours are calculated by multiplying 4.3 weeks per month by weekly working hours. Then the nominal hourly wages are deflated to 2006 Chinese yuan to obtain the average real hourly wages.

In estimating the effects of drinking and smoking status on employment, the total sample used in this chapter is based on the total sample used in Chapter 4. After removing the observations that have missing values for drinking or smoking status, which accounts for 8% of the sample, the total sample used in the employment analysis is 14,718 individuals from 35,791 observations pooled across the six waves (for each $i, t = 1, \dots, T$, where $T \leq 6$ in a case of the unbalanced panel). Then the sample is split by gender and by young and older working-age. Recall from Chapter 4, young working-age adults are defined as adults who are aged between 16 and 45, and older working-age adults are defined as adults who are aged 46 and over. The sub samples within the population involve 7,525 men and 7,193 women (20,345 and 17,246 observations, respectively), and 29,126 young (aged 16-45) and 8,465 older (aged 46-59) working-age adults (12,763 and 4,707 observations, respectively).

As discussed in both Chapter 3 and Chapter 5, the low response rate of wage income is mainly due to the low response rate of people who are agricultural workers and/or who are self-employed. Therefore, following the sample construction of Chapter 5, this chapter also excludes observations who are either agricultural workers or self-employed. Therefore, the sample used in the wage analysis includes 5,911 individuals (10,871 observations) and is comprised of 3,356 men and 2,555 women (6,364 and 4,507 observations, respectively), and 5,058 young and 1,454 older working-age adults (8,716 and 2,155 observations, respectively).

Drinking and Smoking

The CHNS asks respondents whether or not they drank beer or any other alcoholic beverage during the past year. Those who answer in the affirmative are then asked information about their drinking frequency during the same period.[148] Five possible responses are: '*daily or almost every day*', '*3-4 times a week*', '*once or twice a week*', '*once or twice a month*' and '*no more than once a month*'.[149] From 1993, respondents who drank last year were asked how much they drink each week (in units) based on three types of alcohol: beer (unit = bottle), grape wine including various coloured wines and rice wine (unit = liang, where 1 liang = 50 grams), and liquor (unit = liang). According to these alcohol consumption questions, three measures of drinking status, which are similar to those in Tekin (2002), are constructed. The first measure is a dummy variable, *drink*, which takes on the value of one if an individual has consumed any alcoholic beverage during the previous year, and zero otherwise. The second measure is constructed by combining the responses of the first two questions above, that is, six dummy variables which are named '*nodrink* (the reference category), '*drink0to1m*' , '*drink1to2m*', '*drink1to2w*', '*drink3to4w*'and '*drinkdaily*', respectively. The third measure is a continuous variable, *wkdrinkamt*, indicates the average weekly alcohol consumption amount (in units). The three measures of alcohol consumption are used in both the analysis employment and wages.

The CHNS asks respondents whether or not they have smoked cigarettes, including rolled or manufactured. Those who answered in the affirmative are then asked whether or not they are smoking currently and their daily cigarettes consumption amount.[150] Combining the answers of these smoking questions, this chapter uses four measures of smoking status, where the latter three are similar to the measures of smoking status in Heineck and Schwarze (2003). The first measure is a dummy variable, *eversmoked*, which takes the value of one if the respondent has ever smoked

[148] According to the CHNS Work Manual (1993), respondents who drink occasionally once or twice in a year are not considered as drinking alcohol.
[149] For wave 1991, 1993 and 2000, the alcohol consumption questions were listed in the physical examination survey questionnaire. For other three waves, they were listed either in the household survey questionnaire or in the adult survey questionnaire.
[150] For wave 1991, 1993 and 2000, the cigarettes consumption questions were listed in the physical examination survey questionnaire. For other three waves, they were listed either in the household survey questionnaire or in the adult survey questionnaire.

cigarettes, and zero otherwise. The second measure is a dummy variable, *smokenow*, which takes value of one if the respondent is a current smoker, and zero otherwise.[151] The third measure is a set of five dummy variables: *nonsmoke* (the reference category), *smoke1to10*, *smoke11to20*, *smokemt20*, *smokemiss*, which take a value of one if individuals do not smoke, smoke up to 10, 11-20, and more than 20 cigarettes, and do smoke but have missing values on their average daily cigarettes consumption amount, respectively. The last measure of smoking is a continuous variable, *dailysmokeamt*, indicates the number of cigarettes being smoked per day.

Figures 6.2 and 6.3 present the prevalence of drinking and smoking status for all the sample groups within the population by each survey year, respectively. On average, the prevalence of drinking and smoking status for men and for the older working-age adults are higher than those for women and for the young working-age adults, respectively. For women, unlike other countries, the prevalence of both drinking and smoking is very low, which accounts for 10% and 2%, respectively. The low prevalence of smoking of women is even lower than that found in Levy (2006). As shown in Figure 6.2, there is a relatively flat trend of the prevalence of drinking for the total sample of working-age adults and for men between 1991 and 2006. For women, there is a slight declining trend, from 11% in 1991 to 7% in 2006. For the young working-age adults, there is a slight declining trend after 1997 (from about 36% to 33%). For the older working-age adults, the prevalence of drinking alcohol decreases from 55% to about 50% between 1991 and 2006, with a minimum of 48% in 2004. In Figure 6.3, the rate of smoking almost unchanged for the total sample and for women over the period. However, there is a declining trend in smoking for men (from 68% to 58%) and for the older working-age adults (from 55% to 47%). For the young working-age adults, the rate of smoking decreases slightly from 32% in 1991 to 27% in 2006.

Table 6.1a presents summary statistics of the average employment rate across different measures of drinking and smoking status for the Chinese working-age population and sub samples within the population. The average employment rates by discrete measures of drinking status and of smoking status are shown in the upper and the middle panel of Table 6.1a, respectively. The continuous measure of the

[151] According to the CHNS Work Manual (1993) and (2006), smoking is considered as a person smokes at least one cigarette per day. Otherwise, the person should not be considered as smoking.

average weekly alcohol consumption (in units) and the average daily cigarette consumption (in number of cigarettes) amount are separately shown in the bottom panel of the table. For the Chinese working-age population, people who presently drink alcohol have higher employment rate (91%) than those who do not (85%). In addition, people who drink once or twice a week (*drink1to2w*) and 3-4 times a week (*drink3to4w*) have higher employment rate than other alcohol consumption groups. For young working-age adults (aged 16-45), the higher the drinking frequency, the higher the average employment rate, whereas the reverse is almost true for older working-age adults (aged 46-59).

According to the middle panel of Table 6.1a, lifetime/current smokers have a higher employment rate (91%) than lifetime/current non-smokers (85%). This is also true for the sub samples except for women. For both the young and the older working-age sample, people who smoke more than 20 cigarettes a day have a slightly higher employment rate than those who smoke up to 10, and 10 to 20 cigarettes a day. However, for the total and the male sample, the higher employment rate for smokers does not distinguish to the number of cigarettes consumed per day.[152] In addition, 36% of individuals have ever smoked, and 97% of them are still consuming cigarettes. In other words, there are 3% of people (422 observations) who have ever smoked but not current smokers, i.e. they have quit smoking. Appendix 6.4 presents summary statistics for this group of people (refers to as ex-smokers) and for those who still smoke (refers to as current smokers). On average, people who have quit smoking are older, have lower level of self-reported health status and are better educated than those who still smoke. For example, the average age of ex-smokers is about 8 years older than that of current smokers. In addition, 60% of ex-smokers are in the older age group (aged 46 to 49), which is almost double that of current smokers. Compared to current smokers, ex-smokers have higher proportion of having fair and poor health status, especially the proportion of the latter, which is three times as much as that of current smokers. For ex-smokers, the proportion of those with technical or vocational degree qualifications (*techvoc*) and with university or higher degree qualifications (*university*) account for 10% and 5%, respectively.

[152] Since both the number of female smokers who smoke more than 20 and who have missing values on their daily cigarettes consumption amount are so small, the summary statistics for these two groups may not appropriately present the average employment rate for these two types of population.

However, for current smokers these two proportions account for 5% and 3%, respectively.

As shown in the lower panel of Table 6.1a, the sample of working-age adults used in the analysis of the continuous measures of drinking and smoking status are different to that when applying the discrete measures. As mentioned earlier, the wave 1991 did not survey the average amount of alcohol consumed each week, thus observations in this wave are excluded. Additionally, observations that have missing values on alcohol and/or cigarette consumption, which account for 3% of the total sample of 30,015 observations across the five waves, are also excluded. On average, individuals consume approximately 4 units of alcohol per week (with a minimum zero value indicating non-drinkers and a maximum of 128 units) and 6 cigarettes per day (with a minimum zero value indicating non-smokers and a maximum of 60 cigarettes), respectively. For individuals who drink alcohol, an average of 11 units is consumed each week. For those who are current smokers, an average of 16 cigarettes is consumed daily, which is similar to that in Pan and Hu (2008). Moreover, men and the older working-age adults have both higher drinking and smoking intensity than women and the young working-age adults, respectively.

There is a relationship between drinking alcohol and smoking. Table 6.1b presents the sample distribution and the correlation between drinking and smoking status. It shows that 25% of the Chinese working-age adults both drink alcohol and smoke, 14% are drinkers but do not smoke, and 10% smoke but do not drink alcohol. For individuals who drink alcohol, 65% of them are smokers. For individuals who smoke, 72% of them are drinkers. The correlation between drinking alcohol and smoking is 0.50 (in a case of ex-smokers, the correlation is 0.51).

Table 6.1c presents the average real hourly wages across different measures of drinking and smoking status. It shows that the average real hourly wages are higher for drinkers (3.52 Chinese yuan) than non-drinkers (2.98 Chinese yuan). Individuals who drink once or twice a week (*drink1to2w*), 3-4 times a week (*drink3to4w*), and who almost drink every day (*drinkdaily*) have higher average real weekly wages than other three dinking consumption groups except for women. As can been seen from the middle panel of Table 6.1c, individuals who have ever smoked (3.47 Chinese yuan) and who are current smokers have higher hourly wages (3.38 Chinese yuan) than those who are lifetime non-smokers (3.08 Chinese yuan) or who currently do not smoke (3.14 Chinese yuan), respectively. In addition, summary statistics in the lower panel of Table 6.1c are similar to those in Table 6.1a.

6.3.2 Methodology

Methodology – Treating Drinking and Smoking as Exogenous Variables

This chapter firstly uses the pooled probit model to estimate the effect of drinking and smoking status on employment for the Chinese working-age adults. The pooled probit model, which assumes $P(y_{it} = 1|x_{it}) = \Phi(\vartheta_{dk}dk_{it} + \vartheta_{sk}sk_{it} + \vartheta_0 h_{it} + \vartheta_1 x_{it})$, is as follows:

$$y_{it} = \begin{cases} 1 \text{ if } y_{it}^* = \vartheta_{dk}dk_{it} + \vartheta_{sk}sk_{it} + \vartheta_0 h_{it} + \vartheta_1 x_{it} + o_{0it} > 0, \\ \qquad i = 1, \dots, n, t = 1, \dots, T \\ 0 \text{ otherwise} \end{cases} \tag{6.1}$$

where, P indicates probability; i indicates individuals; t = indicates time periods; Φ indicates the standard normal Cumulative Distribution Function (CDF); y_{it} indicates a binary response indicator of employment; y_{it}^* indicates a latent variable of employment; dk_{it} indicates a vector of drinking status variables; sk_{it} indicates a vector of smoking status variables; h_{it} indicates a vector of SRHS variables; x_{it} indicates a vector of other explanatory variables that affect employment, including gender, age, educational attainment, marital status, living in urban/rural area, region of residence, and presence of children in the household;[153] ϑ_{dk}, ϑ_{sk}, ϑ_0, and ϑ_1 indicate the estimated coefficients, where ϑ_{dk} and ϑ_{sk} are measures of the effect of drinking and smoking status on employment, respectively; o_{0it} indicates random errors which are assumed to follow the standard normal distribution with variance 1, and $o_{0it} \sim N(0,1)$. Under the pooled probit, cluster-robust standard errors are used to correct for correlation over time for a given individual.

By allowing for the panel nature of the data, this chapter uses the RE probit model which assumes that the heterogeneity across individuals is time-invariant,[154] and the error term o_{0it} in Equation (6.1) is decomposed as:

[153] For more detail regarding the definition of explanatory variables, see Chapter 3.
[154] This chapter does not use a probit with FE for two reasons. Firstly, it only relies on within-individual variations in the dependent and the explanatory variables. In other words, individuals who do not change their employment status will not be included in the estimation, neither the coefficients of the time-invariant explanatory variables are estimated. In such case, 81% of the sample will not be included. Moreover, more than 70% and 80% of the sample does not change the drinking and

$$o_{0it} = \kappa_{1i} + \mu_{1it} \quad (6.2)$$

where, κ_{1i} indicates the individual specific unobservable effect, assuming that $\kappa_{1i} \sim IN(0, \sigma_{\kappa_1}^2)$ and is independent of the μ_{1it} and x_{it}, μ_{1it} indicates a random error term, assuming $\mu_{1it} \sim IN(0, \sigma_{\mu_1}^2)$ and is independent of x_{it}. The correlation between two successive error terms for the same individual i is a constant, given by:

$$\rho_3 = corr(o_{0i2}, o_{0i1}) = \frac{\sigma_{o_0}^2}{\sigma_{o_0}^2 + \sigma_{\mu_1}^2} \quad (6.3)$$

A likelihood-ratio test of the significance of ρ_3 (H_0: $\rho_3 = 0$) compares the pooled probit estimator with the RE probit estimator (StataCorp, 2009). If $\rho_3 = 0$, the pooled probit parameter estimates will be equal to the RE probit estimates, then it is appropriate to use the pooled probit estimator. If $\rho_3 \neq 0$, then it suggests that the RE probit estimator would be appropriate.

To investigate the effect of drinking and smoking status on wages for the Chinese workers, this chapter adopts the pooled OLS and the FE estimator which controls for the individual unobserved factors. The pooled OLS is based on a modified Mincerian Wage Model, which is:

$$w_{it} = \pi_0 + \pi_{dk} dk_{it} + \pi_{sk} sk_{it} + \pi_1 h_{it} + \pi_2 e_{1it} + \varrho_{0it}, \quad (6.4)$$

where, w_{it} indicates the logarithm of real hourly wage income; e_{1it} indicates a vector of other explanatory variables that affect wages, including gender, age, educational attainment, marital status, living in urban/rural area, region of residence, and presence of children in the household, occupational status, employment position, type of work unit, number of employees in work unit, and whether or not an individual has a second job; π_0 indicates a constant term; π_{dk}, π_{sk}, π_1 and π_2 indicate the estimated coefficients, where π_{dk} and π_{sk} are measures of the effect of drinking and smoking status on wages, respectively; ϱ_{0it} indicates a disturbance term, which is uncorrelated with the explanatory variables, and assumes that $\varrho_{0it} \sim N(0, \sigma_{\varrho_0}^2)$. Under the pooled OLS, cluster-robust standard errors are used to correct for error correlation over time for a given individual.

The formulation of the FE model can be written as:

---

smoking status, respectively. Therefore, the FE coefficients might be imprecise and have large standard errors in a case where within-individual variation is small relative to the between-individual variation. Secondly, FE probit might introduce incidental parameters problem (Heckman, 1981).

$$w_{it} = \pi_0 + \pi_{dk}dk_{it} + \pi_{sk}sk_{it} + \pi_1 h_{it} + \pi_2 e_{1it} + \alpha_{1i} + \eta_{1it}, \qquad (6.5)$$

where, α_{1i} = random individual-specific effects, i.e. a time-invariant disturbance term, η_{1it} = a time-varying (or idiosyncratic) disturbance term. In the FE model, the time-varying disturbance term η_{1it} is assumed to be uncorrelated to the explanatory variables, while the time-invariant disturbance term α_{1i} is allowed to be correlated with the explanatory variables. Further details relating to the model specification can be found in Section 5.3 in Chapter 5.

This chapter employs the standard Hausman test which has also described in Section 5.3 of Chapter 5, to compare the FE and the RE estimators. Under a strong assumption that RE estimator is fully efficient, the null hypothesis is that the individual effects are uncorrelated with the explanatory variables, i.e. RE is consistent. Although the Hausman test does not compare the pooled OLS and FE directly, pooled OLS estimates will be consistent if the RE model is appropriate but will be inconsistent if the FE model is appropriate. The test statistic has a chi-squared asymptotic distribution with degrees of freedom equal to the number of coefficients.

Based on the above methods, which assume drinking and smoking status are exogenous, this chapter estimates separate models to reflect different measures of drinking and smoking status in terms of frequency and intensity. The first two models (Model 1 and Model 2) include the dummy variable of drinking status (*drink*) and different dummy variables of smoking status (*smokenow* and *eversmoked*, respectively). The third model (Model 3) includes the previously defined drinking frequency variables, and the five dummy variables which reflect the different categories of smoking intensity. The final model (Model 4) includes the two continuous variables of drinking and smoking, plus their square terms to capture the potential inverted- U shape relationship between drinking / smoking and labour market outcomes.

This chapter also estimates three specifications of the models above as sensitivity analyses. According to the literature, controlling for several explanatory variables that might be correlated with drinking or smoking status such as health status, educational attainment, and family formation could affect the estimated coefficients of drinking status on labour market outcomes (Mullahy and Sindelar, 1993; French and Zarkin, 1995; Macdonald and Shields, 2001; Tekin, 2002; and Auld, 2005). This implies that drinking and smoking influence labour market outcomes both directly and indirectly through human capital and family formation. To explore this issue further, and following French and Zarkin (1995), Model I, referred to as 'full effect' models, is re-estimated both for the employment and wage analyses (refer to as 'direct effect' models) with SRHS, education attainment, and marital status variables excluded. Additionally, Models II – IV re-estimate Model 1 with the three groups of variables separately excluded from the analyses. Following Auld (2005), Model V re-estimates Model 1 with either smoking or drinking status excluded from the analyses, to examine the effect of each variable separately on labour market outcomes without controlling for the effect of another. According to the literature, (see for example, Lye and Hirschberg, 2004) there is evidence of an interaction effect of drinking and smoking behaviour on labour market outcomes. As discussed in Section 6.3.1, there is a positive correlation between drinking and smoking status. Accordingly, a final specification considers the interaction effect of drinking and smoking status on labour market outcomes by using four dummy variables, *nodksk* (the reference category), *drinkonly*, *smokeonly* and *drinksmoke* are defined as follows: *nodksk* takes the value of one if individuals neither drink alcohol nor smoke, *drinkonly* refers to drink alcohol but do not smoke, *smokeonly* refers to smoke but do not drink alcohol, and *drinksmoke* refers to both drink alcohol and smoke, respectively.

It is important to note that since the prevalence of drinking and smoking is very low for women, some of the analyses cannot be undertaken on the female sample only due to the lack of sufficient observations. More specifically, for Model 3 of employment equation, the measure of smoking status only includes three dummy variables: *nonsmoke* (the reference category), *smoke1to10*, and *smokemt11*, where *smokemt11* is defined as a value of one if individuals smoke more than 11 cigarettes per day. Women who smoke but having missing values on cigarettes consumption amount (a total of 22 observations) are also excluded from the analysis. When estimating the effect of drinking frequency on wages for women, dummy variables for *drink1to2w*, *drink3to4w* and *drinkdaily* are not used due to the same reason as above. Instead a dummy variable indicating whether or not women consume alcohol more often than once or twice a month, *drinkoften*, is used.

Drinking and smoking may affect labour market outcomes through unobserved factors that are correlated with drinking, smoking and labour market outcomes. Therefore, treating drinking and smoking as exogenous may leads to biased estimates. Moreover, the results are also likely to be biased under the presence of measurement error (French and Zarkin, 1995) and reverse causality (Mullahy and Sindelar, 1996). The existing literature has employed an IV method to address the endogeneity of drinking and/or smoking, such as Mullahy and Sindelar (1996), Hamilton and Hamilton (1997), Macdonald and Shields (2001), Barrett (2002), Lye and Hirschberg (2004), Auld (2005), and Lin and Chen (2010). Both the IV method of Lye and Hirschberg (2004) and of Auld (2005) simultaneously account for the endogeneity of drinking and smoking. However, they are unable to test the validity of instruments in both studies, and the former does not perform an endogeneity test. Other studies that fail to employ the IV method are mainly due to lacks of valid instruments (French and Zarkin, 1995; Zarkin *et al.*, 1998; and Tekin, 2002).

To account for potential endogeneity of drinking and smoking in a binary employment equation, this chapter employs the Newey (1987)'s two-step estimator. Based on Cameron and Trivedi (2009) and StataCorp (2009), the model is as follows:

$$y_{it}^{*} = \vartheta_{end} x_{it}^{end} + \vartheta_2 h_{it} + \vartheta_3 x_{it} + o_{1it} \qquad (6.6)$$

$$x_{it}^{end} = \pi_4 h_{it} + \pi_5 x_{it} + \pi_6 x_{it}^{IV} + \varrho_{1it}, \qquad (6.7)$$

x_{it}^{end} is a $1\times k_1$ vector of endogenous variables including continuous measures of alcohol and cigarette consumption and their square terms, following MacDonald and Shields (2001). x_{it}^{IV} is a $1\times k_2$ vector of instruments that are correlated with x_{it}^{end} but not with y_{it}^{*}. ϑ_{end}, ϑ_2, ϑ_3, and π_4-π_6 are estimated coefficients. o_{1it} and ϱ_{1it} are disturbance terms. Equation (6.6) is referred to as a structural equation. Equation (6.7) is referred to as a first-stage equation or reduced-form equation, which serves as a source of identifying instruments. For multiple endogenous explanatory variables, the first stage of the Newey (1987)'s two-step estimator estimates Equation (6.7) using pooled OLS for each endogenous variable and obtains consistent estimates for the reduced form coefficients, indicated as $\hat{\pi}_4$-$\hat{\pi}_6$.

If o_{1it} and ϱ_{1it} are independent, there is no endogeneity problem. If o_{1it} and ϱ_{1it} are correlated, x_{it}^{end} is endogenous. A Wald test of the null hypothesis H_0: $\lambda_{Newey} = 0$ serves as a test of exogeneity of x_{it}^{end}, is as follows:

$$o_{1it} = \lambda_{Newey}\varrho_{1it} + error_{it} \tag{6.8}$$

where $error_{it}$ is independent of ϱ_{1it}. If the test statistic is not significant, there is no evidence of endogeneity. In this case, the pooled probit regression is appropriate.

It is possible to test the validity of instruments in an over identified model (i.e. there are more instruments then endogenous variables) by using an Amemiya (1978, 1979, and 1983), Lee (1981) and Newey (1987)'s test. The joint null hypothesis is that the instruments are valid (i.e. are uncorrelated with the error term), and that the instruments are correctly excluded from the employment equation. The test statistic is Amemiya-Lee-Newey minimum chi-squared statistic, which has an asymptotic chi-squared distribution with degrees of freedom equal to the number of over identifying restrictions, where the number of over identifying restrictions is the number of instrumental variables minus the number of endogenous variables.

To account for potential endogeneity of drinking and smoking in a wage determination equation, this study employs the 2SLS estimator. The model is as follows:

$$w_{it} = \pi_7 + \pi_8 x_{it}^{end} + \pi_9 h_{it} + \pi_{10} e_{1it} + \varrho_{2it} \tag{6.9}$$

$$x_{it}^{end} = \pi_{11} + \pi_{12} h_{it} + \pi_{13} e_{it} + \pi_{14} x_{it}^{IV} + \varrho_{3it}, \tag{6.10}$$

where, π_7 and π_{11} are constant terms, π_8-π_{10}, and π_{12}-π_{14} are estimated coefficients, ϱ_{2it} and ϱ_{3it} are two disturbance terms, assuming that Cov[$x_{it}^{IV}, x_{it}^{end}$] $\neq 0$, Cov[$x_{it}^{IV}, \varrho_{2it}$] = 0, and Cov[$x_{it}^{IV}, \varrho_{3it}$] = 0.

For multiple endogenous explanatory variables, the first stage of the 2SLS is to estimate Equation (6.10) using pooled OLS for each endogenous variable and obtains consistent estimates for the reduced form coefficients, indicated as $\hat{\pi}_{11}$-$\hat{\pi}_{14}$. With these estimates, predicted values of x_{it}^{end} can be obtained as:

$$x_{it}^{end} = \hat{\pi}_{11} + \hat{\pi}_{12} h_{it} + \hat{\pi}_{13} e_{1it} + \hat{\pi}_{14} x_{it}^{IV} \tag{6.11}$$

The F test is used to test the joint significance of the coefficient on instruments for each endogenous explanatory variable. If the F statistic is significant, the instruments have significant explanatory power for the endogenous explanatory variable. If it is not significant, the instruments have no explanatory power for the endogenous explanatory variable. In the second stage, endogenous explanatory variables are replaced by their predicted values to estimate Equation (6.9) using pooled OLS.

The Hausman test is used to test for endogeneity of the suspected endogenous variables by comparing the pooled OLS and the 2SLS estimates. Under the null hypothesis that the specified endogenous explanatory variables can actually treated as exogenous, the test statistic is distributed as chi-squared with degrees of freedom equal to the number of explanatory variables tested. In addition, the Sargan-Hansen test is used to test for the validity of over identifying instruments, where the null hypothesis is that instruments are valid, and are correctly excluded from the employment equation. A more detailed specification of the Sargan-Hansen test can be found in Section 5.3 in Chapter 5.

Applying an IV method to control for potential endogeneity of drinking and smoking requires instruments that are correlated with drinking and smoking but uncorrelated with labour market outcomes. Most of the existing literature use the price of alcohol and/or cigarettes (Hamilton and Hamilton, 1997; and Auld, 2005), measures of drinking and/or smoking in the area where respondents live (Mullahy and Sindelar, 1996; and Barrett, 2002), measures of drinking and smoking of respondents' family members (Mullahy and Sindelar, 1996; and Macdonald and Shields, 2001), individuals' illness status (Macdonald and Shields, 2001; and Lye and Hirschberg, 2004), individuals' knowledge about the negative effects of drinking and/or smoking on health (Macdonald and Shields, 2001; and Lin and Chen, 2010), and individuals' religious status or country of birth (Lye and Hirschberg, 2004; and Auld, 2005). This chapter is unable to use instruments such as individuals' knowledge about the negative effects of drinking and/or smoking on health or their religious status as these are not surveyed by the CHNS.

This chapter considers a number of alternative variables as potentially suitable instruments for drinking and smoking. Recall from the previous chapter, individuals are asked what was the doctor's diagnosis of your illness during the last month, based on a list of specified illnesses[155]. According the medical literature, drinking and smoking are associated with most of the above illnesses, such as alcohol poisoning, infections (Aldoori *et al.*, 1997; Brenner, *et al.*, 1997; Nuorti *et al.*, 2000; and Davies *et al.*, 2006), heart disease (Djousse *et al.*, 2008), respiratory disease (Doll *et al.*, 2005), haematological disease (Savage and Lindenbaum, 1986), mental/psychiatric disorder (Parrott, 1999; Wetterling and Junghanns, 2000; and Cosci, 2007), digestive disease (Freiberg *et al.*, 2004; and Yoon *et al.*, 2004), sexual dysfunction (Peate, 2005; and Arackal and Benegal, 2007), dermatological disease (Kostovic and Lipozencic, 2004; and Qureshi *et al.*, 2010), rheumatologic disease (Myllykangas-Luosujarvi *et al.*, 2000; and Kallerg *et al.*, 2009). Dummy variables are created indicating presence of either one of the above diseases or a combination of several diseases due to the low incidence of those diseases. However, they cannot be used as instruments since they are found to be correlated with drinking or smoking in the first stage estimation.

Following Mullahy and Sindelar (1996), Macdonald and Shields (2001), and Barrett (2002), this chapter uses six continuous variables measuring the share of the respondents' working-age drinkers with a same type of drinking frequency, based on the five types of drinking frequency (*drink1to1m_hh*, *drink1to2m_hh*, *drink1to2w_hh*, *drink3to4w_hh*, and *drinkdaily_hh*)[156], and two continuous variables measuring the average weekly alcohol and daily cigarettes consumption of the same household members (*dkamt_hh*, and *skamt_hh*), respectively, as instruments for drinking and smoking.[157] The rationale for these instruments is that drinking and smoking behaviour of an individual may be influenced by their family members or their social environment (Barrett, 2002). For example, a higher level of alcohol or cigarette consumption of an individual's household members is likely associated with a higher level of alcohol or cigarette consumption for such individual.

6.4 Empirical Results

[155] The specified illnesses are: infectious/parasitic disease, heart disease, tumour, respiratory disease, injury, alcohol poisoning, endocrine disorder, haematological disease, mental/psychiatric disorder, mental retardation, neurological disorder, eye/ear/nose/throat/teeth disease, digestive disease, urinary disease, sexual dysfunction, obstetrical/gynaecological disease, neonatal disease, dermatological disease, muscular/rheumatologic disease, genetic disease, old age/mid-life syndrome, and other.
[156] The share of type of drinking frequency of the respondents' household members is calculated as: a ratio between the number of a respondent's working-age household members in a specified type of drinking frequency and the total number of the respondent's working-age household members.
[157] Summary statistics of the instruments are presented in Appendix 6.5.

6.4.1 The Impact of Drinking and Smoking Status on Employment - based on the Pooled Probit and the RE Probit Model

The marginal effects of drinking and smoking status on employment based on Models 1 to 4 for the total sample of the Chinese working-age adults are shown in Table 6.2. The results show that the marginal effects of drinking status on employment using the RE probit are lower than those using the pooled probit for all measures of drinking. In contrast, the estimated effects of smoking status using the RE probit are higher than those using the pooled probit. The likelihood-ratio test statistics under RE probit estimates suggest that the null hypotheses that the pooled probit is preferred to the RE probit can be rejected (i.e. p-value = 0.00), therefore the results from the RE probit are preferred to the pooled probit estimator. The marginal effects of drinking and smoking status on employment based on Models 1 to 4 for men and women (and for the young and the older working-age adults) are separately shown in Table 6.3 (and in Table 6.4). For all the sub samples within the population, the null hypotheses can be rejected, thus marginal effects results from RE probit is preferable compared to the results from pooled probit estimator.

The results shown in Table 6.2 indicate that after controlling for the effect of SRHS and other explanatory variables, the statistically significant and positive effect of drinking and smoking status on employment are confirmed across different measures of drinking and smoking. The only exception is the marginal effects of the continuous measure of cigarettes consumption. The estimation results of Model 3 show that there is an (slight) inverse U-shaped relationship between drinking and employment for the Chinese working-age adults and the positive effect peaks at drinking 3 to 4 times per week. This inverse U-shaped effect, which is further confirmed by the results of Model 4, is found be consistent with the medical literature. Nevertheless, it might also be a result of the alcohol's positive social effect.

Smoking is found to be positively correlated with employment, which is inconsistent with the literature. A possible explanation is based on the special Chinese culture in terms of showing respect and hospitality during social interactions, smoking may positively affect employment through a way of maintaining good relations with employer, colleagues, and customers. It is also found that the positive effect on smoking on employment increases as the level of daily cigarettes consumption amount increases. According to the results of Model 4, there is no evidence of an significant relationship between smoking and employment.

Table 6.3 shows that there is a statistically significant and positive effect of drinking status on employment for both genders. Based on the results from Models 3 and 4, an (slight) inverse U-shaped relationship between drinking and employment is found for men but not for women. For women, as shown in marginal effect from Model 3, the positive effect of drinking on employment is estimated at the lowest level of drinking frequency (with a marginal effect of 3.1 percentage points which is significantly different from zero at 5% level). Moreover, women who drink once or twice a week are also more likely to be employed than those who are non-drinkers, with a marginal effect of 2.7 percentage points, which is significantly different from zero at 10% level. The effect of smoking on employment is statistically significant and positive for men but is negative for women. The result in terms of the negative effect of smoking on employment for women is also consistent to the theory. For men, the positive effect of smoking on employment increases as the level of daily cigarettes consumption amount increases. Overall, the statistically significant and positive effect of drinking and smoking status on employment for men, as well as the statistically significant and positive effect of drinking status and the negative effect of smoking status on employment for women are confirmed using the different measures of drinking and smoking. The exception is that there is no statistically significant effect of the continuous measure of cigarettes consumption for men. Not surprisingly, the positive relationship between smoking and employment for men is similar to that for the total sample since 97% of smokers are males.

The results for both genders show some differences compared to the existing literature. The impact of drinking on employment for men and women in China are different to those found in Tekin (2002) for Russia. The author finds that the positive effect of drinking on employment based on the pooled OLS disappears for men and reduces a lot for women after applying the FE linear method. The inverse U-shaped effect also disappears for both genders. Based on the pooled OLS, Lin and Chen (2010) find that smoking is negatively correlated to the probability of being employed for both Taiwanese men and women, and the effect is larger for women than for men.

As shown in Table 6.4, a statistically significant and positive effect of drinking on employment for both the young and the older working-age adults is confirmed across Models 1-4, and the effect is larger for the older than for the young working-age adults. In addition, the inverse U-shaped relationship between drinking and employment is found for both age groups. There is a statistically significant and positive effect of smoking on employment for the young working-age adults, and the effect increases as the level of daily cigarettes consumption amount increases. However, for the older counterparts smoking is not an important predictor of employment, except that smoking more than 20 cigarettes per day are 5.2 percentage points more likely to be employed than those who do not smoke.

The results of the three sensitivity analyses based on the preferred RE probit estimator for the total sample are shown in Table 6.5. Results from Model 1 (which is also known as the 'direct effect' model) in Table 6.5 are the same as to that in Table 6.2. The results of the first sensitivity analysis are presented in columns 2-5 (include Models I to IV). Model I (which is also known as the 'full effect' model) shows the marginal effects of drinking and smoking status on employment without controlling for the effect of SRHS, educational attainment and marital status. Models II to IV show the results by excluding either one of the three variables. The second sensitivity analysis includes Models V and VI, where Model V estimates the marginal effects of drinking status on employment by excluding the effect of smoking, and Model VI estimates the marginal effects of smoking when drinking variable is omitted. The third sensitivity analysis, Model VII, examines the interaction effect of drinking and smoking status on employment. The results of the same sensitivity analyses for the sub samples within the population are shown in Appendix 6.6.

Model 1 in Table 6.5 shows that individuals who drink alcohol are 2.6 percentage points more likely to be employed than individuals who do not. Smokers are 1.2 percentage points more likely to be employed than nonsmokers. Omitting the SRHS, controls for education and marital status increase the effect of drinking and decrease the effect of smoking, although the changes are modest. This finding is consistent to that in Macdonald and Shields (2001), Tekin (2002), and Auld (2005). Similar conclusions are made based on the results shown in Models II to IV. Moreover, based on result of the first sensitivity analysis in Appendix 6.6, there is no evidence of indirect effects of drinking or smoking on employment for the sub samples.

According to the results from the second sensitivity analysis in Models V-VI, excluding smoking / drinking status has a limited effect on the marginal effects of drinking status / smoking status on employment, which is consistent with Auld (2005). The same conclusion is also made for the sub samples. For example, when smoking status is omitted, the estimated effect of drinking for men increases from 2.0 percentage points to 2.3 percentage points. When drinking is omitted, the estimated effect of smoking for men increases from 1.1 percentage points to 1.6 percentage points.

As shown in Model VII, for the total sample, individuals who drink alcohol but do not smoke and who smoke but do no drink alcohol are 3.2 and 2.0 percentage points more likely to be employed than those who neither drink alcohol nor smoke, respectively. The estimated result of individuals who both drink alcohol and smoke has the largest effect among the three groups, with a value of 3.8 percentage points. However, this result is equal to the sum of the other two from Model 1 (i.e. 2.6 +1.2 percentage points = 3.8 percentage points), suggesting that there is no evidence on the interaction effect of drinking and smoking on employment. Based on result of the third sensitivity analysis in Appendix 6.6, the same conclusion is made for the sub samples.

6.4.2 The Impact of Drinking and Smoking Status on Wages – based on the Pooled OLS and the FE Model

The pooled OLS and the FE estimation results of the effect of drinking and smoking status on hourly wages for the total sample of Chinese workers based on Models 1 to 4 are presented in Table 6.6. The Hausman test statistics strongly reject the null hypothesis that RE provides consistent estimates, thus FE is preferable to pooled OLS. The estimation results based on Models 1 to 4 for men and women (and for the young and the older working-age adults) are separately shown in Table 6.7 (and in Table 6.8). The Hausman test statistics for all the sub samples also strongly reject the null hypothesis, which implies that FE is the preferred model.

Table 6.6 shows that there is no statistically significant effect of drinking or smoking on wages for the Chinese workers. This result is robust to different measures of drinking and smoking status. Furthermore, it is not sensitive to the models. The only exception is that in Model 3 daily drinkers tend to earn 3.7% more than non-drinkers. However, once the individual unobservable factors are controlled for, the statistically significant effect disappears. Based on the OLS approach, French and Zarkin (1995) find that there is a positive and inverse U-shaped relationship between drinking and wages for the American workers. However, the result obtained in this chapter suggests that drinking or smoking is not an important predictor of wages for the Chinese workers.

As shown in Table 6.7, there is no statistically significant relationship between drinking alcohol and wages for either gender. For women, the pooled OLS estimates show a positive and inverse U-shaped relationship between drinking and wages. Again, the statistically significant effects disappear once the individual unobserved factors are controlled for. There is also no statistically significant relationship between smoking and wages for either gender. The results for both genders are consistent with those in Neumann (2009). The results for women are also consistent with those in Lin and Chen (2010), Van Ours (2002), Zarkin *et al.* (1998), and Heineck and Schwarze (2003), although results for men from these studies are mixed. For example, Lin and Chen (2010) find a negative effect of smoking on earnings for Taiwanese males. Van Ours (2002) applies the analysis of starting rates for drinking and smoking to account for the unobservable factors and finds that OLS overestimates the positive effect of drinking and underestimates the negative effect of smoking on wages for men. Zarkin *et al.* (1998) use the pooled OLS approach and find that there is statistically significant and positive relationship between drinking for American men. Heineck and Schwarze (2003) use the same methods as used in this chapter and find that smoking is positively correlated to male workers.

Table 6.8 shows that there is no statistically significant relationship between drinking and wages for either the young or older working-age workers. For the young working-age workers, smoking does not have a statistically significant effect on wages. For the older working-age workers, smoking is positively correlated with wages after controlling for individual unobservable factors. According the FE estimates in Models 1 and 3, current older working-age smokers tend to earn 9% more than those who do not smoke. The FE estimates in Model 2 show that individuals who have ever smoked tend to earn about 10% more than those who have never smoked.

The results of the three sensitivity analyses based on the FE (the preferred model) for all the sample groups within the population are presented in Appendix 6.7. It shows that there is no evidence of the indirect effect of drinking or smoking status on wages for all the sample groups. Moreover, excluding smoking / drinking from the model does not affect the effect of drinking / smoking on wages. Based on the sample of Canadian men, Auld (2005) finds that failure to control for drinking underestimates the negative effect of smoking on income, and failure to control for smoking also affects the estimated result of the drinking variables in different directions. There is also no evidence of the interaction effect of drinking and smoking on wages for all the sample groups, which is found to be different to that in Lye and Hirschberg (2004). The authors investigate the interaction effect of smoking with drinking alcohol by estimating the effect of drinking status on income separately for Australian male smokers and non-smokers. They find that there is a positive and inverse U-shaped relationship between alcohol consumption and income for non-smokers but not for smokers.

A concern which might be raised in the analysis is to correct for the sample selection bias in the wage model. Tekin (2002) argues that if the selection is caused by unobserved time-invariant factors, the FE will control for such selection and there is no need to correct for selectivity. However, the last chapter applies the Heckman (1979)'s selection model to correct for the sample selection bias and finds a considerable effect of the bias in estimating the effect of SRHS on wages. After correcting for the sample selection bias, the statistically significant and positive effect between SRHS and wages disappears. The author also applies the same model to correct for the sample selection bias in this chapter. However, the author dose not reported the result based on the Heckman (1979)'s selection model, since it is consistent with the result based on the FE, implying that there is no effect of drinking or smoking on wages for the Chinese workers.

6.4.3 Addressing the Potential Endogeneity of Drinking and Smoking – IV Models

The above analyses assume that drinking and smoking are exogenous. In accounting for their potential endogeneity in the employment and wage analysis this chapter applies the Newey (1987)'s two-step and the 2SLS estimator, respectively. The coefficient estimates based on the second stage regression of the Newey (1987)'s two-step estimator for all the sample groups are shown in the top panel of Table 6.9. In addition, the over identification test statistic (the Amemiya-Lee-Newey minimum chi-squared statistic), and the Wald test statistic for the endogeneity of drinking and smoking are also presented. For comparison, the middle and bottom panel of Table 6.9 present the estimation results of marginal effects based on the pooled probit and the RE probit, respectively, where drinking and smoking are treated as exogenous. Estimation results from the first stage of the Newey (1987)'s two-step estimation are shown in Table 6.10.

As shown in Table 6.9, the insignificance of the Amemiya-Lee-Newey minimum chi-squared statistic for all the sample groups suggests that the instruments are correctly excluded from the employment equation. For women, the Wald test statistic reject the null hypothesis of exogeneity of drinking and smoking, suggesting that Newey (1987)'s two-step estimation is more appropriate than the the pooled probit estimation. For the other sample groups, the null hypothesis of exogeneity of drinking and smoking cannot be rejected, suggesting that there is no endogeneity problem and the pooled probit estimation is more appropriate than the Newey (1987)'s two-step estimation. Accordingly, previous conclusion that RE probit is the preferred model holds.

The Newey's (1987) two-step estimates show that there is a positive and inverse U-shaped relationship between drinking and employment for women, and the positive effect is larger for women (with a marginal effect[158] of 6.4 percentage points which is significantly different from zero at 10% level) comparing to the preferred RE probit estimates (with a marginal effect of 0.1 percentage points which is significantly

[158] The Newey's two-step estimation is performed by Stata. However, unlike the pooled probit or the RE probit, the postestimation such as computing marginal effect with command 'margins' are not available after the Newey's two-step estimation. Therefore, the author computes the marginal effect after the Newey's two-step estimation manually.

different from zero at 1% level) for men. In addition, a larger effect of drinking on employment for the young than for the older working-age sample, which is found in Table 6.9, is consistent with that found in Table 6.4. There is no statistically significant relationship between smoking and employment for the total sample or the sub samples within the population. For women, after controlling for the endogeneity of smoking, its negative effect on employment no longer exists, which suggests that smoking negatively affect employment through unobserved factors, thus treating smoking as exogenous would lead to biased estimate.

The effect of the instruments on individual alcohol and cigarette consumption after controlling the effect of SHRS and other explanatory variables are presented in the top and the bottom panel of Table 6.10, respectively. Recall from the last section, a group of five variables, *drink1to1m_hh*, *drink1to2m_hh*, *drink1to2w_hh*, *drink3to4w_hh*, and *drinkdaily_hh* are defined as the share of the respondents' working-age household members with one out of five types of drinking frequency, variables *dkamt_hh*, and *skamt_hh* are defined as the average amount of alcohol and cigarettes consumption of the same household members, respectively.

For the total sample, the share of all types drinking frequency are found to be negatively correlated with individual alcohol consumption. A similar result is also found for the sub samples. Moreover, for all the sample groups, the effect of these instruments on individual cigarettes consumption shown in the bottom panel of Table 6.10 is similar to that on individual alcohol consumption found in the top panel of the table.

The top panel of Table 6.10 also shows that the average amount of alcohol consumption of the respondents' household members is found to be positively correlated with individual's alcohol consumption for all the sample groups. The result of the effect of *skamt_hh* in the smoking equation yields a similar conclusion. However, there is a statistically significant and negative relationship between average amount of cigarettes consumption of the respondents' household members and individual alcohol consumption. The average amount of alcohol consumption of the respondents' household members is found to be positively correlated with individual's cigarettes consumption for the total and the young working-age sample.

However, it is negatively affect individual's cigarettes consumption for men and the older working-age sample

Table 6.11 shows the estimation results of the effect of drinking and smoking on wages for all the sample groups when taking into account the potential endogeneity of drinking and smoking. The top panel of the table shows the results of the second stage of the 2SLS. In addition, the over identification test statistic (the Hansen J statistic), and the Hausman test statistic for the endogeneity of drinking and smoking are also presented. The middle and bottom panel of Table 6.11 show the estimation results based on the pooled OLS and the FE method, where drinking and smoking status are treated as exogenous. Estimation results from the first stage of the 2SLS estimation, along with the F statistic for the joint significance of the coefficients on the instruments, are shown in Appendix 6.8.

The insignificance of the Hansen J statistics suggests that the instruments are correctly excluded from the wage equation. In addition, the F statistics shown in Appendix 6.8 are significant at the 1% level for all the sample groups, indicating that the instruments have significant explanatory power for the potentially endogenous variables after controlling for the other explanatory variables. Moreover, the estimation results from the first stage of the 2SLS are consistent with that from the first stage of the Newey (1987)'s two-step estimator. As shown in Table 6.11, for all the sample groups, the null hypothesis of exogeneity of drinking and smoking cannot be rejected since the Hausman test statistics shown in the top panel of the table are not statistically significant, suggesting that the pooled OLS estimation is more appropriate than the 2SLS estimation. Accordingly, the previous conclusion that FE is the preferred method is also made in this chapter.

For all the sample groups, the pooled OLS and the FE estimation results shown in Table 6.11 are consistent with those in Tables 6.6-6.8. In addition, the finding that there is no statistically significant effect of drinking or smoking intensity on wages is not sensitive whether drinking and smoking are treated as endogenous or exogenous, which is found to be different to that in Auld (2005). Auld (2005) finds treating smoking as endogenous increases the negative effect of smoking status on annual wage income based on the sample of Canadian men. The only exception is that the positive and inverse U-shaped relationship between drinking intensity and wages for women disappears once controlling for the individual unobserved factors.

6.5 Conclusions

This chapter contributes to the literature which examines the simultaneous effect of drinking and smoking on employment and wages by providing new evidence on the relationship between health-related behaviours and labour market outcomes for the Chinese working-age population. It also helps to produce evidence in achieving a balance between economic development and public health, and between short-term and long-term alcohol and tobacco control strategies.

The chapter employs panel data methods, which include the RE probit and FE in estimating the effect of drinking and smoking behaviour on employment and wages, respectively, to allowing for the panel nature of the data. Furthermore, the Newey (1978)'s two-step and the 2SLS estimator in used to account for the potential endogeneity of drinking and smoking in the employment and wage analysis, respectively. The share of the respondents' working-age household members with the same type of drinking frequency and the average amount of alcohol and cigarettes consumption of the respondents' working-age household members are used as instruments of drinking and smoking. The test of instruments validity shows that the instruments are correctly excluded from both employment and wage equation. However, the endogeneity test shows that drinking and smoking are exogenous to wages. There is also no evidence of endogeneity of drinking or smoking in the employment analysis for all the sample groups except for women.

There is a positive and inverse U-shaped relationship between drinking and employment for the Chinese working-age population. This finding is robust across different ways measuring drinking behaviour, which is also consistent with the medical literature. Smoking incidence is found to be positively correlated with employment, whereas there is no statistically significant effect when using average daily cigarettes consumption is used as a measure of smoking. There is no statistically significant effect of drinking and smoking on hourly wages for the Chinese working-age population, and the result is not sensitive across methods or different measures of drinking or smoking behaviour. The results in terms of the effect of smoking on labour market outcomes are found to be different from most of the existing literature. For the Chinese working-age population, the unique cultural background on smoking behaviour is in favour of labour supply, whereas is not much related to the labour productivity.

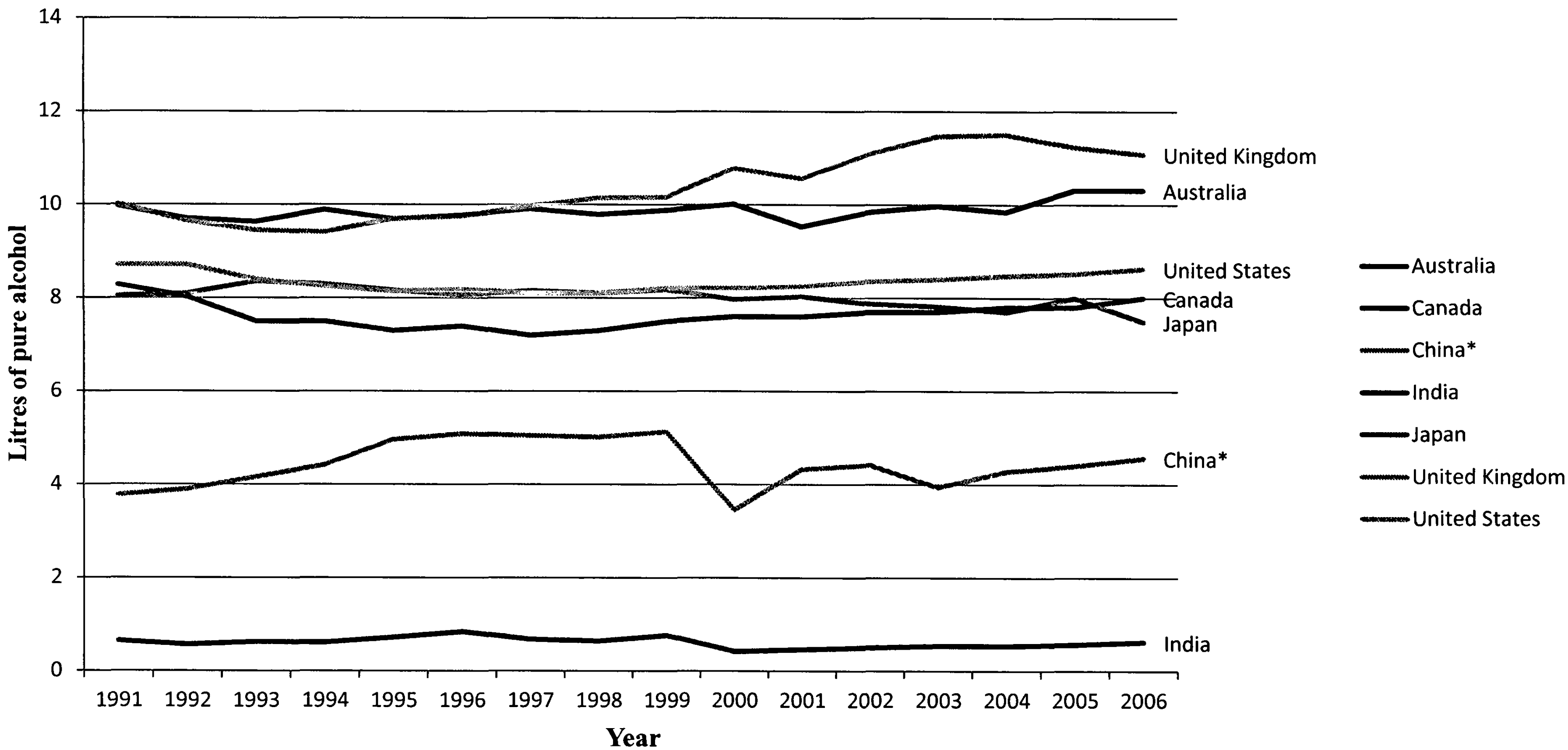
Most of the studies consider on the gender differences in estimating the effect of drinking and/or smoking behaviour on the labour market outcomes, although the results are somewhat mixed. In this chapter, analysis is also carried out for sub samples within the population, which allows us to explore the differences across genders and age groups. There is evidence of a positive for both genders, and the effect is larger for women than for men. Additionally, there is an inverse U-shaped relationship between drinking and employment for men but not for women. For men, the smoking incidence is found to be positively correlated with employment whereas average daily cigarettes consumption is not. For women, after controlling for the endogeneity of drinking and smoking, the negative and U-shaped relationship between smoking intensity and employment disappears. There is no statistically significant effect of drinking and smoking on wages for both genders. For men, the result is not sensitive across models or different measures of drinking and smoking behaviours. For women, the positive and inverse U-shaped relationship between drinking and employment is driven by individual unobserved factors.

There is a positive and inverse U-shaped relationship between drinking and employment for both the young and the older working-age adults, and the effect is larger for the later than for the former. For the young working-age adults, smoking incidence is found to be positively correlated with employment whereas average daily cigarettes consumption is not. For the older working-age adults, only those who smoke more than 20 cigarettes per day are found to be positively correlated with employment. There is no statistically significant effect of drinking on wages for both sample groups, and the result is not sensitive across models or different measure of drinking behaviour. For the young working-age sample, there is no statistically significant relationship between smoking and wages. For the older counterparts, smoking incidence is found to be positively correlated with wages after controlling for individual unobserved factors.

The results of sensitivity analyses provide little evidence of the indirect effect of drinking or smoking status on labour market outcomes through health status, educational attainment or family formations. The analysis also finds that failure to control for drinking/smoking status has a limited effect on the estimated result of smoking/drinking status in the employment equation. In addition, there is no evidence of the interaction effect of drinking and smoking status on employment for the Chinese working-age adults and all the sub samples except for women.

It is worth noting that the sample used in the wage analysis does not include agricultural workers and the self-employed because most of them report their income from other sources. However, drinking and smoking behaviour may have important impact on their productivity. Therefore, future analysis could focus on this group of population by using a net profit earned from agricultural work and self-employment to measure labour productivity. Moreover, this chapter estimates contemporaneous effect of drinking and smoking (e.g. drinking status in the last year, and current cigarette consumption) on labour market outcomes (e.g. current employment status, and averaged monthly wages in the last year). However, drinking and smoking behaviour might not likely to affect or may have a limited effect on labour market outcomes in a short term period (Mullahy and Sindelar, 1996). In other words, the effect may occur with a lag. Therefore, it would be reasonable to explore the dynamic effect of health related behaviour on labour market outcomes.

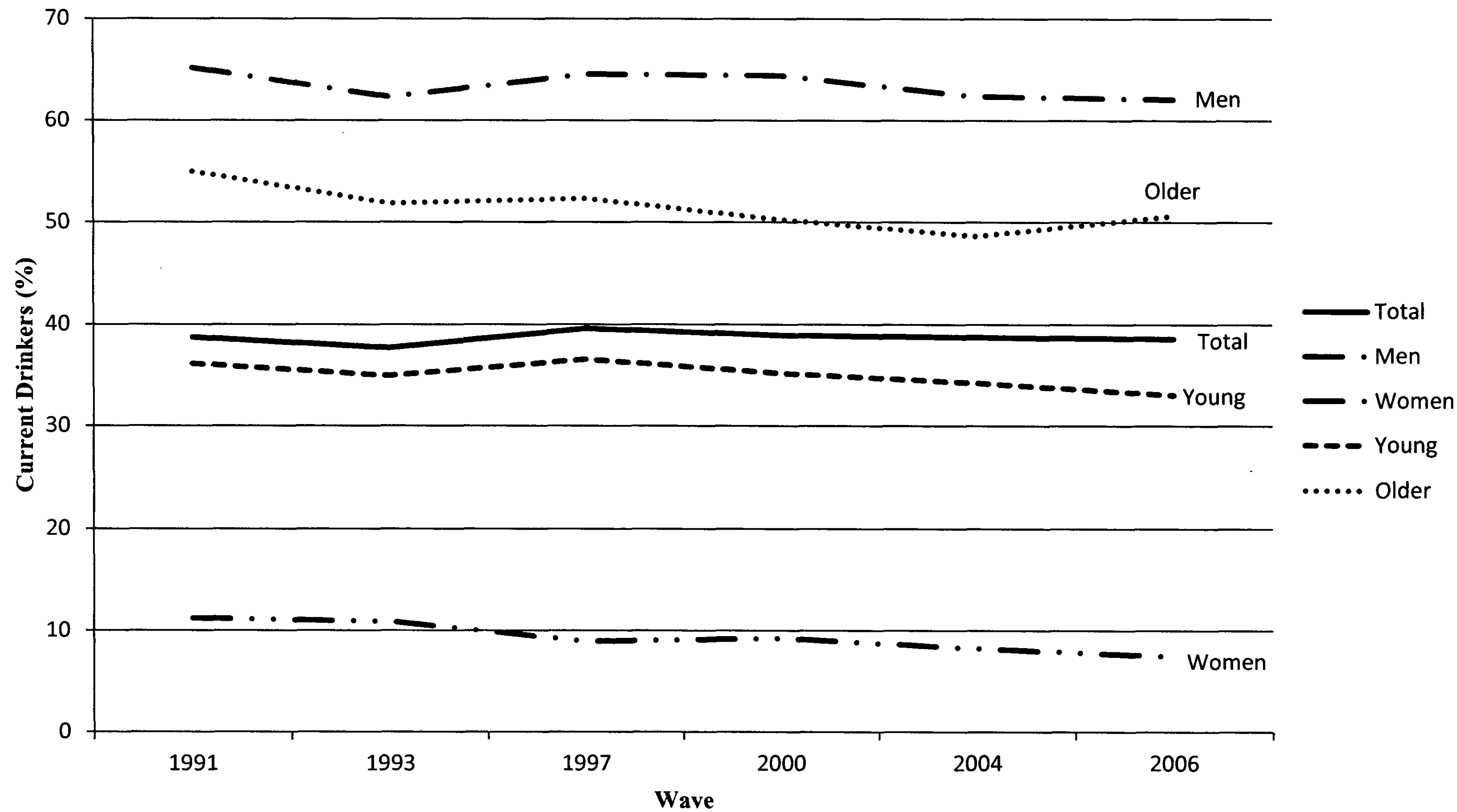
Figure 6.1 Adult (age 15+) per Capita Alcohol Consumption by Country, 1991-2006



Source: http://apps.who.int/gho/data/node.main.A1025?lang=en. [cited 16th Apr 2013].

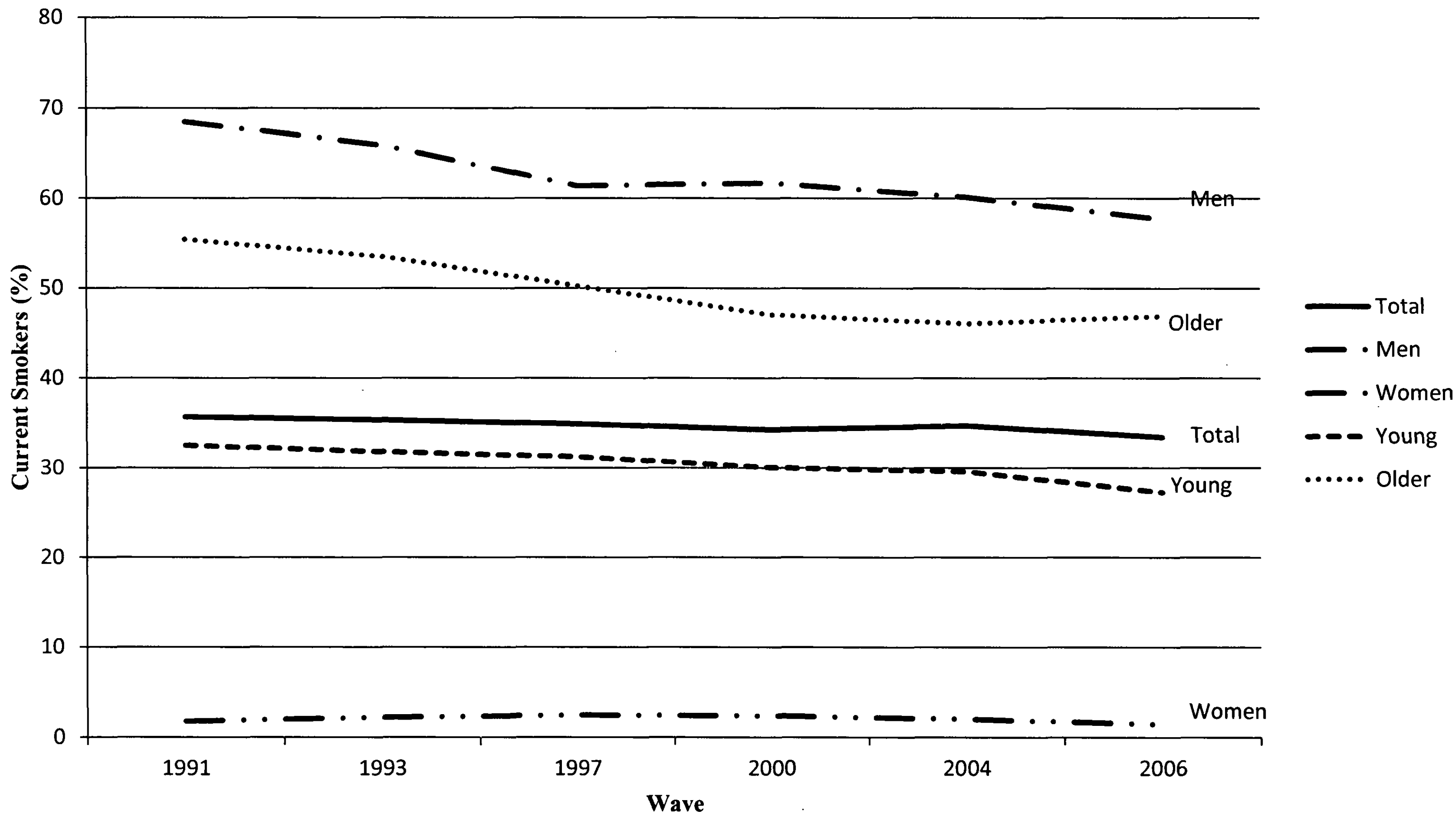
* Note: Data reported from 2000 onwards have been calculated using a different source. Therefore the decline between 1999 and 2000 may not necessarily reflect a real decline in alcohol consumption.

Figure 6.2 Prevalence of Drinking Alcohol by Wave



Source: Author's calculations using CHNS data.

Figure 6.3 Prevalence of Smoking by Wave



Source: Author's calculations using CHNS data.

Table 6.1a Employment Rates by Drinking and Smoking Status

	Employment Rate (%)				
Drinking Status	**Total (N = 37,591)**	**Men (N = 20,345)**	**Women (N = 17,246)**	**Young (aged 16-45) (N = 29,126)**	**Older (aged 46-59) (N =8,465)**
Non-Drinkers **(Sample Size: %)**	85% (61%)	87% (36%)	84% (90%)	86% (65%)	78% (49%)
Drinkers **(Sample Size: %)**	91% (39%)	92% (64%)	89% (10%)	93% (35%)	87% (51%)
Drinkers (*drink0to1m*) **(Sample Size: %)**	91% (4%)	92% (5%)	90% (1%)	91% (8%)	91% (20%)
Drinkers (*drink1to2m*) **(Sample Size: %)**	91% (8%)	92% (12%)	87% (1%)	92% (5%)	89% (7%)
Drinkers (*drink1to2w*) **(Sample Size: %)**	92% (11%)	92% (18%)	89% (2%)	93% (10%)	88% (12%)
Drinkers (*drink3to4w*) **(Sample Size: %)**	93% (6%)	93% (10%)	89% (3%)	93% (8%)	91% (8%)
Drinkers (*drinkdaily*) **(Sample Size: %)**	90% (10%)	91% (19%)	86% (3%)	95% (4%)	85% (4%)
Smoking Status					
Non-Smokers (lifetime) **(Sample Size: %)**	85% (64%)	88% (35%)	84% (98%)	87% (69%)	79% (48%)
Smokers (who have ever smoked) **(Sample Size: %)**	91% (36%)	91% (65%)	79% (2%)	93% (31%)	86% (52%)
Non-Smokers (current) **(Sample Size: %)**	85% (65%)	88% (38%)	84% (98%)	87% (69%)	79% (51%)
Smokers (current) **(Sample Size: %)**	91% (35%)	91% (62%)	79% (2%)	93% (31%)	87% (49%)
Smokers (*smoke1to10*) **(Sample Size: %)**	91% (12%)	91% (22%)	79% (1%)	92% (11%)	86% (16%)
Smokers (*smoke11to20*) **(Sample Size: %)**	91% (18%)	92% (33%)	76% (1%)	94% (16%)	87% (26%)
Smokers (*smokemt20*) **(Sample Size: %)**	93% (4%)	92% (6%)	100% (< 1%)[1]	95% (3%)	89% (6%)
Smokers (*smokemiss*) **(Sample Size: %)**	85% (1%)	86% (1%)	83% (< 1%)[2]	89% (1%)	79% (1%)
	Mean (S.D.)				
Weekly Average Alcohol Consumption	**Total**	**Men**	**Women**	**Young (aged 16-45)**	**Older (aged 46-59)**

Amount (units), CHNS 1993 - 2006:	(N = 30,015)	(N = 16,189)	(N = 13,826)	(N = 22,723)	(N = 7,292)
wkdrinkamt	4.20 (10.32)	7.47 (12.94)	0.37 (2.85)	3.34 (8.89)	6.90 (13.51)
wkdrinkamt (without non-drinkers)[3]	11.38 (14.38)	12.09 (14.66)	4.84 (9.11)	10.10 (13.08)	14.06 (16.47)
Daily Average Cigarette Consumption (no. of cigarettes), CHNS 1993 - 2006:					
dailysmokeamt	5.55 (9.27)	10.09 (10.55)	0.24 (1.94)	4.73 (8.67)	8.12 (10.52)
dailysmokeamt (without non-smokers)[4]	16.49 (8.65)	16.61 (8.66)	11.93 (7.02)[5]	16.12 (8.52)	17.20 (8.86)

Source: Author's calculations using CHNS data.
[1] Small sample size: N = 10.
[2] Small sample size: N = 12.
[3] The weekly average alcohol consumption for individuals who drink alcohol.
[4] The daily average cigarette consumption for individuals who are current smokers.
[5] Small sample size: N = 275.

Table 6.1b Relationship between Drinking and Smoking Status (The Chinese Working-age Adults: N = 37,591)

Smoking Status / Drinking Status		*smokenow* = 0	*smokenow* = 1	Total
drink =	0	51%	10% (28%)	61%
	1	(35%) 14%	(65%) 25% (72%)	(100%) 39%
Total		65%	35% (100%)	100%
The Correlation between *drink* and *smokenow (eversmoked)*:			0.50 (0.51)	

Source: Author's calculations using CHNS data.

Table 6.1c Average Real Hourly Wages by Drinking and Smoking Status

	Average Real Hourly Wages (Chinese yuan)				
Drinking Status	**Total** (N = 10,871)	**Men** (N = 6,364)	**Women** (N = 4,507)	**Young** (N = 29,126)	**Older** (N =8,465)
Non-Drinkers **(Sample Size: %)**	2.98 (55%)	3.44 (32%)	2.75 (87%)	2.86 (58%)	3.71 (41%)
Drinkers **(Sample Size: %)**	3.52 (45%)	3.59 (68%)	3.02 (13%)	3.26 (42%)	4.25 (59%)
Drinkers (*drink0to1m*) **(Sample Size: %)**	3.00 (5%)	3.15 (5%)	2.79 (5%)	2.72 (5%)	4.10 (5%)
Drinkers (*drink1to2m*) **(Sample Size: %)**	3.40 (9%)	3.48 (12%)	3.01 (4%)	3.16 (8%)	4.22 (10%)
Drinkers (*drink1to2w*) **(Sample Size: %)**	3.66 (13%)	3.65 (21%)	3.79 (2%)	3.48 (13%)	4.32 (14%)
Drinkers (*drink3to4w*) **(Sample Size: %)**	3.52 (7%)	3.60 (12%)	2.00 (1%)	3.22 (7%)	4.47 (9%)
Drinkers (*drinkdaily*) **(Sample Size: %)**	3.67 (11%)	3.70 (18%)	3.12 (1%)	3.38 (9%)	4.16 (21%)
Smoking Status					
Non-Smokers (lifetime) **(Sample Size: %)**	3.08 (62%)	3.64 (37%)	2.79 (99%)	2.94 (66%)	3.87 (47%)
Smokers (who have smoked) **(Sample Size: %)**	3.47 (38%)	3.48 (63%)	2.16 (1%)[1]	3.19 (34%)	4.16 (53%)
Non-Smokers (current) **(Sample Size: %)**	3.14 (64%)	3.77 (39%)	2.79 (99%)	2.98 (67%)	4.02 (50%)
Smokers (current) **(Sample Size: %)**	3.38 (36%)	3.39 (61%)	1.99 (1%)[2]	3.13 (33%)	4.03 (50%)
	Mean (S.D.)				
Weekly Average Alcohol Consumption Amount (units), CHNS 1993 - 2006:	**Total (N = 8,396)**	**Men (N = 4,927)**	**Women (N = 3,469)**	**Young (aged 16-45) (N = 6,603)**	**Older (aged 46-59) (N = 1,793)**
wkdrinkamt	4.68 (10.43)	7.60 (12.48)	0.52 (3.55)	4.03 (9.62)	7.05 (12.69)
wkdrinkamt (without non-drinkers)[3]	10.71 (13.58)	11.39 (13.79)	4.78 (9.76)	10.06 (13.06)	12.39 (14.73)
Daily Average Cigarette Consumption (no. of cigarettes) , CHNS 1993 - 2006:					
dailysmokeamt	5.44 (8.84)	9.20 (9.89)	0.10 (1.25)	4.70 (8.26)	8.13 (10.25)
dailysmokeamt (without non-smokers)[4]	15.37 (8.25)	15.41 (8.25)	11.93 (7.38)[5]	14.86 (8.05)	16.59 (8.60)

Source: Author's calculations using CHNS data.

[1] Small sample size: N = 41. [2] Small sample size: N = 39. [3] See footnote 3 in Table 7.1a. [4] See footnote 4 in Table 7.1a.[5] Small sample size: N = 28.

Table 6.2 The Impact of Drinking and Smoking on Employment: Pooled Probit and RE Probit (The Total Sample)

	(Model 1)		(Model 2)		(Model 3)		(Model 4)	
Variables	Pooled probit	RE probit	Pooled probit	RE probit	Pooled probit	RE probit	Pooled probit	RE probit
				The Total Sample				
drink	0.0309***	0.0263***	0.0313***	0.0267***				
	(0.0041)	(0.0036)	(0.0041)	(0.0036)				
drink0to1m					0.0289***	0.0215***		
					(0.0082)	(0.0072)		
drink1to2m					0.0289***	0.0249***		
					(0.0064)	(0.0055)		
drink1to2w					0.0346***	0.0285***		
					(0.0058)	(0.0050)		
drink3to4w					0.0377***	0.0317***		
					(0.0073)	(0.0062)		
drinkdaily					0.0243***	0.0243***		
					(0.0064)	(0.0055)		
wkdrinkamt							0.0020***	0.0019***
							(0.0005)	(0.0004)
$(wkdrinkamt)^2/100$							-0.0023***	-0.0021***
							(0.0007)	(0.0006)
smokenow	0.0089*	0.0123***						
	(0.0049)	(0.0042)						
eversmoked			0.0066	0.0099**				
			(0.0050)	(0.0043)				
smoke1to10					-0.0016	0.0044		
					(0.0063)	(0.0055)		
smoke11to20					0.0124**	0.0150***		
					(0.0057)	(0.0049)		
smokemt20					0.0402***	0.0335***		
					(0.0092)	(0.0073)		
smokemiss					-0.0304	-0.0099		
					(0.0201)	(0.0176)		
dailysmokeamt							0.0005	0.0009
							(0.0007)	(0.0006)
$(dailysmokeamt)^2/100$							0.0018	0.0007
							(0.0022)	(0.0019)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.1721		0.1720		0.1729		0.1569	
Likelihood ratio test (*p*-value)		967.71 (0.00)		966.44 (0.00)		958.10 (0.00)		830.98 (0.00)
Observations	37,591	37,591	37,591	37,591	37,591	37,591	30,015	30,015

Note: Only estimation results for marginal effects are reported. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. Under the RE probit analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] Other explanatory variables are: gender, age, educational attainment, marital status, urban/rural, region of residence, presence of children in the household, and year.

Table 6.3 The Impact of Drinking and Smoking on Employment based the Preferred Model: RE probit (Men vs. Women)

	(Model 1)		(Model 2)		(Model 3)		(Model 4)	
Variables	Pooled probit	RE probit	Pooled probit	RE probit	Pooled probit	RE probit	Pooled probit	RE probit
Men								
drink	0.0238*** (0.0044)	0.0202*** (0.0038)	0.0241*** (0.0044)	0.0205*** (0.0038)				
drink0to1m					0.0261*** (0.0091)	0.0170** (0.0079)		
drink1to2m					0.0277*** (0.0064)	0.0224*** (0.0054)		
drink1to2w					0.0266*** (0.0057)	0.0217*** (0.0049)		
drink3to4w					0.0268*** (0.0071)	0.0229*** (0.0059)		
drinkdaily					0.0136** (0.0063)	0.0151*** (0.0054)		
wkdrinkamt							0.0011*** (0.0004)	0.0011*** (0.0004)
$(\text{wkdrinkamt})^2/100$							-0.0014** (0.0006)	-0.0014*** (0.0005)
smokenow	0.0077* (0.0044)	0.0109*** (0.0038)						
eversmoked			0.0062 (0.0044)	0.0094** (0.0038)				
smoke1to10					0.0028 (0.0055)	0.0071 (0.0048)		
smoke11to20					0.0096* (0.0051)	0.0124*** (0.0043)		
smokemt20					0.0298*** (0.0083)	0.0253*** (0.0064)		
smokemiss					-0.0301 (0.0185)	-0.0103 (0.0153)		
dailysmokeamt							0.0005 (0.0006)	0.0007 (0.0005)
$(\text{dailysmokeamt})^2/100$							0.0008 (0.0019)	0.0001 (0.0015)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.1991		0.1990		0.2005		0.1777	
Likelihood ratio test (*p*-value)		415.37 (0.00)		414.22 (0.00)		406.70 (0.00)		332.70 (0.00)
Observations	20,345	20,345	20,345	20,345	20,345	20,345	16,189	16,189
Women								
drink	0.0290*** (0.0086)	0.0270*** (0.0080)	0.0289*** (0.0086)	0.0268*** (0.0080)				
drink0to1m					0.0340** (0.0148)	0.0307** (0.0131)		
drink1to2m					0.0197 (0.0155)	0.0202 (0.0142)		
drink1to2w					0.0338* (0.0175)	0.0266* (0.0158)		
drink3to4w					0.0441 (0.0284)	0.0369 (0.0255)		
drinkdaily					0.0167 (0.0247)	0.0270 (0.0232)		
wkdrinkamt							0.0036 (0.0024)	0.0029 (0.0024)
$(\text{wkdrinkamt})^2/100$							-0.0036 (0.0041)	-0.0021 (0.0042)
smokenow	-0.0544** (0.0212)	-0.0493** (0.0219)						
eversmoked			-0.0527** (0.0210)	-0.0465** (0.0214)				
smoke1to10					-0.0633** (0.0272)	-0.0563** (0.0273)		
smokemt11					-0.0480 (0.0354)	-0.0488 (0.0351)		
dailysmokeamt							-0.0115** (0.0050)	-0.0092* (0.0052)
$(\text{dailysmokeamt})^2/100$							0.0522** (0.0252)	0.0423 (0.0272)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[2] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.1552		0.1552		0.1552		0.1433	

Likelihood ratio test (*p*-value)		487.12 (0.00)		486.96 (0.00)		490.01 (0.00)		440.21 (0.00)
Observations	17,246	17,246	17,246	17,246	17,234	17,234	13,826	13,826

Note: Only estimation results for marginal effects are reported. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. Under the RE probit analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1-2] Other explanatory variables are: age, educational attainment, marital status, urban/rural, region of residence, presence of children in the household, and year.

Table 6.4 The Impact of Drinking and Smoking on Employment based the Preferred Model: RE probit (Young vs. Older)

	(Model 1)		(Model 2)		(Model 3)		(Model 4)	
Variables	Pooled probit	RE probit	Pooled probit	RE probit	Pooled probit	RE probit	Pooled probit	RE probit
				Young (aged 16-45)				
drink	0.0277***	0.0221***	0.0276***	0.0219***				
	(0.0045)	(0.0038)	(0.0046)	(0.0038)				
drink0to1m					0.0118	0.0066		
					(0.0092)	(0.0081)		
drink1to2m					0.0219***	0.0183***		
					(0.0071)	(0.0059)		
drink1to2w					0.0324***	0.0261***		
					(0.0064)	(0.0053)		
drink3to4w					0.0276***	0.0247***		
					(0.0085)	(0.0069)		
drinkdaily					0.0444***	0.0358***		
					(0.0072)	(0.0058)		
wkdrinkamt							0.0029***	0.0025***
							(0.0006)	(0.0006)
$(wkdrinkamt)^2/100$							-0.0030***	-0.0026***
							(0.0010)	(0.0009)
smokenow	0.0128**	0.0133***						
	(0.0056)	(0.0046)						
eversmoked			0.0127**	0.0138***				
			(0.0057)	(0.0046)				
smoke1to10					-0.0006	0.0030		
					(0.0073)	(0.0062)		
smoke11to20					0.0164**	0.0159***		
					(0.0066)	(0.0054)		
smokemt20					0.0420***	0.0319***		
					(0.0110)	(0.0086)		
smokemiss					-0.0191	0.0035		
					(0.0241)	(0.0190)		
dailysmokeamt							0.0008	0.0010
							(0.0008)	(0.0007)
$(dailysmokeamt)^2/100$							0.0017	0.0005
							(0.0028)	(0.0024)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.1626		0.1626		0.1640		0.1512	
Likelihood ratio test (*p*-value)		611.74 (0.00)		612.37		605.91		471.39
Observations	29,126	29,126	29,126	29,126	29,126	29,126	22,723	22,723
				Older (aged 46-59)				
drink	0.0450*** (0.0093)	0.0442*** (0.0091)	0.0463*** (0.0093)	0.0459*** (0.0091)				
drink0to1m					0.0929*** (0.0174)	0.0779*** (0.0151)		
drink1to2m					0.0583*** (0.0145)	0.0548*** (0.0134)		
drink1to2w					0.0434*** (0.0134)	0.0384*** (0.0124)		
drink3to4w					0.0718*** (0.0152)	0.0613*** (0.0139)		
drinkdaily					0.0152 (0.0126)	0.0211* (0.0122)		
wkdrinkamt							0.0014* (0.0008)	0.0016** (0.0008)
$(\text{wkdrinkamt})^2/100$							-0.0018 (0.0011)	-0.0019* (0.0011)
smokenow	0.0049 (0.0100)	0.0124 (0.0094)						
eversmoked			-0.0014 (0.0101)	0.0032 (0.0097)				
smoke1to10					-0.0045 (0.0130)	0.0056 (0.0125)		
smoke11to20					0.0077 (0.0116)	0.0143 (0.0109)		
smokemt20					0.0482*** (0.0173)	0.0521*** (0.0143)		
smokemiss					-0.0469 (0.0384)	-0.0484 (0.0399)		
dailysmokeamt							0.0003 (0.0013)	0.0007 (0.0011)
$(\text{dailysmokeamt})^2/100$							0.0029 (0.0040)	0.0025 (0.0035)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Other Explanatory Variables[2] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.1862		0.1862		0.1904		0.1621	
Likelihood ratio test (*p*-value)		251.16 (0.00)		249.83 (0.00)		243.02 (0.00)		235.58 (0.00)
Observations	17,246	17,246	17,246	17,246	17,224	17,224	13,826	13,826

Note: Only estimation results for marginal effects are reported. Under pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. Under the RE probit analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1-2] Other explanatory variables are: gender, educational attainment, marital status, urban/rural, region of residence, presence of children in the household, and year.

Table 6.5 Sensitivity Analyses based on the Employment Equation (The Total Sample)

	RE probit							
Variables	**Model 1**	**Model I**	**Model II**	**Model III**	**Model IV**	**Model V**	**Model VI**	**Model VII**
	The Total Sample							
drink	0.0263***	0.0313***	0.0277***	0.0287***	0.0277***	0.0285***		
	(0.0036)	(0.0035)	(0.0036)	(0.0035)	(0.0036)	(0.0035)		
smokenow	0.0123***	0.0108***	0.0127***	0.0092**	0.0134***		0.0184***	
	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)		(0.0040)	
drinkonly								0.0317***
								(0.0046)
smokeonly								0.0196***
								(0.0060)
drinksmoke								0.0375***
								(0.0047)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Likelihood ratio test (*p*-value)	967.71 (0.00)	1084.73 (0.00)	990.07 (0.00)	1054.36 (0.00)	972.24 (0.00)	963.32 (0.00)	974.12 (0.00)	967.50 (0.00)
Observations	37,591	37,591	37,591	37,591	37,591	37,591	37,591	37,591

Note: Only estimation results for marginal effects are reported. For dummy variables, the reported coefficient is the effect of changing from dummy from 0 to 1. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] Other explanatory variables are: gender, age, urban/rural, region of residence, presence of children in the household, and year.

Table 6.6 The Impact of Drinking and Smoking on Wages: Pooled OLS and FE (The Total Sample)

	(Model 1)		(Model 2)		(Model 3)		(Model 4)	
Variables	**Pooled OLS**	**FE**	**Pooled OLS**	**FE**	**Pooled OLS**	**FE**	**Pooled OLS**	**FE**
					The Total Sample			
drink	0.0129	-0.0007	0.0110	-0.0017				
	(0.0125)	(0.0181)	(0.0125)	(0.0181)				
drink0to1m					0.0182	0.0135		
					(0.0220)	(0.0304)		
drink1to2m					0.0045	0.0206		
					(0.0180)	(0.0253)		
drink1to2w					0.0109	-0.0316		
					(0.0181)	(0.0239)		
drink3to4w					-0.0141	-0.0086		
					(0.0217)	(0.0299)		
drinkdaily					0.0369*	0.0024		
					(0.0197)	(0.0299)		
wkdrinkamt							0.0001	0.0004
							(0.0012)	(0.0019)
$(\text{wkdrinkamt})^2/100$							0.0003	-0.0016
							(0.0017)	(0.0028)
smokenow	0.0028	-0.0153			0.0024	-0.0123		
	(0.0147)	(0.0239)			(0.0148)	(0.0240)		
eversmoked			0.0130	-0.0036				
			(0.0147)	(0.0240)				
dailysmokeamt							-0.0006	-0.0019
							(0.0019)	(0.0033)
$(\text{dailysmokeamt})^2/100$							0.0055	0.0105
							(0.0062)	(0.0098)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.6053	0.5788	0.6054	0.5788	0.6055	0.5792	0.5182	0.5355
Standard Hausman Test:								
$\chi^2(K)$ (*p*-value)		129.72 (0.00)		129.39 (0.00)		139.04 (0.00)		125.35 (0.00)
Number of id		5,911		5,911		5,911		5,099
Observations	10,871	10,871	10,871	10,871	10,871	10,871	8,396	8,396

Note: Under the pooled OLS analysis, cluster-robust standard errors are in parentheses. Under the FE analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1] The other explanatory variables are: gender, age, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

Table 6.7 The Impact of Drinking and Smoking on Wages: Pooled OLS and FE (Men vs. Women)

	(Model 1)		(Model 2)		(Model 3)		(Model 4)	
Variables	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE	Pooled OLS	FE
					Men			
drink	-0.0094	-0.0153	-0.0118	-0.0168				
	(0.0150)	(0.0221)	(0.0151)	(0.0221)				
drink0to1m					-0.0167	0.0034		
					(0.0296)	(0.0405)		
drink1to2m					-0.0155	0.0099		
					(0.0214)	(0.0297)		
drink1to2w					-0.0113	-0.0429		
					(0.0198)	(0.0267)		
drink3to4w					-0.0231	-0.0164		
					(0.0230)	(0.0326)		
drinkdaily					0.0088	-0.0149		
					(0.0211)	(0.0324)		
wkdrinkamt							-0.0005	0.0000
							(0.0013)	(0.0020)
$(\text{wkdrinkamt})^2/100$							0.0010	-0.0012
							(0.0017)	(0.0029)
smokenow	0.0016	-0.0134			0.0009	-0.0111		
	(0.0152)	(0.0247)			(0.0154)	(0.0248)		
eversmoked			0.0116	-0.0004				
			(0.0153)	(0.0249)				
dailysmokeamt							-0.0012	-0.0021
							(0.0019)	(0.0034)
$(\text{dailysmokeamt})^2/100$							0.0068	0.0105
							(0.0062)	(0.0099)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.5889	0.5927	0.5889	0.5927	0.5890	0.5932	0.5007	0.5498
Standard Hausman Test:								
$\chi^2(K)$ (p-value)		73.05 (0.00)		72.80 (0.00)		80.53 (0.00)		89.84 (0.00)
Number of id		3,356		3,356		3,356		2,913
Observations	6,364	6,364	6,364	6,364	6,364	6,364	4,927	4,927

				Women				
drink	0.0654***	0.0421	0.0649***	0.0411				
	(0.0230)	(0.0322)	(0.0230)	(0.0322)				
drink0to1m					0.0545	0.0279		
					(0.0337)	(0.0465)		
drink1to2m					0.0319	0.0457		
					(0.0343)	(0.0520)		
drinkoften					0.1071***	0.0607		
					(0.0407)	(0.0567)		
wkdrinkamt							0.0131**	0.0072
							(0.0065)	(0.0084)
$(\text{wkdrinkamt})^2/100$							-0.0168*	-0.0107
							(0.0086)	(0.0112)
smokenow	0.0284	-0.0361			0.0180	-0.0416		
	(0.0605)	(0.1358)			(0.0607)	(0.1369)		
eversmoked			0.0482	0.0136				
			(0.0610)	(0.1309)				
dailysmokeamt							0.0077	-0.0798
							(0.0125)	(0.0532)
$(\text{dailysmokeamt})^2/100$							-0.0428	0.4602*
							(0.0503)	(0.2629)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[2] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.6161	0.5620	0.6161	0.5620	0.6163	0.5621	0.5307	0.5268
Standard Hausman Test:								
$\chi^2(K)$ (p-value)		86.34 (0.00)		86.22 (0.00)		87.29 (0.00)		73.58 (0.00)
Number of id		2,555		2,555		2,555		2,186
Observations	4,507	4,507	4,507	4,507	4,507	4,507	3,469	3,469

Note: Under the pooled OLS analysis, cluster-robust standard errors are in parentheses. Under the FE analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1-2] The other explanatory variables are: age, educational attainment, urban/rural, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

Table 6.8 The Impact of Drinking and Smoking on Wages: Pooled OLS and FE (Young vs. Older)

	(Model 1)		(Model 2)		(Model 3)		(Model 4)	
Variables	**Pooled OLS**	**FE**	**Pooled OLS**	**FE**	**Pooled OLS**	**FE**	**Pooled OLS**	**FE**
					Young (aged 16-45)			
drink	0.0132	-0.0080	0.0115	-0.0088				
	(0.0144)	(0.0214)	(0.0144)	(0.0214)				
drink0to1m					0.0116	0.0115		
					(0.0251)	(0.0359)		
drink1to2m					-0.0085	0.0132		
					(0.0204)	(0.0298)		
drink1to2w					0.0125	-0.0338		
					(0.0208)	(0.0284)		
drink3to4w					-0.0160	-0.0204		
					(0.0256)	(0.0359)		
drinkdaily					0.0666***	-0.0338		
					(0.0239)	(0.0371)		
wkdrinkamt							-0.0003	-0.0010
							(0.0016)	(0.0025)
$(\text{wkdrinkamt})^2/100$							0.0012	0.0002
							(0.0021)	(0.0038)
smokenow	0.0184	-0.0354			0.0159	-0.0312		
	(0.0174)	(0.0295)			(0.0175)	(0.0297)		
eversmoked			0.0281	-0.0240				
			(0.0173)	(0.0294)				
dailysmokeamt							-0.0011	-0.0032
							(0.0022)	(0.0043)
$(\text{dailysmokeamt})^2/100$							0.0153**	0.0226*
							(0.0075)	(0.0130)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.5988	0.5525	0.5989	0.5524	0.5994	0.5528	0.5128	0.5161
Standard Hausman Test:								
$\chi^2(K)$ (p-value)		133.45 (0.00)		132.62 (0.00)		145.42 (0.00)		115.75 (0.00)
Number of id		5,058		5,058		5,058		4,292
Observations	8,716	8,716	8,716	8,716	8,716	8,716	6,603	6,603
					Older (aged 46-59)			

drink	0.0351	-0.0054	0.0333	-0.0063				
	(0.0246)	(0.0410)	(0.0250)	(0.0410)				
drink0to1m					0.0597	0.0040		
					(0.0457)	(0.0690)		
drink1to2m					0.0402	0.0210		
					(0.0390)	(0.0569)		
drink1to2w					0.0246	-0.0571		
					(0.0358)	(0.0541)		
drink3to4w					0.0147	0.0135		
					(0.0400)	(0.0677)		
drinkdaily					0.0399	0.0138		
					(0.0338)	(0.0636)		
wkdrinkamt							0.0029	0.0047
							(0.0020)	(0.0041)
$(\text{wkdrinkamt})^2/100$							-0.0030	-0.0060
							(0.0027)	(0.0069)
smokenow	-0.0156	0.0905*			-0.0145	0.0870*		
	(0.0259)	(0.0488)			(0.0262)	(0.0494)		
eversmoked			-0.0078	0.1008**				
			(0.0271)	(0.0502)				
dailysmokeamt							0.0005	-0.0027
							(0.0032)	(0.0064)
$(\text{dailysmokeamt})^2/100$							-0.0069	0.0082
							(0.0094)	(0.0197)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[2] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.5845	0.4820	0.5845	0.4825	0.5847	0.4839	0.5123	0.5201
Standard Hausman Test:								
$\chi^2(K)$ (*p*-value)		62.51 (0.00)		62.80 (0.00)		65.67 (0.00)		65.79 (0.00)
Number of id		1,454		1,454		1,454		1,274
Observations	2,155	2,155	2,155	2,155	2,155	2,155	1,793	1,793

Note: Under the pooled OLS analysis, cluster-robust standard errors are in parentheses. Under the FE analysis, standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
[1-2] The other explanatory variables are: gender, educational attainment, urban/rural, region of residence, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

Table 6.9 The Impact of Drinking and Smoking on Employment: Newey (1987)'s Two-step, Pooled probit and RE probit Estimator

Variables	Total	Men	Women	Young (aged 16-45)	Older (aged 46-59)
			Newey (1987)'s Two-step Estimator		
The Effect of Drinking and Smoking in the Employment Equation					
wkdrinkamt	0.0128	-0.0684	0.4207*[1]	-0.1052	0.1247
	(0.0794)	(0.0915)	(0.2174)	(0.1128)	(0.0931)
wkdrinkamt2_100	-0.0182	0.1265	-0.8363*[2]	0.2057	-0.2088
	(0.1425)	(0.1639)	(0.4291)	(0.2098)	(0.1585)
dailysmokeamt	-0.0013	0.3064	-1.2722	0.2998	0.0588
	(0.3084)	(0.3887)	(1.2440)	(0.6857)	(0.1605)
dailysmokeamt2_100	0.0220	-1.0236	6.5468	-1.0347	-0.1997
	(1.0647)	(1.3251)	(6.2810)	(2.4553)	(0.5286)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Wald test of exogeneity: χ^2(d.f.) (*p*-value)	2.91 (0.57)	1.74 (0.78)	10.41 (0.03)	4.56 (0.34)	4.98 (0.29)
Amemiya-Lee-Newey minimum chi-sq statistic: χ^2(d.f.) (*p*-value)	2.00 (0.57)	0.60 (0.90)	2.88 (0.41)	1.41 (0.70)	3.24 (0.36)
Observations	30,015	16,189	13,826	22,723	7,292
			Pooled probit		
wkdrinkamt	0.0020***	0.0011***	0.0036	0.0029***	0.0014*
	(0.0005)	(0.0004)	(0.0024)	(0.0006)	(0.0008)
wkdrinkamt2_100	-0.0023***	-0.0014**	-0.0036	-0.0030***	-0.0018
	(0.0007)	(0.0006)	(0.0041)	(0.0010)	(0.0011)
dailysmokeamt	0.0005	0.0005	-0.0115**	0.0008	0.0003
	(0.0007)	(0.0006)	(0.0050)	(0.0008)	(0.0013)
dailysmokeamt2_100	0.0018	0.0008	0.0522**	0.0017	0.0029
	(0.0022)	(0.0019)	(0.0252)	(0.0028)	(0.0040)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[4] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Pseudo-R^2	0.1569	0.1777	0.1433	0.1512	0.1621
Observations	30,015	16,189	13,826	22,723	7,292
			RE probit		
wkdrinkamt	0.0019***	0.0011***	0.0029	0.0025***	0.0016**

	(0.0004)	(0.0004)	(0.0024)	(0.0006)	(0.0008)
wkdrinkamt2_100	-0.0021***	-0.0014***	-0.0021	-0.0026***	-0.0019*
	(0.0006)	(0.0005)	(0.0042)	(0.0009)	(0.0011)
dailysmokeamt	0.0009	0.0007	-0.0092*	0.0010	0.0007
	(0.0006)	(0.0005)	(0.0052)	(0.0007)	(0.0011)
dailysmokeamt2_100	0.0007	0.0001	0.0423	0.0005	0.0025
	(0.0019)	(0.0015)	(0.0272)	(0.0024)	(0.0035)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[5] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Likelihood ratio test (*p*-value)	830.98 (0.00)	332.70 (0.00)	440.21 (0.00)	471.39 (0.00)	235.58 (0.00)
Number of id	13,403	6,921	6,482	11,387	4,314
Observations	30,015	16,189	13,826	22,723	7,292

Notes: The coefficient estimation results based on the Newey (1987)'s two-step are reported. The estimation results for marginal effects based on pooled probit and the RE probit are reported. Under the Newey (1987)'s two-step and the RE probit analysis, standard errors are in parentheses. Under the pooled probit analysis, cluster-robust standard errors are used to correct for the error correlation over time for a given individual. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[1-2] The estimation results for marginal effects are 0.0640 and -0.1272, respectively.

[3-5] The other explanatory variables are: gender (not in models for men and women), age group (not in models for the young and the older sample), educational attainment, urban/rural, region of residence, marital status, presence of children in the household, and year.

Table 6.10 The Impact of Instruments on Drinking and Smoking: First-stage Estimates of the Newey (1987)'s Two-step Estimator

Variables	Total	Men	Women	Young (aged 16-45)	Older (aged 46-59)
Instruments in the Drinking Equation[1]					
drink0to1m_hh	-0.0085***	-0.0075**	-0.0016	-0.0100	-0.0050
	(0.0022)	(0.0035)	(0.0014)	(0.0023)	(0.0059)
drink1to2m_hh	-0.0116***	-0.0019	-0.0028***	-0.0155***	-0.0001
	(0.0016)	(0.0029)	(0.0009)	(0.0017)	(0.0044)
drink1to2w_hh	-0.0278***	-0.0086***	-0.0061***	-0.0324***	-0.0082**
	(0.0014)	(0.0029)	(0.0008)	(0.0014)	(0.0040)
drink3to4w_hh	-0.0567***	-0.0338***	-0.0114***	-0.0610***	-0.0337***
	(0.0018)	(0.0043)	(0.0010)	(0.0019)	(0.0054)
drinkdaily_hh	-0.1197***	-0.1303***	-0.0185***	-0.1196***	-0.1081***
	(0.0016)	(0.0036)	(0.0009)	(0.0017)	(0.0043)
dkamt_hh	1.1165***	1.3685***	0.2280***	1.0749***	1.1723***
	(0.0051)	(0.0061)	(0.0048)	(0.0064)	(0.0091)
skamt_hh	-0.0258**	-0.0634***	-0.0214***	0.0010	-0.0719***
	(0.0059)	(0.0075)	(0.0047)	(0.0069)	(0.0117)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
R-squared	0.6780	0.7759	0.1603	0.6247	0.7397
Observations	30,015	16,189	13,826	22,723	7,292
Instruments in the Smoking Equation[2]					
drink0to1m_hh	-0.0069***	0.0049	-0.0010	-0.0114***	0.0063
	(0.0021)	(0.0031)	(0.0010)	(0.0023)	(0.0050)
drink1to2m_hh	-0.0143***	-0.0009	-0.0014**	-0.0179***	-0.0066*
	(0.0015)	(0.0026)	(0.0006)	(0.0016)	(0.0037)
drink1to2w_hh	-0.0243***	-0.0106***	-0.0029***	-0.0280***	-0.0141***
	(0.0013)	(0.0026)	(0.0005)	(0.0014)	(0.0034)
drink3to4w_hh	-0.0308***	-0.0174***	-0.0030***	-0.0348***	-0.0191***
	(0.0017)	(0.0039)	(0.0007)	(0.0018)	(0.0046)
drinkdaily_hh	-0.0442***	-0.0293***	-0.0030***	-0.0513***	-0.0258***
	(0.0015)	(0.0032)	(0.0007)	(0.0016)	(0.0037)
dkamt_hh	0.0202***	-0.0369***	0.0030	0.0651***	-0.0344***
	(0.0048)	(0.0055)	(0.0034)	(0.0063)	(0.0077)
skamt_hh	0.9233***	1.3039***	0.0944***	0.8663***	1.0378***

	(0.0055)	(0.0068)	(0.0034)	(0.0067)	(0.0099)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
R-squared	0.6494	0.7256	0.0898	0.6212	0.6927
Observations	30,015	16,189	13,826	22,723	7,292

Notes: Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

[1-2] The following explanatory variables are: gender, age group, educational attainment, region of residence, marital status, presence of children in the household and year.

Table 6.11 The Impact of Drinking and Smoking on Wages: 2SLS, Pooled OLS, and FE

Variables	Total	Men	Women	Young (aged 16-45)	Older (aged 46-59)
			2SLS		
The Effect of Drinking and Smoking in the Wage Equation					
wkdrinkamt	0.0291	0.0067	0.0038	0.0131	-0.0313
	(0.0317)	(0.0184)	(0.0600)	(0.0419)	(0.0341)
wkdrinkamt2_100	-0.0519	-0.0125	0.0016	-0.0226	0.0549
	(0.0568)	(0.0328)	(0.1064)	(0.0755)	(0.0600)
dailysmokeamt	-0.0491	-0.0439	-0.2194	-0.0310	-0.0050
	(0.0620)	(0.0433)	(0.4878)	(0.0713)	(0.0886)
dailysmokeamt2_100	0.1721	0.1611	0.8471	0.1189	0.0220
	(0.2269)	(0.1582)	(2.5413)	(0.2733)	(0.3005)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Hausman statistic: χ^2(d.f.) (*p*-value)	6316 (0.19)	2.49 (0.65)	2.34 (0.67)	5.75 (0.22)	5.57 (0.23)
Hansen J statistic: χ^2(d.f.) (*p*-value)	1.16 (0.76)	4.79 (0. 19)	4.70 (0. 20)	2.84 (0. 42)	0.83 (0. 84)
R-squared	0.4317	0.4282	0.5117	0.4920	0.4027
Observations	8,396	4,927	3,469	6,603	1,793
			Pooled OLS		
wkdrinkamt	0.0001	-0.0005	0.0131**	-0.0003	0.0029
	(0.0012)	(0.0013)	(0.0065)	(0.0016)	(0.0020)
wkdrinkamt2_100	0.0003	0.0010	-0.0168*	0.0012	-0.0030
	(0.0017)	(0.0017)	(0.0086)	(0.0021)	(0.0027)
dailysmokeamt	-0.0006	-0.0012	0.0077	-0.0011	0.0005
	(0.0019)	(0.0019)	(0.0125)	(0.0022)	(0.0032)
dailysmokeamt2_100	0.0055	0.0068	-0.0428	0.0153**	-0.0069
	(0.0062)	(0.0062)	(0.0503)	(0.0075)	(0.0094)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[2] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
R-squared	0.5182	0.5007	0.5307	0.5128	0.5123
Observations	8,396	4,927	3,469	6,603	1,793
			FE		

wkdrinkamt	0.0004	0.0000	0.0072	-0.0010	0.0047
	(0.0019)	(0.0020)	(0.0084)	(0.0025)	(0.0041)
wkdrinkamt2_100	-0.0016	-0.0012	-0.0107	0.0002	-0.0060
	(0.0028)	(0.0029)	(0.0112)	(0.0038)	(0.0069)
dailysmokeamt	-0.0019	-0.0021	-0.0798	-0.0032	-0.0027
	(0.0033)	(0.0034)	(0.0532)	(0.0043)	(0.0064)
dailysmokeamt2_100	0.0105	0.0105	0.4602*	0.0226*	0.0082
	(0.0098)	(0.0099)	(0.2629)	(0.0130)	(0.0197)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
R-squared	0.5355	0.5498	0.5268	0.5161	0.5201
Standard Hausman Test: $\chi^2(36)$ (p-value)	125.35 (0.00)	89.84 (0.00)	73.58 (0.00)	115.75 (0.00)	65.79 (0.00)
Number of id	5,099	2,913	2,186	4,292	1,274
Observations	8,396	4,927	3,469	6,603	1,793

Notes: Under the 2SLS and pooled OLS analysis, cluster-robust standard errors are in parentheses. Under the FE analysis, standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
[1-3] The other explanatory variables are: gender (not in models for men and women), age group (not in models for the young and the older sample), educational attainment, region of residence, marital status, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

Appendix 6.1 Patterns of Drinking Alcohol and Health Consequences by Country and Gender

Australia		**Canada**		**China**		**India**		**Japan**		**UK**		**US**	
Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Adult (15+ years) per capita consumption, in litres of pure alcohol, average 2003-2005 (drinkers only):													
16.26	7.23	18.20	7.04	13.68	5.19	23.93	10.35	13.76	5.75	21.58	9.46	19.98	8.45
Lifetime abstainers (15+ years), 2003:													
5.90%	11.20%	5.70%	10.80%	12.50%	44.50%	67.20%	92.00%	4.90%	13.50%	8.90%	15.20%	12.00%	22.90%
Former drinkers, 2003:													
6.10%	8.40%	13.10%	15.00%	14.90%	17.40%	12.80%	5.00%	7.00%	5.10%	1.50%	2.90%	16.30%	17.60%
Abstainers (persons who did not drink in the past 12 months), 2003:													
12.00%	19.60%	18.80%	25.80%	27.40%	61.90%	80.00%	97.00%	11.90%	18.60%	10.40%	18.10%	28.30%	40.50%
Heavy episodic drinkers (15-85 years, had at least 60 grams or more of pure alcohol on at least one occasion weekly), 2003:													
9.90%	2.60%	15.50%	3.50%	9.60%	0.30%	25.00%	14.6%	20.00%	3.90%	NA	NA	13.00%	3.40%

Source: WHO Country Profile. Available: http://www.who.int/countries/en/. [cited 23th Apr 2013].
NA = No information

Appendix 6.2 Alcohol Policy

Alcohol Policy	**Australia**	**Canada**	**China**	**India**	**Japan**	**UK**	**US**
Excise tax on beer/wine/spirits	Yes/No/Yes	Yes/Yes/Yes	Yes/Yes/Yes	Yes/Yes/Yes	Yes/Yes/Yes	Yes/Yes/Yes	No/ No/Yes
National legal minimum age for off-premise sales of alcoholic beverages (selling) (beer/wine/spirits)	18/18/18	Subnational	No/ No/No	Subnational	20/20/20	18/18/18	21/21/21
National legal minimum age for on-premise sales of alcoholic beverages (selling) (beer/wine/spirits)	18/18/18	Subnational	No/ No/No	Subnational	20/20/20	16/16/18	21/21/21
Restrictions for on-/off-premise ales of alcoholic beverages: Time (hours and days) / location (places and density) Specific events / intoxicated persons / petrol stations	Yes/Yes Yes/Yes/Yes	No/ No No/ No/No	No/ No No/ No/No	Yes/Yes Yes/No/No	No/ No No/ No/No	Yes & No Yes/Yes/Yes	No/ No No/ No/No
National maximum legal blood alcohol concentration (BAC) when driving a vehicle (general / young / professional), in %	0.05/ZT/ZT	0.08/0.04/0.08	0.02/0.02/0.02	0.03/0.03/0.03	0.03/0.03/0.03	0.08/0.08/0.08	0.08/0.02/0.04

Legally binding regulations on alcohol advertising / product placement	Yes/No	Yes/No	Yes/No	Yes/Yes	No/No	Yes/Yes	No/No
Legally binding regulations on alcohol sponsorship / sales promotion	No/Yes	No/No	No/No	Yes/Yes	No/No	No/No	NA

Source: WHO Country Profile. Available: http://www.who.int/countries/en/. [cited 23th Apr 2013].
ZT = Zero tolerance.
NA = No information

Appendix 6.3 WHO Age-standardized Estimated Prevalence of Smoking among Adults Aged 15 or More, 2009

Australia		**Canada**		**China**		**India**		**Japan**		**UK**		**US**	
Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Any smoked tobacco (%): *Current (Daily)*													
22 (19)	19 (16)	24 (15)	17 (12)	51 (44)	2 (2)	26 (20)	4 (3)	42 (39)	12 (11)	25 (17)	23 (16)	33 (17)	25 (15)
Cigarettes (%): *Current (Daily)*													
22 (19)	19 (16)	19 (15)	16 (12)	50 (43)	2 (2)	11 (7)	1 (<1)	42 (39)	12 (11)	25 (17)	23 (16)	28 (16)	24 (15)

Source: WHO Country Profile. Available: http://www.who.int/countries/en/. [cited 23th Apr 2013].

Appendix 6.4 Summary Statistics by Ex-smokers and Current Smokers

Variables	**Mean (Standard Deviation)**	
	Ex-smokers (N =422)	**Current Smokers** (N = 13,057)
age (in years)	47.36 (9.76)	39.69 (10.77)
age16_25 (%)	0.03 (0.17)	0.13 (0.33)
age26_35 (%)	0.12 (0.33)	0.25 (0.43)
age36_45 (%)	0.25 (0.43)	0.30 (0.46)
age46_59 (%)	0.60 (0.49)	0.32 (0.47)
excellent (%)	0.12 (0.32)	0.17 (0.37)
good (%)	0.45 (0.50)	0.59 (0.49)
fair (%)	0.32 (0.47)	0.21 (0.41)
poor (%)	0.11 (0.31)	0.03 (0.17)
noedu (%)	0.10 (0.30)	0.12 (0.33)
primary (%)	0.20 (0.40)	0.24 (0.43)
lowermid (%)	0.36 (0.48)	0.40 (0.49)
uppermid (%)	0.19 (0.39)	0.16 (0.37)
techvoc (%)	0.10 (0.30)	0.05 (0.21)
university (%)	0.05 (0.22)	0.03 (0.18)

Source: Author's calculations using CHNS data.

Appendix 6.5 Summary Statistics of the Instruments

Variables	Mean (Standard Deviation)				
	Total (N =30,015)	**Men** (N = 16,189)	**Women** (N = 13,826)	**Young (aged 16-45)** (N = 22,723)	**Older (aged 46-59)** (N = 7,292)
nodrink_hh (× 100%)	54.23 (45.51)	67.64 (42.84)	38.53 (43.48)	54.84 (45.47)	52.34 (45.57)
drink1to1m_hh (× 100%)	3.00 (15.28)	2.58 (14.02)	3.50 (16.63)	3.09 (15.73)	2.74 (13.80)
drink1to2m_hh (× 100%)	6.18 (21.83)	3.88 (16.91)	8.88 (26.20)	6.56 (22.70)	5.00 (18.82)
drink1to2w_hh (× 100%)	9.13 (26.06)	4.35 (17.09)	14.73 (32.78)	10.00 (27.46)	6.42 (20.87)
drink3to4w_hh (× 100%)	4.84 (19.59)	1.93 (11.45)	8.26 (25.65)	5.36 (20.69)	3.24 (15.55)
drinkdaily_hh (× 100%)	8.43 (25.04)	3.14 (14.26)	14.63 (32.43)	9.29 (26.14)	5.77 (21.00)
dkamt_hh (in units)	4.15 (7.48)	4.83 (8.47)	3.34 (6.03)	3.75 (6.55)	5.40 (9.73)
skamt_hh (in no. of cigarettes)	5.50 (6.10)	6.35 (6.76)	4.50 (5.04)	5.19 (5.60)	6.46 (7.36)

Source: Author's calculations using CHNS data.
Note: The share of the respondents' household members with a same type of drinking frequency is calculated as:

$$\frac{\text{The number of a respondent's working-age household members in a specified type of drinking frequency}}{\text{The total number of the respondent's working-age household members}} \times 100\%$$

Appendix 6.6 Sensitivity Analyses based on the Employment Equation (All Sub Samples within the Population)

					RE			
Variables	**Model 1**	**Model I**	**Model II**	**Model III**	**Model IV**	**Model V**	**Model VI**	**Model VII**
					Men			
drink	0.0202***	0.0257***	0.0227***	0.0210***	0.0221***	0.0228***		
	(0.0038)	(0.0039)	(0.0038)	(0.0038)	(0.0038)	(0.0037)		
smokenow	0.0109***	0.0111***	0.0118***	0.0087**	0.0125***		0.0158***	
	(0.0038)	(0.0037)	(0.0038)	(0.0037)	(0.0038)		(0.0037)	
drinkonly								0.0327***
								(0.0057)
smokeonly								0.0242***
								(0.0060)
drinksmoke								0.0346***
								(0.0052)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Likelihood ratio test (*p*-value)	415.37 (0.00)	475.43 (0.00)	437.69 (0.00)	436.43 (0.00)	424.73 (0.00)	410.25 (0.00)	417.34 (0.00)	415.14 (0.00)
Observations	20,345	20,345	20,345	20,345	20,345	20,345	20,345	20,345
					Women			
drink	0.0270***	0.0333***	0.0271***	0.0332***	0.0272***	0.0250***		
	(0.0080)	(0.0075)	(0.0080)	(0.0075)	(0.0080)	(0.0080)		
smokenow	-0.0493**	-0.0580**	-0.0513**	-0.0557**	-0.0497**		-0.0422**	
	(0.0219)	(0.0230)	(0.0220)	(0.0228)	(0.0219)		(0.0211)	
drinkonly								0.0272***
								(0.0082)
smokeonly								-0.0475*
								(0.0251)
drinksmoke								-0.0208
								(0.0349)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes

Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[2] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Likelihood ratio test (*p*-value)	487.12 (0.00)	558.83 (0.00)	489.98 (0.00)	555.49 (0.00)	486.31 (0.00)	489.06 (0.00)	487.36 (0.00)	487.08 (0.00)
Observations	17,246	17,246	17,246	17,246	17,246	17,246	17,246	17,246
Young (aged 16-45)								
drink	0.0221*** (0.0038)	0.0289*** (0.0036)	0.0225*** (0.0038)	0.0238*** (0.0037)	0.0275*** (0.0037)	0.0244*** (0.0037)		
smokenow	0.0133*** (0.0046)	0.0167*** (0.0044)	0.0135*** (0.0046)	0.0102** (0.0046)	0.0195*** (0.0044)		0.0188*** (0.0043)	
drinkonly								0.0258*** (0.0048)
smokeonly								0.0188*** (0.0064)
drinksmoke								0.0336*** (0.0049)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Likelihood ratio test (*p*-value)	611.74 (0.00)	731.29 (0.00)	622.43 (0.00)	679.92 (0.00)	658.31 (0.00)	609.62 (0.00)	616.90 (0.00)	611.50 (0.00)
Observations	29,126	29,126	29,126	29,126	29,126	29,126	29,126	29,126
Older (aged 46-59)								
drink	0.0442*** (0.0091)	0.0553*** (0.0092)	0.0499*** (0.0091)	0.0488*** (0.0092)	0.0444*** (0.0091)	0.0465*** (0.0089)		
smokenow	0.0124 (0.0094)	0.0097 (0.0096)	0.0141 (0.0095)	0.0079 (0.0096)	0.0125 (0.0094)		0.0214** (0.0092)	
drinkonly								0.0577*** (0.0124)
smokeonly								0.0284* (0.0145)
drinksmoke								0.0589***

								(0.0120)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[4] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Likelihood ratio test (*p*-value)	251.16	262.42	254.34	259.09	251.00	249.47	252.34	251.37
Observations	8,465	8,465	8,465	8,465	8,465	8,465	8,465	8,465

Note: Only estimation results for marginal effects are reported. For dummy variables, the reported coefficient is the effect of changing from dummy from 0 to 1. Cluster-robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[1-2] Other explanatory variables are: age, urban/rural, region of residence, presence of children in the household, and year.

[3-4] Other explanatory variables are: gender, urban/rural, region of residence, presence of children in the household, and year.

Appendix 6.7 Sensitivity Analyses under the Wage Equation (All Sample Groups within the Population)

Variables	Model 1	Model I	Model II	Model III	Model IV	Model V	Model VI	Model VII
					FE			
					The Total Sample			
drink	-0.0007	0.0020	0.0006	0.0002	-0.0002	-0.0020		
	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0180)		
smokenow	-0.0153	-0.0149	-0.0150	-0.0157	-0.0147		-0.0154	
	(0.0239)	(0.0240)	(0.0239)	(0.0240)	(0.0239)		(0.0237)	
drinkonly								0.0119
								(0.0220)
smokeonly								0.0098
								(0.0345)
drinksmoke								-0.0140
								(0.0283)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[1] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Hausman Test:								
$\chi^2(K)$ (p-value)	129.72 (0.00)	142.52 (0.00)	122.89 (0.00)	149.80 (0.00)	130.20 (0.00)	129.49 (0.00)	128.29 (0.00)	129.37 (0.00)
R-squared	0.5788	0.5747	0.5777	0.5760	0.5788	0.5788	0.5788	0.5789
Number of id	5,911	5,911	5,911	5,911	5,911	5,911	5,911	5,911
Observations	10,871	10,871	10,871	10,871	10,871	10,871	10,871	10,871
					Men			
drink	-0.0153	-0.0127	-0.0136	-0.0158	-0.0141	-0.0169		
	(0.0221)	(0.0221)	(0.0220)	(0.0221)	(0.0221)	(0.0219)		
smokenow	-0.0134	-0.0118	-0.0131	-0.0142	-0.0112		-0.0156	
	(0.0247)	(0.0248)	(0.0247)	(0.0248)	(0.0247)		(0.0245)	
drinkonly								-0.0111
								(0.0302)
smokeonly								-0.0077
								(0.0373)
drinksmoke								-0.0271
								(0.0318)

SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[2] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Hausman Test:								
$\chi^2(K)$ (p-value)	73.05 (0.00)	82.55 (0.00)	70.40 (0.00)	84.57 (0.00)	75.01 (0.00)	72.91 (0.00)	73.20 (0.00)	72.99 (0.00)
R-squared	0.5927	0.5875	0.5920	0.5888	0.5922	0.5927	0.5926	0.5927
Number of id	3,356	3,356	3,356	3,356	3,356	3,356	3,356	3,356
Observations	6,364	6,364	6,364	6,364	6,364	6,364	6,364	6,364
Women								
drink	0.0421	0.0389	0.0407	0.0422	0.0402	0.0414		
	(0.0322)	(0.0322)	(0.0323)	(0.0322)	(0.0322)	(0.0321)		
smokenow	-0.0361	-0.0395	-0.0508	-0.0284	-0.0328		-0.0218	
	(0.1358)	(0.1358)	(0.1358)	(0.1358)	(0.1358)		(0.1354)	
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Standard Hausman Test:								
$\chi^2(K)$ (p-value)	86.34 (0.00)	78.49 (0.00)	80.36 (0.00)	86.88 (0.00)	83.88 (0.00)	86.19 (0.00)	84.93 (0.00)	
R-squared	0.5620	0.5569	0.5598	0.5597	0.5615	0.5620	0.5616	
Number of id	2,555	2,555	2,555	2,555	2,555	2,555	2,555	
Observations	4,507	4,507	4,507	4,507	4,507	4,507	4,507	
Young (age 16-45)								
drink	-0.0080	-0.0068	-0.0078	-0.0072	-0.0078	-0.0109		
	(0.0214)	(0.0214)	(0.0214)	(0.0214)	(0.0213)	(0.0212)		
smokenow	-0.0354	-0.0349	-0.0361	-0.0342	-0.0352		-0.0366	
	(0.0295)	(0.0296)	(0.0295)	(0.0296)	(0.0295)		(0.0293)	
drinkonly								0.0234
								(0.0255)

smokeonly								0.0330 (0.0422)
drinksmoke								-0.0416 (0.0345)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[4] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Hausman Test: $\chi^2(K)$ (p-value)	133.45 (0.00)	113.42 (0.00)	125.10 (0.00)	126.08 (0.00)	132.49 (0.00)	131.38 (0.00)	131.52 (0.00)	134.41 (0.00)
R-squared	0.5525	0.5480	0.5515	0.5492	0.5525	0.5523	0.5525	0.5531
Number of id	5,058	5,058	5,058	5,058	5,058	5,058	5,058	5,058
Observations	8,716	8,716	8,716	8,716	8,716	8,716	8,716	8,716
Older (aged 46-59)								
drink	-0.0054 (0.0410)	-0.0008 (0.0410)	-0.0021 (0.0409)	-0.0085 (0.0409)	-0.0018 (0.0412)	0.0011 (0.0409)		
smokenow	0.0905* (0.0488)	0.0921* (0.0489)	0.0903* (0.0488)	0.0935* (0.0487)	0.0895* (0.0490)		0.0899* (0.0486)	
drinkonly								-0.0803 (0.0542)
smokeonly								-0.0158 (0.0701)
drinksmoke								0.0614 (0.0618)
SRHS Variables Controlled: (Yes/No)	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Qualification Variables Controlled: (Yes/No)	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Marital Status Variable Controlled: (Yes/No)	Yes	No	Yes	Yes	No	Yes	Yes	Yes
Other Explanatory Variables[5] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard Hausman Test: $\chi^2(K)$ (p-value)	62.51 (0.00)	65.75 (0.00)	59.45 (0.00)	72.84 (0.00)	59.41 (0.00)	57.47 (0.00)	61.98 (0.00)	68.31 (0.00)
R-squared	0.4820	0.4701	0.4785	0.4798	0.4762	0.4793	0.4820	0.4854

Number of id	1,454	1,454	1,454	1,454	1,454	1,454	1,454	1,454
Observations	2,155	2,155	2,155	2,155	2,155	2,155	2,155	2,155

Note: Cluster-robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

[1] Other explanatory variables are: gender, age, urban/rural, region of residence, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

[2-3] Other explanatory variables are: age, urban/rural, region of residence, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

[4-5] Other explanatory variables are: gender, urban/rural, region of residence, presence of children in the household, occupation, employment position, type of work unit, size of work unit, having second job and year.

Appendix 6.8 The Impact of Instruments on Drinking and Smoking: First-stage Estimates of the 2SLS

Variables	Total	Men	Women	Young (aged 16-45)	Older (aged 46-59)
Instruments in the Drinking Equation[1]					
drink0to1m_hh	-0.0112***	-0.0127***	-0.0039**	-0.0127***	-0.0041
	(0.0026)	(0.0039)	(0.0017)	(0.0026)	(0.0073)
drink1to2m_hh	-0.0137***	-0.0065	-0.0068***	-0.0158***	-0.0035
	(0.0024)	(0.0040)	(0.0019)	(0.0027)	(0.0062)
drink1to2w_hh	-0.0288***	-0.0112**	-0.0109***	-0.0329***	-0.0060
	(0.0023)	(0.0046)	(0.0031)	(0.0026)	(0.0061)***
drink3to4w_hh	-0.0517***	-0.0271	-0.0144***	-0.0544***	-0.0353
	(0.0038)	(0.0096)	(0.0055)	(0.0041)	(0.0129)
drinkdaily_hh	-0.1164***	-0.1321***	-0.0256***	-0.1183***	-0.0894***
	(0.0053)	(0.0137)	(0.0076)	(0.0060)	(0.0158)
dkamt_hh	1.1620***	1.3875***	0.3373***	1.1254***	1.2519***
	(0.0335)	(0.0399)	(0.0933)	(0.0438)	(0.0548)
skamt_hh	-0.0264*	-0.0516**	-0.0340**	0.0079	-0.1094***
	(0.0156)	(0.0203)	(0.0177)	(0.0170)	(0.0300)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
R-squared	0.6928	0.7778	0.2458	0.6679	0.7487
F statistic (*p*-value)	210.08 (0.00)	227.76 (0.00)	4.45 (0.00)	108.71 (0.00)	99.17 (0.00)
Observations	8,396	4,927	3,469	6,603	1,793
Instruments in the Smoking Equation[2]					
drink0to1m_hh	-0.0034	0.0546***	0.0049	-0.0064*	0.0035
	(0.0033)	(0.0042)	(0.0036)	(0.0036)	(0.0076)
drink1to2m_hh	-0.0090***	0.0509***	0.0025	-0.01.4***	-0.0123*
	(0.0025)	(0.0040)	(0.0034)	(0.0027)	(0.0069)
drink1to2w_hh	-0.0203***	0.0376***	-0.0071	-0.0248***	-0.0046
	(0.0023)	(0.0048)	(0.0044)	(0.0023)	(0.0070)
drink3to4w_hh	-0.0269***	0.0319***	-0.0088	-0.0308***	-0.0109
	(0.0032)	(0.0082)	(0.0079)	(0.0032)	(0.0112)
drinkdaily_hh	-0.0466*	0.0019	-0.0354***	-0.0512***	-0.0325***
	(0.0032)	(0.0070)	(0.0066)	(0.0034)	(0.0100)
dkamt_hh	0.0201***	0.0070	-0.0327***	0.0536***	-0.0499***
	(0.0107)	(0.0114)	(0.0124)	(0.0130)	(0.0170)

skamt_hh	1.0132***	1.4426***	1..3819***	0.9573***	1.1440***
	(0.0189)	(0.0196)	(0.0220)	(0.0228)	(0.0289)
SRHS Variables Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
Other Explanatory Variables[3] Controlled: (Yes/No)	Yes	Yes	Yes	Yes	Yes
R-squared	0.6641	0.7475	0.0541	0.6395	0.7079
F statistic (*p*-value)	454.40 (0.00)	573.49 (0.00)	1.53 (0.14)	268.13 (0.00)	257.84 (0.00)
Observations	8,396	4,927	3,469	6,603	1,793

Notes: Cluster-robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
[1-2] The other explanatory variables are: gender, age group, educational attainment, region of residence, marital status, presence of children in the household and year.

Chapter 7 Conclusion

The Chinese economy has carried out progressive reforms over the decades and the Chinese labour market has also experienced substantial change in line with the progress of the reforms. It implied a transition from a centrally planned to a more flexible and increasingly market-oriented labour market. Changes in both systems of job assignment and of wage determination were the most noticeable. Over the same period there were three main national health insurance systems as well as the growing private insurance system to jointly provide comprehensive healthcare benefits for the Chinese population, which lead to considerable progress in improving the health outcomes. These changes have implications for the relationship between health and labour market outcomes.

So far, much of the work in studying the effect of health on labour market outcomes has been based on developed countries such as UK, US, Germany and Austria. For developing countries, very few studies exist (for example, Behrman and Deolalika 1988; and Strauss and Thomas, 1998), and especially for China, such studies are relatively scarce (Wei, 2004). Some researchers suggest that health is thought to be a major determinant of labour market outcomes in developing countries. This is probably because the relationship between health and labour market outcomes is more apparent in societies in which many prime age adults are under-nourished and in poor health, and also because the theory of efficiency wages provides a natural starting point for examination of this issue (Currie and Madrian, 1999).

Although there is a possibility that the direction of causality runs from better labour market outcomes to health, or unobserved factors affect both health and labour market outcomes in either the same or different directions. Many empirical studies have evidenced that better health is conducive to employment and raises individual's productivity. Based on the panel data from the China Health and Nutrition Survey (CHNS), this thesis draws on those studies, and investigates the impact of health and health-related behaviour on employment and wages for the Chinese working-age population.

Chapter 4 examines the effect of self-reported health status and obesity on the probability of being employed. The panel nature of the CHNS data enables estimation of a RE probit model and a FE logit model to control for unobservable individual factors. In estimating the effect of having poor health status on employment, this thesis also employs a recursive bivariate probit model to take into account the potential endogeneity of health. The same model is also employed when estimating the effect of obesity on employment. The analysis draws the following major findings. First, there is a statistically significant and positive relationship between having better health status and employment for the Chinese working-age population. Second, having an unhealthy BMI is found to be negatively associated with employment. These two findings are consistent with the international literature. Based on the estimated results, people who have excellent, good, or fair health status are predicted to increase the probabilities of being employed compared to those who have poor health status. However, they have similar impact on employment. In other words, any changes among the excellent, good or fair health status would not significantly change the probabilities of being employed. Third, the results on effect of health on labour market outcomes for the sub samples within the Chinese working-age population are also found to be consistent with the international literature, the positive relationship between better health and employment is larger for men and older working-age population than for women and the young working-age population, respectively. In addition, the results also suggest that the effect for urban residents is found to be larger than the effect for their rural counterparts, which is consistent with the expectation. Fourth, controlling for the effect of SRHS only slightly reduce the marginal effects of BMI on the probability of being employed Although SRHS may not completely capture the impact of health on productivity, the result still reflects the presence of discrimination against people who have unhealthy body weight. As for the sub samples, being underweight is negatively correlated with employment for men, whereas the negative effect is found for women who are overweight or obese. Fifth, failure to control for individual unobserved factors would bias the estimates of the effect of health on employment. Comparing the marginal effects estimates after the pooled probit, estimates after RE probit are similar in terms of sign but smaller in terms of magnitude. Sixth, there is no evidence of endogeneity of health. Based on the bivariate probit model, the analysis fails to reject the null hypothesis of exogeneity of SRHS for all the sample groups except for women. In addition, being overweight is found to be endogenously related to employment. The finding that failure to account for the endogeneity of being overweight underestimates its impact on employment is consistent with international literature.

Chapter 5 estimates the impact of self-reported health status and obesity on hourly wages. The RE and the FE model are used to control for the individual unobservable factors. Based on an index measure of the self-reported health status and a continuous measure of BMI, the relationship between health and wages also uses a Two Stage Least Squares (2SLS) and a Heckman (1979) selection correction method to separately accounting for the potential endogeneity of health and the sample selection bias. The key major findings are concluded as follows. First, there is limited evidence that better health status increases labour productivity and therefore wages. Only having excellent health status are found to be positively affect hourly wages for the Chinese workers. Second, having an unhealthy BMI does not have a statistically significant effect on wages. This finding is robust to the use of the alternative BMI classifications for the Chinese population. Third, the sensitivity analysis based on two constructed variables of the real hourly wage income shows that using the real hourly wage plus bonuses to measure labour productivity yields more pronounced results than the use of the real hourly wage rate. As such it can be argued that the use of the real hourly wage rate may not fully reflect productivity gains form health. Fourth, there is no statistically significant effect of having unhealthy body weight on hourly wages for male workers. However, for female workers being obese is predicted to reduce hourly wages by 22.40% compared to their normal weight counterparts, which is larger than that of the international literature. There are two potential implications. On one hand, obesity is a relatively more important factor in predicting labour productivity for the Chinese women. On the other hand, if controlling for the effect of SRHS that fully captures the effect of health on wages, discrimination against obesity for the Chinese women is more pronounced than those from other countries. Fifth, there is no evidence of endogeneity of health. Based on the 2SLS method, the null hypothesis of exogeneity of health cannot be rejected for all sample groups within Chinese population. However, the result based on the Heckman method suggests the presence of sample selection bias in estimating the effect of health on hourly wages for the Chinese workers. The statistically significant and positive relationship between better health and wages largely disappears after controlling for sample selection bias.

Chapter 6 examines the simultaneous effect of drinking and smoking status on employment and wages. To allow for the potential endogeneity of drinking and smoking simultaneously, this chapter adopts a Newey (1978)'s two-step and a 2SLS method to estimate the effect on employment and wages, respectively. Most of the relevant work has been conducted in the Western countries, particular in the US, and comparatively fewer studies based on other western countries such as the UK, Germany, Canada, and Australia. This chapter asks how drinking and smoking impact on labour market outcomes in China and whether these impacts are comparable with those found in other studies. In Chinese culture drinking and smoking are popular social customs and this is important in understanding the results of such research.

For the overall population, both drinking and smoking status are found to be positively correlated with employment but not with hourly wages. The result confirms the literature that there is a positive and inverse U-shaped relationship between drinking and employment, which is also robust across different ways measuring drinking behaviour. Second, there is evidence of a positive and inverse U-shaped relationship between drinking and employment for all the sub samples, and the effect is larger for women / the young working-age adults than for men / the older working-age adults. Smoking incidence is found to be positively correlated with employment for men and the young working-age sample but not for women or the older working-age sample. There is no statistically significant effect of health-related behaviour on hourly wages for the sub samples, except for the older working-age adults, where smoking incidence is found to be positively correlated with hourly wages. The results in terms of the effect of smoking on labour market outcomes are found to be different from most of the existing literature. For the Chinese working-age population, the positive effect of smoking on labour market outcomes may be because that smoking serves as a bridge of communication with colleagues, employers and customers and thus helps to increase employment opportunities, staying employed and providing opportunity for promotion. Third, the presence of individual unobserved factors yields different result in estimating the effect of health related behaviour on employment and wages. For women, the positive and inverse U-shaped relationship between drinking and employment is driven by individual unobserved factors. For the older working-age adults, smoking incidence is found to be positively correlated with wages after controlling for individual unobserved factors. Fourth, there is no evidence of endogeneity of drinking or smoking.

Although establishing a link between empirical findings and policy initiatives is not always straightforward (Qiu, 2007), the results of this thesis still provide evidence to policy makers, including the Eighteenth and the Seventeenth National Congress of the CPC on improving the health of the whole nation. Although the Chinese health care system has made considerable progress in improving the health outcomes of the Chinese population, it is still less competitive in international terms. The positive effect of health on labour market outcomes suggests there is an economic benefit to better level of health. The national (both government and private) expenditure on health is low by international standards. Accordingly, it is important to improve the Chinese health care system along with increase the expenditure on health. This is also potentially important to public policymakers to conduct a more efficient allocation of public resources, where the improvement of people who have poor health status should be given more consideration. Based on the estimated results found in Chapter 4, improving health status from poor to fair would increase the employment probability by 8.8 percentage points, which is similar to the marginal effect (with a value of 9.0 percentage points) of changing from poor to good health status on the employment probability.

This thesis has several limitations. It fails to test the validity of exclusion restriction for instrumental variable in estimating in addressing the potential endogeneity of health to employment. This is primarily limited by the application of econometric methodologies. To the author's knowledge there is still no formal test for instrument validity for the bivariate probit model. In estimating the effect of health on wages by allowing the its potential endogeneity, the analysis does not consider the discrete nature of self-reported health status. Lee (1982) argues that failure to take into account the discrete nature of the indicators might introduce econometric problems. In addition, this study applies different methods to separately address different econometric issues (such as endogeneity and sample selection bias). The current methods to account for sample selection bias in panel models are still in an experimental stage and remain questionable. In this context, further efforts are expected to extend the analysis to consider the above problems and a more satisfactory method to address those econometric issues is required.

The measures health and labour market outcomes used in this thesis are not without drawbacks. For example, the self-reported health status might introduce justification bias, while even the objective BMI measure cannot efficiently distinguish between muscle and fat, thereby producing measurement error. The dependent variable used in the employment analysis simply relies on a binary choice, i.e. either being employed or not. On one hand, the employment status cannot distinguish self-employment from paid employment, and job seeking from being out of the labour force. On the other hand, the effect of health on hours of work is not captured. For example, being in unfavourable health status or BMI category might only reduce individuals' hours of work, while remaining in employment. Accordingly, another possible extension to the study is to focus the hours of work. The sample used in the wage analysis does not include the agricultural workers and self-employed because most of them reporting their income from other sources. Nevertheless, this thesis would have benefited from a question regarding the net profit earned from agricultural work and self-employment on measure labour productivity. it is also worth noting that adverse health outcomes such as obesity and excessive alcohol consumption might not likely to affect or have a limited effect on labour market outcomes in a short term period due to the presence of lagged effect. Therefore, it would be reasonable to explore the dynamic effect of health and health-related behaviour on labour market outcomes.

Bibliography

Aldoori, W.H., Giovannucci, E.L., Stampfer M.J., Rimm, E.B. Wing A.L., and Willett, W.C. (1997) 'A Prospective study of Alcohol, Smoking, Caffeine, and the Risk of Duodenal Ulcer in Men', *Epidemiology*, Vol. 8, No. 4 pp. 420-424.

Allison, P. D. (2009) *Fixed Effects Regression Models*. Sage. Chapter 6.

Amemiya, T. (1978) 'The Estimation of A Simultaneous Equation Generalized Probit Model', *Econometrica*, Vol. 46, No. 5, pp. 1193-1205.

Amemiya, T. (1979) 'The Estimation of A Simultaneous Equation Generalized Tobit Model', *International Economic Review*, Vol. 20, No. 1, pp. 169-181.

Amemiya, T. (1981) 'Qualitative Response Models: A Survey', *Journal of Economic Literature*, Vol. 19, No. 4, pp. 1483-1536.

Anderson, S. and Newell, R. G. (2003) 'Simplified Marginal Effects in Discrete Choice Models' *Economics Letters*, Vol. 81, No. 3, pp. 321-326.

Arackal, B.S. and Benegal, V. (2007) 'Prevalence of Sexual Dysfunction in Male Subjects with Alcohol Dependence', *Indian Journal of Psychiatry*, Vol. 49, No. 2, pp. 109-112. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2917074/. [cited 23rd Aug 2013].

Arcand, J. and Labar, K. (2005) 'Health and Wages: Instrumental Variables Estimates for a Panel of Chinese Workers', *CERDI Papers*. Available: http://www.cerdi.org/uploads/sfCmsContent/html/199/Arcand_Labar.pdf. [cited 28th Sept 2011].

Arulampalam, W. and Booth, A. L. (1998) 'Training and Labour Market Flexibility: Is There a Trade-off?', *British Journal of Industrial Relations*, Vol. 36, No. 4, pp. 521-536.

Auld, M. C. (2005) 'Smoking, Drinking, and Income', *The Journal of Human Resources*, Vol. 40, No. 2, pp. 505-518.

Baldwin, M. and Johnson, W. G. (1994) 'Labour Market Discrimination Against Men with Disabilities', *Journal of Human Resources*, Vol. 29, No. 1, pp. 1-19.

Banister, J. (2009) 'Health, Mortality, and Longevity in China Today', The XXVI IUSSP International Population Conference, Marrakech, Morocco. Available: http://iussp2009.princeton.edu/papers/90481. [citied 5th Jul 2012].

Bao, Y., Lu J., Wang C., Yang M., Li H., Zhang X., Zhu J., Lu H. and Jia W. (2008) 'Optimal Waist Circumference Cutoffs for Abdominal Obesity in Chinese', *Atherosclerosis*, Vol. 201, No. 2, pp. 378-384.

Barrett, G. (2002) 'The Effect of Alcohol Consumption on Earnings', *The Economic Record*, Vol. 78, No. 1, pp. 79-96.

Baum-Baicker, C. (1985a), 'The Health Benefits of Moderate Alcohol Consumption: A Review of the Literature', *Drug and Alcohol Dependence*, Vo. 15, No. 3, pp. 207-227.

Baum-Baicker, C. (1985b), 'The Psychological Benefits of Moderate Alcohol Consumption: A Review of the Literature', *Drug and Alcohol Dependence*, Vo. 15, No. 4, pp. 305-322.

Beaglehole, R. and Jackson, R. (1992) 'Alcohol, Cardiovascular Diseases and All Causes of Mortality: A Review of the Epidemiological Evidence', *Drug and Alcohol Review*, Vol. 11, No. 3, pp. 275-289.

Becker, G. S. (1962) 'Investment in Human Capital: A Theoretical Analysis', *Journal of Political Economy*, Vol. 70, No. 5, pp. 9-49.

Becker, G. S. (1964) *Human Capital.* 3rd ed. New York: Columbia University Press for the National Bureau of Economic Research.

Becker, G. S. (1971) *The Economics of Discrimination*, 2nd ed, Chicago, London: University of Chicago Press.

Benjamin, D., Brandt, L. and Fan, J. Z. (2003) 'Ceaseless Toil? Health and Labour Supply of the Elderly in Rural China' *Working Papers*, benjamin-03-01, University of Toronto, Department of Economics. Available: http://www.princeton.edu/rpds/papers/pdfs/benjamin_ceaseless_toil.pdf. [cited 23th Sept 2010].

Boffetta, P., and Garfinkel, L. (1990) 'Alcohol Drinking and Mortality among Men Enrolled in an American Cancer Society Prospective Study', *Epidemiology*, Vol. 1, No. 5, pp. 342-348.

Bound, J. (1991) 'Self-reported versus Objective Measures of Health in Retirement Models', *Journal of Human Resources*, Vol. 26, No. 2, pp. 361-386.

Bound, J., Schoenbaum, M. and Waidmann, T. (1996) 'Race Difference in Labour Force Attachment and Disability Status', The Gerontologist, Vol. 36, No. 3, pp. 311-321. Available: http://gerontologist.oxfordjournals.org/content/36/3/311.full.pdf. [cited 15th Mar 2012].

Blundell, R., Dearden, L., Goodman, A. and Reed, H. (2000) 'The Returns to Higher Education in Britain: Evidence from a British Cohort', *The Economic Journal*, Vol. 110, No. 461, pp. F82-F99.

Branigan, T. (2011) 'China's village of the bachelors: no wives in sight in remote settlement', The Guardian, 2rd Sept. Available: http://www.guardian.co.uk/world/2011/sep/02/china-village-of-bachelors. [cited 21st March 2012].

Brenner, H. Rothenbacher, D. Bode, G. and Adler, G. (1997) 'Relation of Smoking and Alcohol and Coffee Consumption to Active Helicobacter Pylori Infection: Cross Sectional Study', *BMJ*, Vol. 315, No. 7121, pp. 1489-1492. Available:

http://www.bmj.com/content/315/7121/1489?view=long&pmid=9420488. [cited 22nd Aug 2013].

Brooks, R. and Tao, R., 2003, 'China's Labour Market Performance and Challenges', *IMF Working Paper*, No. 03/210.

Brown, R., Pinkerton, R., and Tuttle, M. (1987) 'Respiratory Infections in Smokres', *American Family Physician*, Vol. 36, No. 5, pp. 133-140.

Brunello, G. and D'hombres, B (2007) 'Does body weight affect wages? Evidence from Europe', *Economics and Human Biology*, Vol. 5, No. 1, pp 1-19.

Cai, L. (2009) 'Effects of Health on Wages of Australian Men', *The Economic Record*, Vol. 85, No. 270, pp. 290-306.

Cai, L. (2010) 'The Relationship between Health and Labour Force Participation: Evidence from a Panel Data Simultaneous Equation Model', *Labour Economics*, Vol. 17, No. 1, pp. 77-90.

Cai, L. and Kald, G. (2006) 'Health status and labour force participation: evidence from Australia', *Health Economics*, Vol. 15, No. 3, pp.241-261.

Cameron, A. C. and Trivedi, P. K. (2009) *Microeconometrics Using Stata.* Stata Press.

Card, D. (1999) 'The Causal Effect of Education on Earnings', *Handbook of Labour Economics*, Vol. 3, Part A, pp. 1801-1863.

Cawley, J. (2004) 'The Impact of Obesity on Wages', *Journal of Human Resources*, Vol. 39, No. 2, pp. 451-474.

Chan, K.M. and Zhang, L. (1999) 'The *Hukou* System and Rural-Urban Migration in China: Processes and Changes' *The China Quarterly*, No. 160, pp. 818-855.

Chang, H. and Yen, S. (2011) 'Mental Health and Employment of the Elderly in Taiwan: A Simultaneous Equation Approach', Pacific Economic Review, Vol. 16, No. 4, pp. 504-519.

Chen, J. (2005) 'Improving the Nutrition and Health Status of Chinese People, for Facilitating the Development of Social Economy', *Chinese Journal of Preventive Medicine*, Vol. 39, No. 5, p. 299. Available:
http://www.pubhealth.org.cn/magazine_data/zhyf2005/0505pdf/050501.pdf. [Chinese Version]. [cited 29th Jun 2012].

Cherpitel, C.J., Ye, Y., Bond, J., Borges, G., Chou, P., Nilsen, P., Ruan, J. and Xiang, X. (2012) 'Multi-level Analysis of Alcohol-related Injury and Drinking Patterns: Emergency Department Data from 19 Countries', *Addiction*, Vol. 107, No. 7, pp. 1263-1272.

China Health and Nutrition Survey. Available:
http://www.cpc.unc.edu/projects/china. [cited 9th March 2010].

China Ratifies FCTC (2005), Tobacco Journal International. Available:

http://www.tobaccojournal.com/China_ratifies_FCTC.47216.0.html. [cited 17th May 2013].

Chirikos, T. N. (1993) 'The Relationship between Health and Labour Market Status', *Annual Review of Public Health*, Vol. 14, pp. 293-312.

Chow, G. C. (2006) 'An Economic Analysis of Health Care in China ', *CEPS Working Paper*, No. 132. Available:
http://www.princeton.edu/ceps/workingpapers/132chow.pdf. [cited at 30 Aug 2011].

Cochrane, J., Chen, H., Conigrave, K. M. and Hao, W. (2003) 'Alcohol Use in China', *Alcohol and Alcoholism*, Vol. 38, No. 6, pp. 537-542.

Commission on Macroeconomics and Health (CMH) (2005) 'China: Health, Poverty, and Economic Development', *Report of the Commission on Macroeconomics and Health (CMH) in China*, World Health Organization, Office of the WHO Representative in China, and China State Council, Development Research Center (DRC), Social Development Department. Available:
http://www.who.int/macrohealth/action/CMH_China.pdf. [cited 29th Aug 2011].

Contoyannis, P. and Rice, N. (2001) 'The Impact of Health on Wages: Evidence from the British Household Panel Survey', *Empirical Economics*, Vol. 26, No. 4, pp. 599-622.

Cooperative Meta-analysis Group of the Working Group on Obesity in China (C-WGOC). (2002) 'Predictive Values of Body Mass Index and Waist Circumference to Risk Factors of Related Diseases in Chinese Adult Population', Chinese Journal of Epidemiology, Vol. 23, No. 1, pp. 5-10.

Cosci, F., Schruers, K.R., Abrams, K. and Griez, E.J. (2007) 'Alcohol Use Disorders and Panic Disorder: A Review of the Evidence of a Direct Relationship', *Journal of Clinical Psychiatry*, Vol. 68, No. 6, pp. 874-880.

Currie, J. and Madrian, B. C. (1999) 'Health, Health Insurance and the Labour Market', in Ashenfelter, O. and Card, D. (1999) *Handbook of Labour Economics*, New York: Elsevier, Vol. 3c, Chapter 50, pp. 3309-3416. Available:
http://www.princeton.edu/~jcurrie/publications/Health_Health_Insurance.pdf. [cited 28th Sept 2011].

Davies, P.D., Yew, W.W., Ganguly, D., Davidow, A.L., Reichman L.B., Dheda, K. and Rook, G.A. (2006) 'Smoking and Tuberculosis: the Epidemiological Association and Immunopathogenesis', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol. 100, No. 4, pp. 291-298.

Davis, J. M. (1972) 'Impact of Health on Earnings and Labour Market Activity', *Monthly Labour Review*, Vol. 95, No. 10, pp. 46-49.

Ding, D.Z. and Warner, M. (2001) 'China's Labour-Management System Reforms: Breaking the 'Three Old Irons' (1978-1999)', *Asia Pacific Journal of Management*, Vol. 18, pp. 315-334.

Djousse, L. and Gaziano, J.M. (2008) 'Alcohol Consumption and Heart Failure', *Curr Atheroscler Rep*, Vol. 10, No. 2, pp. 117-120. Available:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2365733/. [cited 22nd Aug 2013].

Doll, R. (1986) 'Tobacco: An Overview of Health Effects', *IARC Scientific Publications*, Vol. 74, pp. 11-22.

Doll, R., Peto, R. Hall, E., Wheatley, K. and Gray, R. (1994) 'Mortality in Relation to Consumption of Alcohol: 13 Years' Observations on Male British Docotrs', *British Medical Journal*, Vol. 306, No. 6959, pp. 911-918.

Doll, R., Peto, R., Boreham, J. and Sutherland, I. (2005) 'Mortality in Relation to Alcohol Consumption: A Prospective Study among Male British Doctors', *International Journal of Epidemiology*, Vol. 34, No. 1, pp. 199-204. Available: http://ije.oxfordjournals.org/content/34/1/199.full. [cited 22nd Aug 2013].

Domencich, T. and McFadden, D. L. (1975) Urban Travel Demand: A Behavioral Analysis, Amsterdam Elsevier: North-Holland Publishing Co..

Du, J. (2009) 'Economic reforms and health insurance in China', *Social Science & Medicine*, Vol. 69, No. 3, pp. 387-395.

Du, S., Lu, B., Zhai, F. and Popkin, B. M. (2002) 'A New Stage of the Nutrition Transition in China', *Public Health Nutrition*, Vol. 5, No. 1a, pp. 169-174. Available: http://www.cpc.unc.edu/Plone/projects/nutrans/research/bellagio/papers/PHNChina-Dushufa.pdf. [cited 22nd Jun 2012].

Everett, M. (1990) 'Let an Overweight person Call on Your Best Customers? Fat Chance', *Sales and Marketing Management*, Vol. 142, pp. 66-70.

Fan, M., Fan, Z. and Zhang, J. (2008) 'The Impact of Poor Health on Wage Losses in China', *IITA, Second International Symposium on Intelligent Information Technology Application,* Vol. 3, pp. 872-875.

Food and Argriculture Organization of the United Nations, World Health Organization, and United Nations University (2001) *Human Energy Requirements*. Food and Nutrition Technical Report Series 1, Chapter 5. Available: ftp://ftp.fao.org/docrep/fao/007/y5686e/y5686e00.pdf. [cited 17th Feb 2011].

Freiberg, M.S., Cabral, H.J., Heeren, T.C., Vasan R.S. and Ellison, R.C. (2004) 'Alcohol Consumption and the Prevalence of the Metabolic Syndrome in the U.S.: A Cross-section Analysis of Data from the Third National Health and Nutrition Examination Survey', *Diabetes Care*, Vol. 27, No. 12, pp. 2954-2959. Available: http://care.diabetesjournals.org/content/27/12/2954.full. [cited 23rd Aug 2013].

French, M. T. and Zarkin, G. A. (1995) 'Is moderate alcohol use related to wages? Evidence from four worksites', *Journal of Health Economics*, Vol. 14, No. 3, pp. 319-344.

Gambin, L. (2005) 'The impact of health on wages in Europe – does gender matter?', *Health, Econometrics and Data Group (HEDG) Working Papers*, No. 05/03. University of York. Department of Economic. Available: http://www.york.ac.uk/res/herc/research/hedg/wp.htm. [cited 4th Aug 2010].

Gao, W. and Smyth, R. (2009) 'Health Human Capital, Height and Wages in China', *Monash University Discussion Paper,* No. 05/09, Department of Economics. Available:
http://www.buseco.monash.edu.au/eco/research/papers/2009/0509healthhumangaosmyth.pdf. [cited 19th Aug 2010].

Gao, Y., Griffiths, S. and Chan, E. (2007) 'Community-based Interventions to Reduce Overweight and Ovesity in China: A Systematic Review of the Chinese and English Literature', *Journal of Public Health*, Vol. 30, No. 4, pp. 436-448.

Garcia-Gomez, P. (2011) 'Institutions, Health Shocks and Labour Market outcomes', *Journal of Health Economics*, Vol. 30, No. 1, pp. 200-213.

Garcia-Gomez, P. Jones, A. and Rice, N. (2010) 'Health Effects on Labour Market Exits and Entries', *Labour Economics*, Vol. 17, No. 1, pp. 62-76.

Garcia-Gomez, P. and Lopiz-Nicolas, A. (2006) 'Health Shocks, Employment and Income in the Spanish Labour Market', *Health Economics*, Vol. 15, No. 9, pp. 997-1009.

Gonçalves, A., (2006) 'China's Swing from a Planned Soviet-Type Economy to an Ingenious Socialist Market Economy: An Account of 50 Years'. Available: http://ssrn.com/abstract=949371. [cited 11st Aug 2011].

Greene, W. H. (2011) *Econometric Analysis*, 7th ed, Prentice Hall. Chapter 17.

Gregory, C. A. and Ruhm, C. J. (2009) 'Where Does the Wage Penalty Bite?', *NBER Working Paper*, No. 1498. Available:
http://www.nber.org/papers/w14984.pdf?new_window=1. [cited 11st Aug 2010].

Greve, J. (2008) 'Obesity and Labour Market Outcomes in Denmark', *Economics and Human Biology*, Vol. 6, No. 3, pp. 350-362.

Grossman, M. (1972a) *The Demand for Health: A Theoretical and Empirical Investigation*. New York: Columbia University Press for the National Bureau of Economic Research.

Grossman, M. (1972b) 'On the Concept of Health Capital and the Demand for Health', *Journal of Political Economy*, Vol. 80, No. 2, pp. 223-255.

Grossman, M. and Benham, L. (1974) *Health, hours and wages.* In: Mark Perlman, Editor, The Economics of Health and Medical Care, Macmillan and Co., London.

Gujarati, D.N. (1995) *Basic Econometrics.* 3rd ed. London: McGraw-Hill.

Hamilton, V. And Hamilton, B. H. (1997) 'Alcohol and Earnings: Does Drinking Yield a Wage Premium?', *The Canadian Journal of Economics*, Vol. 30, No. 1, pp. 135-151.

Hansen, L., Healton, J. and Yaron, A. (1996) 'Finite-Sample Properties of Some Alternative GMM Estimators', *Journal of Business and Economic Statistics*, Vol. 14, No. 3, pp. 262-280.

Hao, W., Derson, Y., Xiao, S., Li, L. and Zhang, Y. (1999) 'Alcohol Consumption and Alcohol-related Problems: Chinese Experience from Six Area Samples', *Addiction*, Vol. 94, No. 10, pp. 1467-1476.

Hao, W. and Young, D. (1995a) 'Drinking in China: present situation, trends and counter-measures', *Chinese Journal of Clinical Psychology*, No. 3, pp. 243-248.

Hao, W., Young, D., Li, L., Xiao, S., Tu, J., Shen, H., Yu, Z., Teng, X. and Wan, X. (1995b) 'Psychoactive Substance Use in Three Sites in China: Gender Differences and Related Factors', *Addiction*, Vol. 90, No. 11, pp. 1503-1515.

Haskins, K. M. and Ransford, H.E. (1999) 'The Relationship between Weight and Career Payoffs among Women', *Sociological Forum*, Vol. 14, No. 2, pp. 295-318.

Haveman, R., Wolfe, B., Kreider, B. and Stone, M. (1994) 'Market Work, Wages, and Men's Health', *Journal of Health Economics*, Vol. 13, No. 2, pp. 163–182.

Hausman, J. (1978) 'Specification Tests in Econometrics', *Econometrica*, Vol. 46, No. 6, pp. 1251-1271.

Heckman, J., J. (1981) 'The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process', In *Structural Analysis of Discrete Data with Econometric Applications*, ed. Manski, C., F. and McFadden, D., L., Cambridge: The MIT Rress. Chapter 4. Avalible: http://elsa.berkeley.edu/~mcfadden/discrete/ch4.pdf. [cited 22nd Jul 2013].

Heineck, G. and Schwarze, J. (2003) 'Substance Use and Earnings: The Case of Smokers in Germany', *IZA Discussion Papers*, No. 743. Available: ftp://ftp.iza.org/dps/dp743.pdf. [cited 27th May 2013].

Hew, C. (2006) 'Healthcare in China: Toward Greater Access, Efficiency and Quality ', *IBM Institute for Business Value.* IMB Business Consulting Services. Available: http://www-935.ibm.com/services/us/imc/pdf/g510-6268-healthcare-china.pdf. [cited 4th Apr 2012].

Jackobson, L. and Pursiainen, C. (2001) 'At the Crossroads of Post-Communist Modernisation: Russia and China in Comparative Perspective', *UPI Working Papers*, No. 30. Available: http://se1.isn.ch/serviceengine/Files/ISN/19292/ipublicationdocument_singledocument/089DE55C-9CAE-4564-B2E1-C8A1B1586634/en/WP30.pdf. [cited 16th Aug 2011].

Joffe, E. (1966) 'China in Mid-1966: "Cultural Revolution" or Struggle for Power?' *The China Quarterly*, No. 27, pp. 123-131.

Johansson, E., Bockerman, P., Kiiskinen, U. and Heliovaara, M. (2009) 'Obesity and Labour Market Success in Finland: The Difference between Having a High BMI and Being Fat', *Economics and Human Biology*, Vol. 7, No. 1, pp. 36-45.

Jones, A. (2007) *Applied econometrics for health economists: a practical guide.* 2nd ed. Oxford: Radcliffe. Chapter 2. Available: http://www.york.ac.uk/res/herc/teach/OHE-Booklet.pdf. [cited 4th May 2010].

Jones, M. K. (2011) 'Disability, Employment and Earnings: An Examination of Heterogeneity', Applied Economics, Vol. 43, No. 8, pp. 1001-1017.

Kallerg, H., Jacobsen, S., Bengtsson, C., Pedersen, M., Padyukov, L., Garred, P., Frisch, M., Karlson, E.W., Klareskog, L. and Alfredsson, L. (2009) 'Alcohol Consumption is Associated with Decreased Risk of Rheumatoid Arthritis: Results from Two Scandinavian Case-control Studies', *Annals of the Rheumatic Diseases*, Vol. 68, No. 2, pp. 222-227. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2937278/#!po=3.84615. [cited 23rd Aug 2013].

Kenkel, D. S. and Ribar, D. C. (1994). 'Alcohol Consumption and Young Adults' Socioeconomic Status', *Brookings Papers on Economic Activity. Microeconomics*. Vol. 1994, pp.119-175.

Kidd, M.P. and Meng, X. (2001) 'The Chinese State Enterprise Sector: Labour Market Reform and the Impact on Male-Female Wage Structure', *Asian Econommic Journal*, Vol. 15, No. 4, pp. 405-423.

Kline, B. and Tobias, J. L. (2008) 'The wages of BMI: Bayesian analysis of a skewed treatment- response model with non-parametric endogeneity', *Journal of Applied Economics*, Vol. 23, No. 6, pp. 767-793.

Kostovic, K. and Lipozencic, J. (2004) 'Skin Diseases in Alcoholics', *Acta Dermatovenerologica Croatica: ADC*, Vol. 12, No. 3, pp. 181-190.

Kristen, E. (2002) 'Addressing the problem of weight discrimination in employment', *California Law Review*, Vol. 90, No. 1, pp. 57-109.

Labour Law of the People's Republic of China. Chapter II, Article 15. Available: http://www.npc.gov.cn/englishnpc/Law/2007-12/12/content_1383754.htm. [cited 26th March 2010].

Last, J. M. (1998) *Public Health and Human Ecology*, Prentice Hall, London. Chapter 6.

Lee, L. (1981) *Simultaneous Equations Models with Discrete and Censored Dependent Variables*, in Structural Analysis of Discrete Data with Econometric Applications, ed. C. F. Manski and D. McFadden, Cambridge: MIT Press.

Lee, L. (1982) 'Health and Wages: a Simultaneous Equation Model with Multiple Discrete Indicators', *International Economic Review*, Vol. 23, No. 1, pp. 199–221.

Lee, L. (1992) 'Amemiya's Generalized Least Squares and Tests of Overidenfication in Simultaneous Equation Models with Qualitative or Limited Dependent Variables', *Econometric Reviews*, Vol. 11, No. 3, pp. 319-328.

Lei, X. and Lin, W. (2009) 'The New Cooperative Medical Scheme in Rural China: Does More Coverage Mean Service and Better Health?', *Health Economics*, Vol. 18, No. S2, pp. S25-S46.

Leo, P. (2006) 'In China, Smoking's the Big Thing'. Available:

http://www.post-gazette.com/stories/local/morning-file/in-china-smokings-the-big-thing-432434/. [cited 17th May 2013].
Levine, P. B., Gustafson, T. A., and Velenchik, A. D. (1995) 'More Bad News for Smokers? The Effects of Cigarette Smoking on Labour Market Outcomes', *NBER Working Paper*, No. 5270. Available:
http://www.nber.org/papers/w5270.pdf?new_window=1. [cited 05/03/2013].

Levine, P. B., Gustafson, T. A., and Velenchik, A. D. (1997) 'More Bad News for Smokers? The Effects of Cigarette Smoking on Labour Market Outcomes', *Industrial and Labour Relation Review*, Vol. 50, No. 3, pp.493-509. Available:
http://www.utdallas.edu/~imcfarlin/ECO_3304/Resources/Ch3_SmokingWages.pdf. [cited 05/03/2013].

Levy, D. (2006) 'The role of public policies in reducing smoking and deaths caused by smoking in China: results from the China Tobacco Policy Simulation Model', New York: World Health Organization.

Li, H. and Zax, J. (2003) "Economic Transition and the Labor Market in China." *Working paper, School of Economics, Georgia Institute of Technology.* Available: http://www.prism.gatech.edu/~hl45/research/lbmkt.pdf. [cited 28th Oct 2010].

Li, L. and Dong, X. (2008) 'The Role of Firm Effects in the Gender Earnings Gap: Evidence from China.' *Economic Research*, No. 9, pp.122-135.

Li, L. and Wu X. (2011) 'Gender of Children, Bargaining Power, and Intrahousehold Resource Allocation in China', *University of Wisconsin Press*, Vol. 46, No. 2, pp. 295-316.

Lin, T. and Chen, J. (2010) 'Smoking and Labour Market Outcomes: Empirical Evidence from Taiwan', *Population Association of America 2010 Annual Meeting Progarm*, Poster Session 6. Available:
http://paa2010.princeton.edu/abstracts/100163. [cited 28/02/2013].

Lin, W., Liu, G. G., and Chen, G. (2009) 'The Urban Resident Basic Medical Insurance: a landmark reform towards universal coverage in China?', *Health Economics*, Vol. 18, No. S2, pp. S83-S96.

Liu, G. G., Dow, W. H., Fu, A. Z., Akin, J. and Lance, P. (2008) 'Income productivity in China: On the role of health', *Journal of Health Economics*, Vol. 27, No. 1, pp. 27-44.

Liu, G. G. and Zhao, Z. (2006) 'Urban employee health insurance reform and the impact on out-of-pocket payment in China', *The International Journal of Health Planning and Management*, Vol. 21, No. 3, pp. 211-228.

Liu, Y. (2002) 'Reforming China's urban health insurance system', *Health Policy*, Vol. 60, No. 2, pp. 133-150.

Liu, Y. (2010) 'Great Leap Forward and Chinese Famine of 1958-1961: State, Collective and Peasants in Centralized System' *China Economic Quarterly*, Vol. 3, pp. 1119-1142.

Lu, X., Luo, J., Tao, M., Gen, Y., Zhao, P., Zhao, H., Zhang, X. and Dong, N. (2004) 'Risk Factors for Alcoholic Liver Disease in China', *World Journal Gastroenterol*, Vol. 10, No. 16, pp. 2423-2426.

Lucas, R. (1988) 'On the Mechanics of Economic Development', *Journal of Monetary Economics*, Vol. 22, No. 1, pp. 3-12.

Luft, H. (1975) 'The Impact of Poor Health on Earnings', *The Review of Economic and Statistics*, Vol. 57, No. 1, pp. 43-57.

Lv, W. (2009) 'The Distribution of Medical Insurance – Based on CHNS 1989-2006 Micro-level Data', Journal of Shanxi Finance and Economics University, Vol. 31, No. 8, pp. 23-33.

Lye, J. and Hirschberg, J. (2004) 'Alcohol Consumption, Smoking and Wages', *Applied Economics*, Vol. 36, No. 16, pp. 1807-1817.

MacDonald, Z. and Shields, M. A. (2001) 'The Impact of Alcohol Consumption on Occupational Attainment in England', *Economica*, Vol. 68, No. 271, pp. 427-453. Available: http://onlinelibrary.wiley.com/doi/10.1111/1468-0335.00254/pdf. [cited 7th Jun 2013].

Marmot, M. and Brunner, E. (1991) 'Alcohol, and Cardiovascular Disease: The Status of the U Shaped Curve', *British Medical Journal*, Vol. 303, No. 6802, pp. 565-568.

Mattson, M. E., Pollack, E. S., and Cullen, J. W. (1987) 'What are the Odds that Smoking Will Kill You', *American Journal of Public Health*, Vol. 77, No. 4, pp. 425-431.

Meng, X., Gong, X. and Wang, Y. (20094) 'Impact of Income Growth and Economic Reform on Nutrition Intake in Urban China: 1986-2000', *IZA Discussion Papers*, No. 1148. Available: http://ftp.iza.org/dp1448.pdf. [cited 8th Feb 2011].

Miao, X., and Wang, C. (2010) 'Ahead of World Expo, China Acts to Promote Smoking Controls', Xihua News Agency. Available: http://news.xinhuanet.com/english2010/china/2010-03/10/c_13205579.htm. [cited 17th May 2013].

Mullahy, J. and Sindelar, J. L. (1993) 'Alcoholism, Work, and Income', *Journal of Labour Economics*, Vol. 11, No. 3, pp. 494-520.

Myllykangas-Luosujarvi, R., Kimmo, A. Kautiainen, H., and Hakala, M. (2000) 'Reduced Incidence of Alcohol Related Deaths in Subjects with Rheumatoid Arthritis', *Annals of the Rheumatic Diseases*, Vol. 59, No. 1, pp. 75-76. Available: http://ard.bmj.com/content/59/1/75.long. [cited 23rd Aug 2013].

Misra, A., Wasir, J. S. and Vikram, N. K. (2005) 'Waist Circumference Criteria for The Diagnosis of Abdominal Obesity are Not Applicable Uniformly to All Populations and Ethnic Groups', *Nutrition*, Vol. 21, No. 9, pp. 969-976.

Morris, S. (2004) 'The Impact of Obesity on Employment in England', *Tanaka Business School Discussion Papers*, No. TBS/DP04/4. Available: http://www3.imperial.ac.uk/pls/portallive/docs/1/39668.PDF. [cited 14th Mar 2012].

Morris, S. (2007) 'The Impact of Obesity on Employment', *Labour Economics*, Vol. 14, No. 3, pp. 413-433.

Morz, T. (1987) 'The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions', *Econometrica*, Vol. 55, No. 4, pp. 764-799.

Mullahy, J. and Sindelar, J. (1996) 'Employment, Unemployment, and Problem Drinking', *Journal of Health Economics*, Vol. 15, No. 4, pp.409-434.

Mushkin, S. (1962) 'Health as an Investment', *Journal of Political Economy*, Vol. 70, No. 5, pp. 129-157.

National Heart, Lung and Blood Institute. (1998) 'Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults', *National Institutes of Health (NIH) Publication*, No. 98-4083, New York: National Institutes of Health. Available: http://www.ncbi.nlm.nih.gov/books/NBK2003/pdf/TOC.pdf. [cited 8th May 2012].

Nerlove, M., Press, S. J. (1973) *Univariate and Multivariate Log-Linear and Logistic Models*, Santa Monica: Rand R-1306-EDA/NIH.

Neumann, T. C. (2009) 'The Effects of Drinking and Smoking on the Labour Market Outcomes of Low Income Young Adults', Applied Economics, Vol. 45, No. 5, pp. 541-553.

Newey, W. K. (1987) 'Efficient Estimation of Limited Dependent Variable Models with Endogeneous Explanatory Variables', *Journal of Econometrics*, Vol. 36, No. 3, pp. 231-250.

Norton, E. C. and Han, E. (2007) 'Genetic Information, Obesity, and Labour Market Outcomes', *Health, Econometrics and Data Group (HEDG) Working Papers*, No. 07/15. University of York. Department of Economic. Available: http://www.york.ac.uk/res/herc/documents/wp/07_15.pdf. [cited 23th Mar 2012].

Norton, E. C. and Han, E. (2008) 'Genetic Information, Obesity, and Labour Market Outcomes', *Health Economics*, Vol. 17, No. 9, pp. 1089-1104.

Nuorti, J.P., Butler, J.C., Farley M.M., Harrison, L.H., McGeer, A., Kolczak, M.S. And Breiman R.F. (2000) 'Cigarette Smoking and Invasive Pneumococcal Disease. Active Bacterial Core Surveillance Team', *New England Journal of Medicine*, Vol. 342, No. 10, pp. 681-689.

O'Leary, N. and Sloane, P. (2006) 'Rates of Return to Degrees across British Regions', *IZA Discussion Papers*, No. 1947. Available: http://ftp.iza.org/dp1947.pdf. [cited 4th Oct 2011].

Pan, J. and Qin, X. (2010) 'Does Body Size Matter in Urban Employment? Evidence from China', *School of Economics Peking University Working Papers*, No. C-2010-06-003, pp. 1-19. Available: http://econ.pku.edu.cn/upload/20100728/3418043.pdf. [Chinese Version]. [cited 21st May 2012].

Pan Z. and Hu D. (2008) 'Hierarchical linear modelling of smoking prevalence and frequency in China between 1991 and 2004', *Health Policy and Planning*, Vol. 23, No. 2, pp. 118-124.

Park, S.H., Li, S. and Tse, D.K. (2006) 'Market liberalization and firm performance during China's economic transition', *Journal of International Business Studies*, Vol. 37, No. 1, pp. 127-147.

Parrott, A.C. (1999) 'Does Cigarette Smoking Cause Stress?', *American Psychologist*, Vol. 54, No. 10, pp. 817-820.

Peate, I. (2005) 'The Effect of Smoking on the Reproductive Health of Men', *British Journal of Nursing (Mark Allen Publishing)*, Vol. 14, No. 7, pp. 362-365.

Popkin, B. M., Du, S., Zhai, F. and Zhang, B. (2009) 'Cohort Profile: The China Health and Nutrition Survey--Monitoring and Understanding Socio-Economic and Health Change in China, 1989-2011', *International Journal of Epidemiology*, Available: http://ije.oxfordjournals.org/cgi/content/full/dyp322v1#T4. [cited 10th March 2010].

Pelkowski, J.M. and Berger, M.C. (2004) 'The Impact of Health on Employment, Wages, and Hours Worked over the Life Cycle', *The Quarterly Review of Economics and Finance*, Vol. 44, No. 1, pp. 102-121.

Qian, J., Cai, M., Gao, J., Tang, S., Xu, L. and Critchley, J. (2010) 'Trends in Smoking and Quitting in China from 1993 to 2003', *Bulletin of the World Health Organization*, Vol. 88, No. 10, pp. 769-776. Available: http://www.who.int/bulletin/volumes/88/10/09-064709.pdf. [cited 17th May 2013].

Qiu, T. (2007a) 'The Adjusted Measure of Body Mass Index for the Chinese and its Impact on Health', *Munich Personal RePEc Archive (MPRA) Paper*, No. 6720. Available: http://mpra.ub.uni-muenchen.de/6270/1/MPRA_paper_6270.pdf. [cited 2th April 2010].

Qiu, T. (2007b) 'Private Returns to Education: Earnings, Health and Well-being' Doctor of Philosophy (PhD). Thesis. University of Bath. Available: http://opus.bath.ac.uk/12283/1/Qiu_PHD.pdf. [cited 25th April 2011].

Qureshi, A.A., Dominguez, P.L., Choi, H.K., Han, J. and Curhan, G. (2010) 'Alcohol Intake and Risk of Incident Psoriasis in US Women', *JAMA Dermatology*, Vol. 146, No. 12, pp. 1364-1369. Available: http://archderm.jamanetwork.com/article.aspx?articleid=422554. [cited 23rd Aug 2013].

Rice, D. P., Kelman, S. and Miller, L. S. (1991). 'Estimates of Economic Costs of Alcohol and Drug Abuse and Mental Illness, 1985 and 1988', *Public Health Reports*, Vol. 106, No. 3 pp. 280-292 .

Rosenbaum, P. R., and Rubin, D. B. (1983) 'The central role of the propensity score in observational studies for causal effects', *Biometrika*, Vol. 70, No.1, pp. 41-55.

Russell S. W. (1981), "Attrition Bias in the Estimation of Econometric Models From Panel Data", *Association for Consumer Research*, Advances in Consumer Research Volume 08, pp. 220-226. Available:
http://www.acrwebsite.org/search/view-conference-proceedings.aspx?Id=9814.

Savage, D. and Lindenbaum, J. (1986) 'Anemia in Alcoholics', *Medicine (Baltimore)*, Vol, 65, No. 5, pp. 322-338.

Schultz, T. W. (1961) 'Investment in Human Capital', *American Economic Review*, Vol. 51, No. 1, pp. 1-17.

Shaper, A. G. (1993) 'Alcohol, the Heart, and Health', *American Journal of Public Health*, Vol. 3, No. 6, pp. 799-801.

StataCorp. (2009) *Stata 11 Longitudinal Data/Panel Data Reference Manual*. College Station, TX: StataCorp LP.

State Council Provisional Regulations on Retirement and Resignation of Workers. Article 1. *State Council*, 2th Jun, 1978. Available:
http://www.labournet.com.cn/english/fagui/19990803_1.htm. [cited 27th March 2010].

Stern, S. (1989) 'Measuring the Effect of Disability on Labour Force Participation', *Journal of Human Resources*, Vol. 24, No. 3, pp.361-395.

Swart, J. C. (1990) 'An Overlooked Cost of Employee Smoking', *Personnel*, Vol. 67, No. 8, pp. 53-55.

Tang, Y., Xiang, X., Wang, X., Cubells, J. F., Babor, T. F. and Hao, W. (2013) 'Alcohol and Alcohol-related Harm in China: Policy Changes Needed', *Bulletin of the World Health Organization*, Vol. 91, No. 4, pp. 237-312.

Tekin, E. (2002) 'Employment, Wages, and Alcohol Consumption in Russia: Evidence from Panel Data', *IZA Discussion Papers*, No. *432*. Available:
ftp://repec.iza.org/RePEc/Discussionpaper/dp432.pdf. [cited 18th Oct 2010].

Thomas, D. and Strauss, J. (1997) 'Health and wages: Evidence on men and women in urban Brazil', *Journal of Econometrics*, Vol. 77, No. 1, pp. 159-185.

Thompson, W. (2010) 'China's Rapidly Aging Population', *Population Reference Bureau*, No. 20. Available:
http://www.prb.org/pdf10/TodaysResearchAging20.pdf. [cited 1st Jan 2012].

Tuan. N. T., Adair L. S., He, K., and Popkin, B. M. (2008) 'Optimal Cutoff Values for Overweight: Using Body Mass Index to Predict Incidence of Hypertension in 18- to 65-Year-Old Chinese Adults', *The Journal of Nutrition*, Vol. 138, No. 7, pp. 1377-1382.

Van de Poel, E., O'Donnell, O. and Van Doorslaer, E. (2009) 'The Health Penalty of China's Rapid Urbansization', *Tinbergen Institute Discussion Paper*, No. 09-016/3. Available:
http://www.tinbergen.nl/discussionpapers/09016.pdf. [cited 27th Mar 2012].

Van Ours, J. C. (2002) 'A Pint a day Raises a Man's Pay; But Smoking Blows that Gain Away', *IZA Discussion Papers*, No. 473. Available:
http://ftp.iza.org/dp473.pdf. [cited 3rd Jun 2013].

Walker, I. and Zhu, Y. (2003) 'Education, Earnings and Productivity: Recent UK Evidence', *Labour Market Trends*, Vol. 111, No. 3, pp. 145-152.

Wei, Z. (2004) 'The Role of Health on Off-farm Employment and Wage Decision', *Economic Research Journal*, Vol. 2, pp. 64-74. [Chinese Version].

Wetterling, T. and Junghanns, K. (2000) 'Psychopathology of Alcoholics during Withdrawal and Early Abstinence', *European Psychiatry*, Vol, 15, No. 8, pp. 483-488.

Williams, R. (2011a) 'Panel Data 3: Conditional Logit/Fixed Effects Logit Models', University of Notre Dame, Department of Sociology, Soc 73994. Available:
http://www.nd.edu/~rwilliam/stats3/Panel03-FixedEffects.pdf. [cited 9th Feb 2012].

Williams, R. (2011b) 'Panel Data 4: Fixed Effects vs Random Effects Models', University of Notre Dame, Department of Sociology, Soc 73994. Available:
http://www.nd.edu/~rwilliam/stats3/Panel04-FixedVsRandom.pdf. [cited 9th Feb 2012].

Wong, C., Tang, K. and Lo, V. (2007) 'Unaffordable healthcare amid phenomenal growth: the case of healthcare protection in reform China', *International Journal of Social Welfare*, Vol. 16, No.2, pp. 140-149.

Worden, R.L., Savada, A.M. and Ronald E. Dolan, editors. *China: A Country Study*. Washington: GPO for the Library of Congress, 1987. Available:
file:///E:/papers/china_labour_market/China.htm. [cited 15th Aug 2011].

World Health Organization (2000) 'Obesity: Preventing and Managing the Global Epidemic', WHO Technical Report Series 894, Geneva: World Health Organization.

World Health Organization (2003) 'Diet, Nutrition and the Prevention of Chronic Diseases'. WHO Technical Report Series 916. Chapter 5. Available:
http://whqlibdoc.who.int/trs/WHO_TRS_916.pdf. [cited 18th Feb 2011].

Xu, Y. (2007) 'Jintao Hu : Speeding Up the Establishment of Social Security System and Securing the Basic Livings of People'. Available:
http://www.china.com.cn/17da/2007-10/15/content_9055682.htm. [Chinese Version]. [cited 21st Jun 2012].

Yang F., Lv J., Lei S., Chen X., Liu M., Jian W., Xu H., Tan L., Deng F., Yang Y., Wang Y. Sun X., Xiao S., Jiang C., Guo Y., Guo J., Li Y., Zhu X., Papasian C. J. and Deng H. (2006) 'Receiver-Operating Characteristic Analyses of Body Mass Index, Waist Circumference and Waist-To-Hip Ratio for Obesity: Screening in Young Adults in Central South of China', *Clinical Nutrition*, Vol. 25, No. 6, pp. 1030-1039.

Yoon, Y.S., Oh, S.W., Baik, W.H., Park, H.C. and Kim, W.Y. (2004) 'Alcohol Consumption and the Metabolic Syndrome in Korean Adults: the 1998 Korean National Health and Nutrition Examination Survey', *American Journal of Clinical Nutrition*, Vol. 80, No. 1, pp. 217-224. Available: http://ajcn.nutrition.org/content/80/1/217.full. [cited 23rd Aug 2013].

Yueh, L., 2004, 'Wage Reforms in China during the 1990s', *Asian Economic Journal,* Vol. 18, No. 2, pp. 149-164.

Zarkin, G. A., French, M. T., Mroz, T. And Bray, J. W. (1998) 'Alcohol use and wages: New results from the national household survey on drug abuse', *Journal of Health Economics*, Vol. 17, No. 1, pp. 53-68.

Zhang, G. (2009) 'The Analysis of the Background of Corporate Governance in China', Available: http://www.govyi.com/lunwen/2009/200903/301440.shtml. [Chinese Version]. [cited 10th Aug 2011].

Zhang, J. (2003) 'Nutrition, Health and Productivity – Evidence from Poor Areas of Rural China', *Economic Research Journal*, Vol. 1, pp. 3-12. [Chinese Version].

Zhang, W., Zhao, J., Yu, J., Wu, F., Liu, W., Zhou, Z. and Guan, D. (2007) 'Alcohol and Other Contributing Factors among Drivers Who Died in Traffic Accidents', *Chinese Journal of Forensic Medicine*, Vol. 22, No. 2, pp. 120-123.

Zhang, X., Zhao, X. and Harris, A. (2009) 'Chronic Diseases and Labour Force Participation in Australia', *Journal of Health Economics*, Vol. 28, No. 1, pp. 91-108.