

PLACE IN RETURN BOX to remove this checkout from your record.
 TO AVOID FINES return on or before date due.
 MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
SEP 25 2005		
SEP 25 2007		
JAN 07 2008		
12 14 07		

THREE ESSAYS ON HEALTH AND MACRONUTRIENT CONSUMPTION
AMONG CHINESE ADULTS

By
ZHEHUI LUO

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
DEPARTMENT OF ECONOMICS

2003

ABSTRACT

THREE ESSAYS ON HEALTH AND MACRONUTRIENT CONSUMPTION
AMONG CHINESE ADULTS

By
ZHEHUI LUO

The first chapter contains an in-depth analysis of the socioeconomic determinants of adult (twenty years of age or older) body mass index (BMI) in China using the China Health and Nutrition Survey (CHNS) 1989-1997. There is a dramatic increase in overweight and obesity among adult Chinese in this period (for men 6% to 17% and for women 11% to 21%). A clear theoretical framework was utilized and descriptive and multivariate reduced form and dynamic demand function analyses were undertaken to find a variety of factors affecting individual BMI. The key findings are that for women the effect of education is very strong and inversely U-shaped. Productive assets, food prices, water sources and sanitation conditions are all important determinants of adult BMI in China and these factors affect men and women in different age groups and across regions differently. With careful exposition of the socioeconomic determinants of adult BMI in China in the 1990s, the paper finds strong protective effects of education on BMI in women, particularly in rural areas. The knowledge of this relationship can assist public policy makers to identify target groups for improving their health status. As the Chinese economy undergoes rapid structural transition it is extremely important to find factors that can make such transition as smooth as possible.

The second chapter studies the associations between adult food consumption and several socioeconomic factors such as education level, household resources, and community characteristics in the early 1990s in China. In the overall sample education does not have significant impact on calorie intakes but does affect percent of calories

from fat, from protein and from carbohydrates differently in different region and at different age. The effect of productive assets is nonlinear and in inverted U-shape for male calorie, fat and protein intakes; whereas for women more productive assets are associated with more fat and protein intakes and more percents of calories from fat and from protein. In rural areas the effect of productive assets is stronger than that in the urban areas. Prices of foods, community water and sanitation conditions are also studied. The effects of prices on calorie, fat and protein intakes and the quality of diet measures can go in either direction. Improvements in sanitation are associated with more energy and protein intakes in urban areas. A simple health production function analysis on weight and BMI is carried out and both health measures in the short period of two years can be described as a random walk process.

The third chapter presents identification of the shape of the age, cohort and time effect profiles of male and female BMIs in 1990s in China. The analysis is used to help pin down the model specification in the BMI demand functions later. The age profile for women is of inverse U-shape. The year effect for men is strong. There are not enough data to identify cohort effects due to the short length of the survey. In our main analysis for socioeconomic determinants of BMI we will be using the five-year cohort identification strategy.

Copyright by
ZHEHUI LUO
2003

To my family and friends

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor John Strauss, for all the time and encouragement he gave me. Without his patient guidance this dissertation would not have been done. I aspire to emulate his diligence in research and love for developing countries. He taught me that many difficulties in life can be overcome by a strong will.

I am also in great debt to my “boss”, Professor Joseph Gardiner, who throughout the years supported me financially and emotionally. His believing in me made my transition from a child to a scholar easier.

I thank my other committee members, Professors John Goddeeris and John Giles, who took the time reading my drafts and made many helpful comments and suggestions. I also wish to gratefully acknowledge my appreciation to Professors Jeffrey Wooldridge and Steven Haider at MSU for suggestions made at various stages of the research, and to Dr. Barry Popkin at CPC-UNC for providing the data used in this dissertation. Thanks are due to professors who taught me various courses in economics, especially Professor Warren Samuels, who benefited me greatly with his profound knowledge and sense of history and showed me kindness beyond words.

I have been privileged to know many great people during my long years as a graduate student at Michigan State University. Some are my teachers in my academic pursuit and in my life, some are friends by my side through ups and downs, and some have become like family whom I trust and love deeply. Many friends enriched my life on campus, Facundo, Jing-I, Daiji, Ren, Pungpond and Firman, in church, Qiang, Xiangshu, Shuangwen, Dora and many more from the Lansing Chinese Christian Ministry. Some friends in China are still as close to my heart as ever: Sier, Zhiyong, Fengchen, and Yongfeng.

I give special thanks to my family, my mother and my sister, for their unconditional love and unfailing support, without which it would have been impossible for me to

survive.

Finally, to James, whose humor, intelligence, good heart and love captivated me, this dissertation is dedicated.

Table of Contents

LIST OF TABLES	xii
-----------------------	------------

LIST OF FIGURES	xvi
------------------------	------------

1 Socioeconomic Determinants of Body Mass Index of Adult Chinese in the 1990s	1
1.1 Introduction	1
1.2 Theoretical Framework	4
1.2.1 The Model	5
1.2.2 Estimation Strategies	10
1.3 Literature Review	12
1.4 Data and Descriptive Statistics	18
1.4.1 The Data	18
1.4.2 Summary Statistics	24
1.5 Reduced Form Analyses	39
1.5.1 Education	41
1.5.2 Household Resources	46
1.5.3 Marital Status	48
1.5.4 Community Characteristics	50
1.6 Dynamic Conditional BMI Demand	71
1.6.1 With Community Dummies	72

1.6.2	With Community Characteristics	73
1.7	Concluding Remarks	87
2	Adult Chinese Macronutrient Consumption and Socioeconomic Determinants in the Early 1990s	91
2.1	Introduction	91
2.2	Literature Review	97
2.2.1	Reduced Form Demands	97
2.2.2	Descriptive Studies	98
2.2.3	Income and Nutrients	102
2.2.4	Price and Nutrients	105
2.2.5	Education and Nutrients	107
2.3	Data and Econometric Issues	108
2.3.1	The Data	108
2.3.2	Measurement Errors	112
2.3.3	Patterns and Trends	115
2.3.4	Estimation Methods	118
2.4	Results: Determinants of Levels	123
2.4.1	Basic Models	123
2.4.2	Augmented Models	128
2.5	Results: Determinants of Changes	147
2.5.1	Basic Models	147

2.5.2	Augmented Models	149
2.6	Results: Production Function	154
2.7	Discussions	158
3	Age, Cohort and Year Analysis in the Socioeconomic Determinants of Adult BMI in the 1990s	178
3.1	Introduction	178
3.2	Identification of Second Differences	179
3.3	Simulation Study	184
3.3.1	Models with linear and/or squared terms	185
3.3.2	Models with dummy variables	187
3.3.3	Profiles for y and y_1	189
3.4	BMI Profiles in CHNS	210
3.5	Estimate the Restricted Level Effects	211
3.6	Concluding Remarks	221
	APPENDIX	222
	A Survey instruments in CHNS	223
	B First Stage Regressions for Section 1.6	229
	C Linear Probability Models for Undernourishment and Overweight	237
	D AR1 Model of Determinants of BMI	245

E	Descriptive Statistics of Nutrient Intakes in CHNS89-93	246
F	First Stage Regressions for section 2.6	250
G	The Design Matrix and Constraints Needed	255

List of Tables

1.1	Adult BMI in CHNS 89-97	25
1.2	Prevalence of Adult High Blood Pressure in CHNS 1989-1997	31
1.3	Adult Smoking Status in CHNS 1991-1997	33
1.4	Basic Characteristics of Households and Individuals in CHNS 1989-97	36
1.5	Determinants of Adult BMI in CHNS 1989-97: Overall	53
1.6	Determinants of Adult BMI in CHNS 1989-97: Urban Areas	55
1.7	Determinants of Adult BMI in CHNS 1989-97: Rural Areas	57
1.8	Determinants of Adult BMI in CHNS 1989-97: Age 20 - 39	59
1.9	Determinants of Adult BMI in CHNS 1989-97: Age 40 - 59	61
1.10	Determinants of Adult BMI in CHNS 1989-97: Age 60+	63
1.11	Adult BMI by CHNS Round 1991, 93, and 97: Overall	65
1.12	Determinants of Adult BMI in CHNS 1989-93 with Community Char- acteristics: Overall	67
1.13	Determinants of Adults BMI in 1993 Conditional on BMI in 1991: Overall	75
1.14	Determinants of Adults BMI in 1997 Conditional on BMI in 1993: Overall	77
1.15	Determinants of 1993 BMI Conditional on 1991 BMI with Changes in Community Characteristics	79
1.16	Determinants of 1997 BMI Conditional on 1993 BMI with Changes in Community Characteristics	83
2.1	Individual, Household and Community Characteristics in CHNS 89-93	113

2.2	Average Daily Calorie, Fat, Protein and Carbohydrates Intakes: Patterns and Trends in CHNS 89-93	114
2.3	Average Percent of Calorie from Fat, Protein and Carbohydrates: Patterns and Trends in CHNS 89-93	117
2.4	Daily Calorie Intakes (Kcal) in CHNS 89, 91, 93 From All Food Groups	131
2.5	Daily Fat Intakes (gram) in CHNS 89, 91, 93 From All Food Groups .	133
2.6	Daily Protein Intakes (gram) in CHNS 89, 91, 93 From All Food Groups	135
2.7	Percent Calorie From Fat in CHNS 89, 91, 93 From All Food Groups	137
2.8	Percent Calorie From Protein in CHNS 89, 91, 93 From All Food Groups	139
2.9	Percent Calorie From Carbohydrates in CHNS 89, 91, 93 From All Food Groups	141
2.10	Food Consumption in CHNS 89,91,93 with Community Characteristics: Overall	143
2.11	Food Consumption in CHNS 89,91,93 with Community Characteristics in Urban Areas	149
2.12	Food Consumption in CHNS 89,91,93 with Community Characteristics in Rural Areas	154
2.13	Changes in Daily Calorie Intakes	161
2.14	Changes in Daily Fat Intakes	163
2.15	Changes in Daily Protein Intakes	165
2.16	Changes in Percent Daily Calorie From Fat	167
2.17	Changes in Percent Daily Calorie From Protein	169
2.18	Changes in Percent Daily Calorie From Carbohydrates	171
2.19	Changes in Food Consumption in CHNS 89,91,93 with Community Characteristics	173

2.20 A Simple Production Function Analysis on Log(weight) in 1993 and Log(BMI) in 1993	177
3.1 Short Panel: Estimated restricted age, cohort and year effects for x .	190
3.2 Short Panel: Estimated restricted age, cohort and year effects for x1 .	191
3.3 Long Panel: Estimated restricted age, cohort and year effects for x . .	192
3.4 Long Panel: Estimated restricted age, cohort and year effects for x1 .	193
3.5 Short Panel: Estimated restricted age, cohort and year effects for y .	194
3.6 Short Panel: Estimated restricted age, cohort and year effects for y1 .	195
3.7 Long Panel: Estimated restricted age, cohort and year effects for y . .	196
3.8 Long Panel: Estimated restricted age, cohort and year effects for y1 .	197
3.9 Second Differences of Cohort Profiles	212
A.1 Survey instruments for health related topics	224
A.2 ADL type questions in CHNS 93-97	225
A.3 Adult BMI by Education Levels in Rural and Urban Areas	226
A.4 Community Characteristics in CHNS 1989-97	228
B.1 First stage regressions for Table 1.13	229
B.2 First stage regressions for Table 1.14	233
C.1 Probability Model for Overweight: Overall, Urban and Rural Areas .	237
C.2 Probability Model of Overweight: By Age Groups	239
C.3 Probability Model of Undernourishment: Overall, Urban and Rural Areas	241

C.4	Probability Model of Undernourishment: By Age Groups	243
D.1	AR1 Models for BMI	245
E.1	Patterns and Trends of Individual Daily Nutrient Intakes in CHNS 1989-1993	246
F.1	First Stage Regression for Health Production Function Analysis . . .	251

List of Figures

1.1	Male and Female Age Effects Identified through the specification with age dummies, five-year cohort dummies and year dummies.	40
2.1	FAO Food Balance Sheet Estimates of calories, protein and fat availabilities and percent of protein and fat from vegetable and animal products in 1961-2001.	97
3.1	Lowess Smoothed Male and Female BMI of Each Five- and Three-Year Cohort vs. Age	183
3.2	Estimated Age and Cohort Effects for Men and Women with Arbitrarily Imposed Constraints	184
3.3	Short Panel Estimates of Age, Cohort and Year Effects for x and x_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.	200
3.4	Long Panel Estimates of Age, Cohort and Year Effects for x and x_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.	203
3.5	Short Panel Estimates of Age, Cohort and Year Effects for y and y_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.	206
3.6	Long Panel Estimates of Age, Cohort and Year Effects for y and y_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.	209

3.7	The Second Derivatives of Age and Cohort Effects for Men and Women for BMI using two-year spaced CHNS 89, 91 and 93. * The approach for repeated cross-sections based on means. – The approach for genuine panel data based on individual level data.	215
3.8	The Second Derivatives of Age and Cohort Effects for Men and Women for BMI using four-year spaced CHNS 89, 93 and 97. * The approach for repeated cross-sections based on means. – The approach for genuine panel data based on individual level data.	217
3.9	Age Effects for Men and Women, identified through (a) exact age at interview (b) three-year cohorts, (c) five-year cohorts specifications, (d) three-year cohorts without breaking down at certain groups into single years, (e) five-year cohorts without breaking down at certain groups into single years.	218
3.10	Cohort Effects for Men and Women, identified through (a) exact age at interview (b) three-year cohorts, (c) five-year cohorts specifications, (d) three-year cohorts without breaking down at certain groups into single years, (e) five-year cohorts without breaking down at certain groups into single years.	219
3.11	Year Effects for Men and Women, identified through (a) exact age at interview (b) three-year cohorts, (c) five-year cohorts specifications, (d) three-year cohorts without breaking down at certain groups into single years, (e) five-year cohorts without breaking down at certain groups into single years.	220

Chapter 1: Socioeconomic Determinants of Body Mass Index of Adult Chinese in the 1990s

1.1 Introduction

China, like a number of other developing countries, is experiencing a wide range of transitions in her social, economic, and cultural structures. Since 1978 economic reforms opened the Chinese economy to the world and revitalized a stagnant economy. Productivity has risen and standard of living has increased substantially. Per capita GNP increased from 170 US dollars in 1980 to 668 in 1997 (both in 1995 dollars) and life expectancy increased from 68.3 years to 71.3 years for female and from 65.7 to 68.1 years for male during the same period (*World Development Indicators* 1999). These gains were most evident in the rural areas as a result of the implementation of the household responsibility system (Lin 1988) as well as other reform measures such as increases in the prices of agricultural products (Ke 1999). China's community-based health and sanitation programs extended services to almost all urban and rural residents. During the 80s and 90s China has also experienced a shift in the leading cause of death from acute infectious disease to chronic conditions, due partly to an aging population and partly to the successful reduction in infectious disease morbidity (Popkin et al. 1993, Zhai et al. 2002).

Like other countries in transition, such as Brazil, Russia, and some Asian countries, during this time, researchers found that in China shifts, desirable or undesirable, in diet, physical activity and overweight status are among the most rapid ever docu-

mented (Popkin and Doak 1998, Popkin 1999, Popkin 2002).¹ In China the dietary pattern is moving toward one in which the proportion of energy intake from fat increases each year. The intake of cereals decreased considerably during the eight-year period of 1989 to 1997 by 127 g per person per day; the intake of vegetables decreased by 32 g per person per day; and the intake of animal foods increased by 46.7 g and 36.8 g per person per day for urban and rural residents (Du et al. 2002). In a cohort of prime age men and women the prevalence of overweight² doubled for females (10.4 to 20.8%) and almost tripled in males (5.0 to 14.1%) in the same period (Bell et al. 2001). The increase in the level of overweight in adults is largely confined to the urban areas. It is projected that diet-related chronic diseases, such as obesity and coronary heart disease, will present a huge health care burden for China in the near future (Popkin, Paeratakul, Zhai and Ge 1995a). In urban China mortality attributable to cardiovascular diseases increased from 86.2 per 100,000 (12.1% of total deaths) in 1957 to 214.3 per 100,000 (35.8% of all deaths) in 1990. The overall prevalence of hypertension with threshold values of 140/90 mmHg, was 12.5% in adults aged 35–64 years in the 1990s (Reddy 2002). Increasing economic opulence brings new challenges to the health care system which is going through a transition itself (Gu 2001, Wu 1997). Preventive actions against the increase in obesity and diet-related non-communicable diseases are called for. Identifying the important socioeconomic factors related to obesity helps the designers of public health policies direct their efforts in the right direction.

The anthropometric measure body mass index (BMI), defined by weight (kg) divided by height squared (m^2), compared with the self-reported morbidity, is an objective measure of adult health status. To date studies on adult BMI (kg/m^2) in China have been primarily descriptive in nature. This paper employs a clear

¹The Bellagio Conference organized by the International Union of Nutritional Sciences Committee on the Nutrition Transition in 2001 published a series of papers on the nutrition transition in several developing countries. Source: <http://www.cpc.unc.edu/nutrition.transition/>.

²Overweight is defined by body mass index (BMI) greater than or equal to 25 kg/m^2 .

theoretical framework to estimate socioeconomic determinants of adult BMI using an eight-year panel data set, adding to a small number of studies that examine the reduced form BMI demand function for adult men and women in household surveys in developing countries. BMI demand functions provide useful information on how exogenous factors such as prices and the community infrastructure affect health. A dynamic or conditional BMI demand function is also estimated, which takes advantage of the longitudinal nature of the data and avoids certain specification difficulties in the pure reduced-form analysis (see section 1.2 for detail), although the estimates are only partial effects of the explanatory variables in this case. The key findings of this paper are for women the effect of education is very strong and inversely U-shaped. The education effect is stronger in rural areas than in urban areas and for younger generations than for older generations. Adult men in China had higher levels of education than women, however, their educational effects on BMI were not significant. Productive assets, prices, and environmental health conditions are all important determinants of adult BMIs in China and these factors affect men and women in different age groups and across regions differently. Controlling for lagged BMI the effects of education and household resources were no longer significant in both the OLS and 2SLS estimates.

The rest of the paper is organized as follows: the next section describes the theory and empirical strategy. Section three reviews studies on associations between adult BMI and socioeconomic status. Section four provides a summary of the data and descriptive evidence for the association. Results for the reduced-form analysis are in section five and for the conditional demand function analysis see the penultimate section. The last section concludes.

1.2 Theoretical Framework

Three types of measurement have been used as indicators of adult health in the literature, anthropometric measures, self-reported morbidity or general health conditions, and measures of nutrient intakes. They measure different aspects of a person's physical condition. For example, height is an indicator of childhood nutritional and health accumulation and BMI reflects both past and current nutritional and health status in fat content and active tissue mass (Willett 1998). Higher BMI has been used to diagnose obesity and lower BMI has been used as an indicator for chronic energy deficiency, as proposed by the International Dietary Energy Consultancy Group (Ge et al. 1994).

Self-reported indicators of health, such as being in excellent, good, fair and poor health, may reflect differences in perceptions of health among people in different educational or socioeconomic groups, rather than differences in health itself. Hence the self-assessment measures may generate biased results (Strauss and Thomas 1996). There is no known single index that measures a person's physical and psychological well-being at the same time. Anthropometric measures, particularly weight and height, are commonly accepted as measures of nutritional and health status by epidemiologists and nutritionists. They have also been used by economists to study long-term trends in nutrition and mortality (Fogel 1986, Fogel 1994, Costa and Steckel 1995) and to measure standard of living over time (Whitewell and Nicholas 2001). Whether they are good measures for health status in general is arguable. However, they are strong markers for mortality and morbidity from certain illnesses such as hypertension, diabetes, coronary heart disease and some cancers (NIH 1998). Based on the Waaler curve (iso-mortality curves on height and weight plain) constructed with data from Norwegian men aged 50-64, Fogel (1994) finds that factors associated with improvements in height and BMI predicted the decline in mortality between 1910 and 1980 among men 65 years and older. Extreme values of BMI are signifi-

cantly associated with an increasing risk of mortality. Hence this paper uses BMI as an indicator for adult health.

1.2.1 The Model

To understand health determination, it is necessary to specify the process by which health is produced.³ The economic model considers a rational individual with perfect information.⁴ Assume that the preferences of household members are intertemporally separable and that the household maximizes the present discounted value of a common utility over health of all family members, $h_t = (h_{1t}, \dots, h_{nt})$, consumption, $x_t = (x_{1t}, \dots, x_{nt})$, and leisure, $l_t = (l_{1t}, \dots, l_{nt})$ in time period t :

$$\begin{aligned} \max \quad & \sum_{t=0}^T \beta^t U(h_t, x_t, l_t; \theta, \xi_h) \\ \text{w.r.t } & \{x_{it}, l_{it}, m_{it}\}_{i=1 \dots n}^{t=0 \dots, T} \end{aligned} \quad (1.1)$$

where h_{it} is member i 's health status in period t , x_{it} are k dimensional vectors of food and non-food commodities consumed by member $i = 1, \dots, n$, $\theta = (\theta_1, \dots, \theta_n)$ includes a vector of exogenous individual characteristics known to family members but not controlled by them, such as genetic traits, and ξ_h indicates unobserved household heterogeneity in preferences or environmental factors. Single period utility is assumed to be increasing and concave in h_{it} ($U'_h > 0, U''_h < 0$). h_{i0}, x_{i0}, l_{i0} and m_{i0} are given and positive at t_0 when a person enters adulthood. At $T + 1$, the person dies and $h_{iT+1}, x_{iT+1}, l_{iT+1}$ and m_{iT+1} are zero.

Adequate to the application of BMI, the health status at the beginning of period $t + 1$ is assumed to be influenced by health status at t , health inputs in period t ,

³The set up in this model is similar to that in Foster (1995).

⁴Uncertainty is ignored in this model. There is evidence that in developing countries, for idiosyncratic risks households are natural units of risk-pooling and consumption smoothing through the accumulation and depletion of assets over time is not uncommon. Uncertainty could be important if overweight increases the risk of mortality and the person is aware of it.

m_{it} , such as the balance between energy expenditure (basic metabolism, working or exercising) and intakes, medical care, or illness spell, which does not bring utility directly. More leisure time may bring feelings of well-being and sound state of mind, which can have a positive influence of one's physique. Individual characteristics, θ_{it} , such as age,⁵ gender and community or environmental characteristics, z_{ct} , may also have a direct impact on health outcomes:

$$h_{it+1} = f_i(h_{it}, m_{it}, l_{it}, \theta_{it}, z_{ct}; v_{it}) \quad (1.2)$$

where v_{it} are unobservable individual, household, and community factors that affect member i 's health. It is assumed that $\partial f_i / \partial h_{it} > 0$, and $\partial f_i / \partial m_{it} > 0$. Equation (1.2) is not a general form of health production function in that it assumes that h_{it} is a sufficient statistic that summarizes the effect of all past inputs and choices in periods $1, \dots, t-1$ and there is no direct lagged effects from them on h_{it+1} . A more general health production function would allow all past inputs to be included.

As in Grossman's (1972) original work on health production, time, as in the number of healthy days available for leisure l_{it} and work l_{it}^w , may also be affected by health:⁶

$$T_{it} = l_{it} + l_{it}^w = \Omega_i(h_{it}) \quad (1.3)$$

with $\Omega'_i > 0$.

The household maximizes utility under the standard intertemporal budget con-

⁵The solution to the maximization problem is a time-invariant policy function that determines the control variable (x_{it}, l_{it}, m_{it}) and the state variable h_{it} . It is conceivable as a person grows old the effect of inputs on health (body size) varies. To incorporate such effect we can assume a health production function that varies with age at time t , a_{it} :

$$h_{it} = f_i(h_{it-1}, m_{it}, l_{it}, \theta_{it}, z_{ct}; v_{it}) + g_{it}(a_{it})m_{it}$$

⁶For the case where total time available for leisure and work is fixed, or not affected by health status, labor is the same as the effective labor in Grossman's sense. Then the equilibrium condition in (1.5) will become $\frac{U_x^i}{U_z^i} = \frac{(1+r)p_{t-1}^m}{f_m^{i-1}p_t^i} - \frac{f_h^i p_t^m}{f_m^i p_t^i}$

straints:

$$\begin{aligned}
A_{t+1} &= (1 + r_{t+1}) \left(A_t - p_t^x \sum_{i=1}^n x_{it} - p_t^m \sum_{i=1}^n m_{it} \right) + \sum_{i=1}^n w_{it} l_{it}^w + y_t \\
A_{T+1} &= 0 \text{ and given } A_0 > 0
\end{aligned} \tag{1.4}$$

where A_t is assets, p_t^x , p_t^m are vectors of prices, w_{it} is wage for individual i , and y_t is household non-wage income between t and $t + 1$ received at the end of period t . There is no bequest motive and no debts in the final period T , or $A_{T+1} = 0$. The lifetime of the household is assumed to be exogenously fixed.⁷ The time and budget constraint can be collapsed to derive the full-income constraint. From the first order conditions of maximizing (1.1) subject to (1.2) to (1.4) and $r_t = r_{t+1} = r$ we can derive the following:

$$\frac{U_h^t}{U_x^t} = \frac{\pi_{it}}{p_t^x} \equiv \frac{1}{p_t^x} \left(\frac{(1+r)p_{t-1}^m}{f_m^{t-1}} - \frac{f_h^t p_t^m}{f_m^t} - \frac{\Omega'_{it} w_{it}}{(1+r)} \right) \tag{1.5}$$

and

$$\frac{\beta(1+r)U_h^{t+1}}{U_h^t} = \frac{\pi_{it+1}}{\pi_{it}} \tag{1.6}$$

In equilibrium in each period the household will equate marginal rate of substitution between health and consumption with the shadow price of health (1.5), which is lower than the costs for medical care because bettering current health decreases the future medical costs and improves the healthy days or effective labor. Equation (1.6) states that over time the marginal rate of substitution of health today and tomorrow equates the ratio of the shadow prices.

The dynamic problem can be solved recursively and the reduced-form demand

⁷For a discussion of the length of individual lifetime as endogenous in the model of own-health production see Grossman (2000). For simplicity in the collective household model the household's utility is assumed to be the utility of a dictator in the household and the lifetime of the household becomes the lifetime of this representative and is assumed to be fixed in this model.

functions for consumption goods, health, leisure and health inputs (x, h, l, m) are:

$$(x, h, l, m)_i^t = g_{x,h,l}^t(p_x, p_m, w_i, y_i, \theta_i, \theta^h, z_c; v_i, \xi_h) \quad (1.7)$$

where $p_x = (p_T^x, p_{T-1}^x, \dots, p_0^x)$, $p_m = (p_T^m, p_{T-1}^m, \dots, p_0^m)$, $w_i = (w_{iT}, w_{iT-1}, \dots, w_{i0})$, $y = (y_{iT}, y_{iT-1}, \dots, y_{i0})$ and v_i, ξ_h are unobserved individual and household heterogeneity over the life time. As noted in Rosenzweig and Schultz (1983) without specifying the exact form of the utility function in (1.1) a closed-form solution for the demand equation in (1.7) cannot be obtained and the properties of these reduced-form demand functions are “identical to those from models positing no household production of health”.

From this clear theoretical framework we can see that estimating (1.2) by OLS is biased – each input component is determined by and correlated to the unobserved individual, household and community factors as shown in equation (1.7). It would be of interest if we can identify parameters of the health production function. It is extremely difficult to find proper instruments for all health inputs. However, in (1.7) all the determinants, including individual, household and community characteristics and prices, except for wage, are assumed to be orthogonal to v_{it}, ξ_h . Using productive assets as a proxy for income we estimate the relationship (1.7) using OLS, which is consistent under standard regulation conditions, such as no measurement errors and no omitted variables.

Little is known about the effects of underlying socioeconomic factors on adult health in China, especially during time of transition. What factors have driven or helped contain the dramatic increases in overweight and obesity? What is the effect of schooling on adult health? Are there gender or regional differentials in health? An unconditional reduced-form demand function for health provides a helpful first step in finding answers to such important questions.

Clearly, in order to estimate the reduced-form demand function we need the entire paths of prices and wage rates for each individual, which is not available in any household survey and may create a saturated model problem. If we pool all the waves together and use a vector of current prices and a dummy variable for each community assuming all past and future prices and other unobservables within a community follow a unique path captured by this dummy variable, then we can avoid the omitted variable bias. Obviously one dummy variable can not stand in for all past and future prices. We use community and year interaction dummies to capture some variations for each community in time.

As is specified in the health production function, assuming the lagged health is a sufficient statistic for all past information, we can also get around the need of the entire history of prices by estimating a conditional dynamic health demand function. Assuming the utility function is not linear in h_{it} in (1.6) using the first order expansion of the marginal utility of health between today and tomorrow and substituting other control variables in shadow prices from (1.7) we can write the conditional health demand function as:

$$h_{it+1} = c(h_{it}, p_{t+}^x, p_{t+}^m, w_{it+}, y_{t+}, \theta_{it+}, \theta_{t+}^h, z_{ct+}; v_{it+}, \xi_{ht+}) \quad (1.8)$$

where $p_{t+}^x = (p_T^x, p_{T-1}^x, \dots, p_{t+1}^x)$, $p_{t+}^m = (p_T^m, p_{T-1}^m, \dots, p_{t+1}^m)$, $w_{it+} = (w_{iT}, w_{iT-1}, \dots, w_{it+1})$, $y_{t+} = (y_{iT}, y_{iT-1}, \dots, y_{it+1})$ and v_{it+}, ξ_{ht+} are unobserved individual and household heterogeneity from time t forward. As we can see the conditional health demand function does not depend on all lagged prices and wage rates any more. By assuming all future prices within each community could be summarized by a community dummy variable we can estimate the following relation

$$h_{it+1} = c(h_{it}, p_{t+1}^x, p_{t+1}^m, w_{it+1}, y_{t+1}, \theta_{it+1}, \theta_{t+1}^h, z_{ct+1}; \nu_i) \quad (1.9)$$

with the composite error term ν_i that includes all unobserved individual and household heterogeneity and future wage and nonwage income that is not observable to researchers. The dynamic demand model estimates the partial effects of individual, household and community variables conditional on past BMI.

In this paper we focus on the reduced-form and conditional demand for BMI of adult household members. By construction, all covariates in (1.7) are exogenous, assuming there is no selective migration to a better health environment and no program placement in response to poor health. The first assumption is not unreasonable in the early 1990s in China. Though increasingly the restriction of household registry has been relaxed, the migration from rural to urban area in China is still largely temporary. The second assumption may be problematic in rural areas if the government did intervene in responding to local need by investing more in health facilities or increasing the number or the training of health professionals.

1.2.2 Estimation Strategies

The basic estimating equation takes the form $h_{it} = X_{it}\beta + v_i + \epsilon_t$, where h_{it} is person i 's BMI in year t , X_{it} is a vector of exogenous variables including the individual's age, cohort, education level and household resources, and v_i is the individual random effect. Although we have longitudinal data we still don't observe all past and future prices and other households and community characteristics. When prices and community characteristics are not explicitly controlled for, using community dummy variables we effectively assume all past and future prices and other unobservables within a community follow a unique path captured by this dummy variable. One dummy variable may not stand in for all past and future prices. Thus we use community and year interaction dummies to capture some variations in each community over time. When prices and community characteristics are explicitly controlled for, we also include the community dummy variables for unobserved heterogeneity and

allow the interaction of community and year to be captured by variations in prices and other community characteristics in different years.

Since it is established that both the upper and lower tails of the BMI distribution are related to worse health, a natural question is why not estimating a dichotomized outcome of being overweight or undernourished. There are two reasons. First and most importantly, if the goal is to estimate the differential impacts of X (including a constant) at different points or ranges of the BMI distribution, then a linear probability model (LPM) or a probit model for a dichotomized variable does not achieve this goal. Think of a latent variable framework. Let $y^* = Xb + e$ and $y = 1$ if $y^* > 0$. A linear probability model estimates $\hat{P}(y = 1 | X) = X\hat{b}_{LPM}$ and a probit model estimates $\hat{P}(y = 1 | X) = \Phi(X\hat{b}_{probit})$. Since $P(y = 1) = P(y^* > 0) = P(Xb + e > 0) = P(e < Xb)$, if $e \sim \text{uniform}(0, 1)$ then $P(y = 1) = Xb$. Similarly, if $e \sim \text{normal}(0, 1)$ then $P(y = 1) = \Phi(Xb)$. Hence both the linear probability model and the probit model are estimating the same latent relationship and the direction in the estimates of BMI greater than certain cutoff point should be the same as in the OLS estimates of BMI itself. In appendix C we provide some results from a linear probability estimation of undernourishment and overweight.

Secondly, dichotomization loses information and results in inefficiency. The non-linear effects of any X variables on BMI at different points or ranges of the BMI distribution cannot be efficiently estimated. In order to identify such nonlinearity one solution might be gleaned from the dynamic demand function with interactions of the lagged y and the main variable of interest, such as $y_t^* = f(y_{t-1}^*)g(Xb) + e$.

So far we have not assumed any distribution of ϵ_t in $h_{it} = X_{it}\beta + v_i + \epsilon_t$. In estimating the basic model, we do allow arbitrary correlations between ϵ_t and ϵ_{t-1} at the individual level and use cluster robust standard errors in all regressions for inferences. We could also assume ϵ_t follows an AR(1) process and estimate a linear model with such disturbance $\epsilon_t = \rho\epsilon_{t-1} + u_t$, where $|\rho| < 1$ and u_t is independent and identically

distributed with mean 0 and variance σ_u^2 . We have unbalanced panels whose observations are unequally spaced over time, i.e., not all individuals were surveyed in 1989, 1991, 1993 and 1997. We use the Baltagi and Wu (1999) approach to estimate the model with AR(1) disturbance in which ρ represents autocorrelation between every two-yearly spaced data. The results are included in Appendix D for comparisons.

1.3 Literature Review

Studies on association between adult health and socioeconomic status in the U.S. concentrate on mortality or self-reported morbidity, functional limitation or general health (Attanasio and Hoynes 2000, Smith 1998, Smith and Kington 1997). In developing countries, studies on adult health determinants are scant and sometimes difficult to interpret due to lack of theoretical or modelling clarity (for a comprehensive review see Behrman and Deolalikar 1988 and Strauss and Thomas 1998). Studies on adult BMI, according to their nature – the questions asked and the methodologies used to achieve those goals, can be classified into three categories: (i) descriptive and monitoring studies; (ii) determinants of health in reduced form demand function studies; and (iii) health production function studies. In this paper we focus on the first two types of studies.

Obesity (defined as BMI>30) is increasingly a global health issue. According to WHO, the prevalence of obesity has increased about 15-20 percentage points in Western and Southern Europe and is even higher in Eastern Europe (Sundquist and Johansson 1998). The proportion of overweight US adults (BMI>25) increased from 25.4% to 33.3% between 1976-80 and 1988-91 periods (Kuczmarski et al. 1994).

Using a longitudinal Swedish Annual Level-of-Living Survey of men and women aged 25-74 between 1980/81 and 1988/89, Sundquist and Johansson (1998) found that there was a clear gradient for the attained level of education and BMI, with

the highest BMI for people with a low level of education. All educational groups increased their BMI over 8 years except for men with a low education level, however, only poorly educated females had a higher BMI adjusting for age, smoking status, exercise level, marital status, general health status, ethnicity and time. Since most of these variables in the multivariable analysis are endogenous, this may bias the results. The authors thus look at the change in BMI between two surveys. They also found that exercise habits were more changeable than smoking habits and men and women who stopped smoking had a larger increase in BMI than never smokers. However, for the change of BMIs they only performed univariate regressions.

Using individual level data in 1987 and 1995 Ruhm (2000) examined the variation of risky behaviors, smoking, and time-intensive health investments in physical activity, diet and preventive medical care with the status of the economy controlling for fixed-effects at state level. He found higher jobless rate is associated with reduced smoking and obesity and with increased physical activity and improved diet. When the economy starts to grow the opportunity cost of time will increase and some risky activities may become normal goods. He found strong evidence that “health improves when the economy temporarily deteriorates”... and some evidence that “the unfavorable health effects of temporary upturns are partially or fully offset if the economic growth is long-lasting.”

Realizing the long-run growth in body weight has been accompanied by almost stable total calorie consumption and significant decreases in the relative price of food, Lakdawalla and Philipson (2002) outlined a theory of weight management that will predict such combined results through technological changes on both the supply (agricultural innovation) and the demand (more sedentary production) side. Based on their hypothesis the partial effect of the price of food on weight is negative, which was interpreted as the supply side effect of technological change. The effect of the strenuousness of production (S) on weight is also negative, which indicates that the

reduction in S due to sedentary technological change results in the increase in weight. The relationship between income and weight can be positive, negative or inverted U-shaped. A negative pure income effect combined with a positive indirect income effect through changing the strenuousness of work can still produce increased weight over time. Using the National Longitudinal Survey of Youth (NLSY) the authors showed evidence that the partial effect of exercises is concave and occupation is exogenous with respect to weight. Individual fixed effect models of BMI on job-related strenuousness and a one-year BMI on average level of strenuousness and strength across all years reveal that a one-year-one-unit increase in average strenuousness lowers women's BMI by 0.19 unit and the 14-year effect is about six times as large as the one-year effect. Switching into less strenuous jobs are not preceded by increases in BMI. People who switch into less strenuous occupations do not already weigh more than others. Using the National Health Interview Survey 1976-94 the authors estimated individual's BMI as a function of year trend, the strength requirement of a job (>0), job strenuousness (<0), income quartiles, education, age and age squared, race and marital status. The main findings were: the effect of job strenuousness is larger than other economic factors stressed as key determinants such as income and education; income effect for men is inverse U-shaped and for women is negative.

Is obesity an inevitable price to pay for economic growth and technological innovation in developing countries as well or is there a unique feature that distinguishes them from the developed countries at a similar stage of economic development (Popkin 2002)? Scholars in the U.S. and China joined their effort in finding some answers for such questions. They describe the changing dietary and body weight patterns among the Chinese and document the association between dietary, environmental factors and obesity (Popkin et al. 1993, Ge et al. 1994, Popkin, Paeratakul, Zhai and Ge 1995a, Popkin, Paeratakul, Zhai and Ge 1995b, Popkin, Paeratakul and Zhai 1995, Paeratakul et al. 1998, Stookey et al. 2000). In a cohort of prime age

men and women the prevalence of overweight doubled for females (10.4 to 20.8%) and almost tripled in males (5.0 to 14.1%) in the same period (Bell et al. 2001). The intake of cereals and vegetables decreased considerably and the intake of animal foods increased. These studies provide various indicators on whether the health status of the target population is improving or deteriorating over time. There is evidence that dietary energy and fat intakes were positively and significantly correlated with the BMI; urban residence and higher income were correlated with lower energy intake, higher fat intake and lower physical activity level compared to rural residence and other income categories. These studies have a common problem which makes the interpretation of their results difficult or misleading. They are estimates of correlations between the inputs and BMI and should not be interpreted as causal effect from a health production function. The inclusion of endogenous factors as exogenous explanatory variables tends to bias the parameters in a production function estimation.

In Ge et al. (1994), using the Chinese Health and Nutrition Survey (CHNS) in 1989 (adults aged 20 to 45) and a large Chinese Nutrition Survey in 1982 (adults of all ages), the authors found that increased income was significantly associated with reduced BMI in the urban sample, while for the rural and overall samples the opposite was true. Subpopulations consuming greater proportions of energy from animal sources were more likely to be overweight. The percentage of underweight (BMI<18.5) was low (7 to 13%). The results were based on simple analysis of variance, which does not establish causal relationships nor control for other factors.

In Popkin et al. (1993), again using the CHNS 89 and statistics from the Chinese bureau, the authors found that per capita cereal and vegetable consumptions increased in the early 80s and levelled off later. Per capita meat, edible oils, sugar, eggs and fish increased each year, resulting a marked shift in the structure of the diet to one with high proportion of energy from fat. Cross tabulations of income and major food groups suggested that greater consumption of higher fat food was associated

with higher income levels.

For the elderly in China, the CHNS 91 and 93 were used (Stookey et al. 2000) to estimate the determinants of nutrition intakes and BMI. The dependent variables included energy, fat, protein intakes, consumption of rice, high-fat red meat, eggs, plant oils, and BMI. The explanatory variables were age groups, sex, income tertiles, and rural-urban residence. The prevalence of low BMI (<18.5 or <22.0) exceeded 15% and the prevalence of overweight (BMI > 25.0 or > 27.0) ranged between 4% and 24%. Being overweight was found to be significantly and positively correlated with urban residence in 1991 and 1993 and higher income in 1991. The authors attempted to correct sample selection bias due to missing values in dietary or anthropometric measures in 1993 using such variables as household size, province dummies, and types of roads. Their correction may not be very convincing because if they are estimating reduced form demand functions there is no reason that their selection correction variables should not be included in the main regression. The authors found that the group with missing BMI or dietary information in 1991 was significantly older, from larger families, and were less likely to live in Jiangsu province in 1991. The group with missing values in 1993 was significantly older, from larger families, in rural areas, and in the top income tertile. Bearing the possibility that their correction is poorly identified we note their results below. Male and younger age groups have significantly higher energy intakes while urban residents have significantly lower energy intakes but higher proportion of energy from fat and protein. Increasing income was significantly associated with greater energy, fat and protein intakes. Rice consumption declined with older age and consumption of plant oils, high-fat red meat and eggs increased by income tertiles. Income was positively associated with the probability of having very low physical activity level and inversely associated with very high physical activity level. Undernutrition was concluded to be a more pressing problem for the elderly as BMI declined significantly with age. The coexistence of overweight and underweight

in the elderly suggests more future demand of health care provision.

Simple analysis of covariance was employed to find the role of diet and socioeconomic factors as the determinants of BMI in Popkin, Paeratakul and Zhai (1995), using the CHNS 1989 and 1991 for adults 20 to 45 years of age. For BMI in 1991 the covariates were total energy intakes, fat intakes of 1989, energy from sources other than fat, household income, age, place of residence, physical activity level, and smoking habits. This is an example where all covariates in a health production function were treated as exogenous. The parameters may be under or overestimated. BMI increased with energy and fat intakes and was negatively correlated with smoking and physical activity levels.

In a subsequent study (Paeratakul et al. 1998), the authors employed the fixed-effect method at individual level to estimate the effect of changes of the above variables on the changes in BMI. They found only the change in fat intakes was positively associated with changes in BMI in men and changes in physical activity level in women were negatively associated with changes in BMI in women and all other factors were no longer significant. Their corresponding cross-section analysis using OLS indicated that both fat and energy intakes were positively associated with BMI in men and only energy intakes were positively correlated with BMI in women. Age, urban residence and household assets (possession of certain major durable goods) were positively related to BMI. Education level was negatively related to BMI in women and marital status positively related to BMI in men. The authors acknowledged the inconsistency of the results from the OLS and the fixed-effect model yet did not give satisfactory explanations. One source of the discrepancy may be due to the potential endogeneity in their OLS estimation. Unobserved shocks to BMI, for example income shocks, will affect both intakes and physical activity level resulting in endogeneity bias in all estimates.

Among the few reduced form adult health determinants studies using anthropo-

metric measures, Thomas et al. (1996) estimates a BMI demand function for adults in Côte d'Ivoire. The authors control for the joint determination of health and consumption by instrumenting per capita expenditures with value of land, livestock and assets of the household under the assumption of weak separability between leisure and consumption. The results of this study indicate that higher food prices are associated with lower BMI, especially among rural residents of Côte d'Ivoire. Doubling the price of beef and fish in rural area will reduce BMI by 10% to 20% among rural dwellers. Increasing the price of plantains, eggs and manioc would have an even larger impact on BMI. The authors also find a significant positive impact of per capita expenditure on BMI and find evidence to indicate that the health of men would be less affected than that of women by declines in household resources. Education only had some impact on BMI for urban females controlling household resources.

In this paper, taking advantage of the detailed community level price, health facility and environment data, employing a clear theoretical model, we estimate the reduced form and conditional BMI demand function and find potential socioeconomic determinants. This paper fills in the gap of previous research due to lack of distinction between endogenous and exogenous variables in the model and becomes the first paper on Chinese adult BMI estimations with emphasis on individual as well as community characteristics.

1.4 Data and Descriptive Statistics

1.4.1 The Data

The Chinese Health and Nutrition Survey (CHNS), conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill (CPC-UNC), the Institute of Nutrition and Food Hygiene (INFH) and the Chinese Academy of Preventive Medicine (CAPM), is designed “to examine the effects of health, nutri-

tion, and family planning policies and programs implemented by national and local governments and to see how the social and economic transformation of Chinese society is affecting the health and nutritional status of its population.”⁸ It is important for the government to understand the impact of the economic transition on people’s health and be prepared for increasing demand for certain health care services. The wide range of survey instruments of the CHNS provides researchers an useful tool to better describe and understand the associations between socioeconomic and health status and to find out potential determinants for health and nutrition status so that the government may adjust its program emphases and provide better care for the people.

The survey covered 8 provinces (Guangxi, Guizhou, Henan, Hubei, Huan, Jiangsu, Liaoning, and Shandong) that vary substantially in geography, economic development, public resources, and health indicators in 1989, 1991, 1993 and 1997. A multistage, random cluster process was used to draw the sample. Counties in the 8 provinces were stratified by income (low, middle, and high) and a weighted sampling scheme was used to randomly select 4 counties (one in low, two in middle and one in high income counties) from each province. In addition, the provincial capital and a lower income city were selected. Villages and townships within the counties and urban and suburban neighborhoods within the cities were selected randomly. There are about 190 primary sampling units and 3,800 some households covering approximately 16,000 individuals. For the survey design and sample structure see related articles (Popkin et al. 1993, Stookey et al. 2000). The same household was contacted in each survey. In 1993 an effort was made to locate the individuals who started a new household or change to a different household in the same region. In 1997 Liaoning province was replaced by Heilongjiang province, which is similar to Liaoning in many characteristics. According to the Chinese census definition, small towns and city neighborhoods

⁸<http://www.cpc.unc.edu/projects/china/>

are defined as urban areas, and villages and suburbs as rural areas.

This paper focuses on adults 20 years and older. In the reduced-form BMI demand analysis all four waves of the data are pooled together. In 1989, health and nutrition data were only collected from preschoolers and adults age 20 to 45. The first wave in 1989 thus includes only adults age 20 to 45 with consistent sex and age measures from the basic module and the physical exam module.⁹ The number of adult Chinese aged 20 to 45 in CHNS-89 is 6,578 and the number of individuals with BMI measures is 5,058.¹⁰ In the second wave, the number of all adults aged 20 and above with consistent age information from the roster and from the physical exam section¹¹ in CHNS-91 is 10,461, among whom 8,260 have BMI measures. The number of adults aged 20 and above with nonmissing gender¹² in CHNS-93 is 9,839 and a subsample of 7,706 people have BMI measures. Finally in CHNS-97 there are 11,338 observations for adults age 20 and above with nonmissing gender information and 8,356 adults with BMI measures. Each year after 1991 missing values in BMIs are more likely to be men and people who are considerably younger and more educated. It may be because at time of interview these people are more likely to be out or off for work. For the final sample size in our analysis after accounting for all missing values see Table 1.1.

The household survey section includes basic demographic characteristics of all

⁹In the web-based 1989 adult physical examination data set there are 5,183 observations for adults age 20 to 49, among whom 56 people have different gender information from the physical exam section than from the roster and 95 individuals have more than 2 years of age difference (age calculation based on lunar calendar and western calendar may result in up to 2 years of difference) between age in the physical exam section and the roster. Cleaning and correction are based on date of birth and relationship to the household head.

¹⁰Pregnant women are excluded in all years.

¹¹There are 330 individuals with the age difference more than 2 years. They are excluded for data quality consideration. Because age in 1991 and forward is calculated based on date of birth it can not be used for data cleaning again. 137 adult members in the 1989 survey died by the time of the second round of CHNS.

¹²Sequentially excluded from the original data set are those who are less than 20 years of age, who are dead, who have missing gender information. Missing gender cannot be identified based on relationship to the household head because in 1993 and 97 son and daughter to the household head were given the same code in the variable.

household members, detailed income from different sources, time allocation at home and economic activities, ownership of consumer durable goods and living environments.

In most of the previously reviewed papers, the focus was on the correlation between BMI and income controlling for intakes and/or physical activities. They are hybrid models that differ from a pure production function estimation and a pure reduced-form demand function estimation. Incomes in developing countries have been found to be particularly error-ridden. Improved health status may enhance productivity and income in reverse causality. Attenuation bias and endogeneity bias of incomes are not easily corrected. As a result this paper uses productive assets in the main analysis.¹³ The nominal value of productive assets are self-reported ownership and value of tricycles, motorcycles, tractors or walking tractors, irrigation equipment, power threshers and water pumps. Items not available in all years are not included.¹⁴ Productive assets are discounted to the 1988 value. All price indices are based on the China Statistical Yearbooks (SSB 1988-1998) for urban and rural areas at the provincial level. The CHNS research team at CPC compiled their own price index based on an urban consumer basket, the SSB urban price data, and the CHNS urban rural price ratio for each year in each province. The amount of land (mu) cultivated in the previous year is used as an independent variable.

One important factor in health production is education. In CHNS both years of formal education and highest level of attainment were asked. For about less than one percent of the people years of education were top-coded at 18 years. I chose and categorized the highest level of education by none, some primary schooling, primary degree, middle school degree, high school diploma, and technical, vocational and

¹³Results for comparison purpose on some limited analysis of the effect of income on BMI treating income as endogenous and instrumented by productive assets are available upon request

¹⁴Productive assets in 1989 are calculated differently from other years. All productive assets owned by 1989 are evaluated at last year's purchasing prices. In the other years the total value was self-reported current worth.

college or higher degree.

The health services section contains health insurance coverage of surveyed individuals, medical providers, and health facilities that the household might use. Information on illness and uses of the health system during the previous month is asked for some household members. Physical examinations that include weight, height and blood pressures were given to all adults and children starting from 1991 and to adults aged 20 to 45 in 1989. Physical functioning data were collected since 1993 for the elderly. Three parts of the CHNS concern an interviewee's health (Table A.1 in Appendix A summarizes what is available year-wise in each part of the survey instruments). First, in the household health and medical services module, there are questions about a person's self-evaluation of her health status, recent illness and injury, medical insurance and expenditure (treatment and transportation costs), diagnoses and facilities. Secondly, in the anthropometric measurement and physical examination module, weight, height, upper arm circumference, waist and hip circumferences (available in 1993 and 1997), and blood pressure are measured for adults (in 1989 only for adults aged 20 to 45 years old). For different age and gender groups, illness or symptoms pertaining to them were asked. Past and current smoking and alcohol behaviors, injury history are recorded. For the elderly (age 50 and older in 1993, and over 55 in 1997) activities of daily living (ADL)-type questions were asked (Table A.2 in Appendix A lists the 15 ADL-type questions). Thirdly, in the ever-married women module, marriage history and birth history are surveyed in detail starting from 1993, so is infant food record.

Internationally accepted adult BMI cutoffs are used to define underweight as a BMI < 18.5 kg/m², overweight as a BMI ≥ 25 kg/m² and obese as a BMI ≥ 30 kg/m² for adults 20 years and older (the WHO standards).¹⁵ The National Institute

¹⁵For the elderly, Stookey et al. (2000) have used a different set of cutoffs and suggested them to be more pertinent, which defines overweight and obesity for 60 years and older people as a BMI ≥ 22 kg/m² and BMI ≥ 27 kg/m². Recently the China Center for Disease Control and Prevention (CCDC) proposed a cutoff for Chinese population to classify overweight as a BMI ≥ 24 kg/m² and

of Heart, Lung and Blood clinical guidelines (NIH 1998) for identifying and treating obesity suggests if a person's waist circumference is greater than 88 cm (for women), or 102 cm (for men),¹⁶ the risk of having certain illnesses are greatly increased when they are overweight. The guideline points out that waist circumferences greater than the cutoffs for overweight adults indicate excess fat in the abdomen out of proportion to total body fat, which is an independent predictor of morbidity for men and women with a BMI of 25 to 34.9 kg/m². These cutoff points lose their incremental predictive power in persons with a BMI ≥ 35 kg/m².

The community survey section includes information on infrastructure, services, populations in the village or neighborhood, percentage of land with poor quality, daily wage for unskilled farmers and construction workers, percent of work force engaged in agriculture or working out of town for more than one month, and hospital and clinic infrastructures and personnel. All questions were answered by a knowledgeable respondent. In each community, state ration coupon, retail and free market prices for most commonly consumed rice, wheat, egg, pork, beef and fish were collected in stores. Since free market prices reflect the value of each produce better they were in the reduced-form analysis with community characteristics. Every health service and family planning provider or facility were identified and information about personnel, prices and availability of services was collected. The percentages of households within each community with certain types and sources of water and toilet facility are aggregated from household level data.

obesity as a BMI ≥ 28 kg/m². Since this decision is not finalized and for purpose of comparison with other studies we still use the WHO cutoffs in this paper.

¹⁶Similarly, CCDC proposed a set of different cutoff points for women at 80 cm and for men at 85 cm.

1.4.2 Summary Statistics

BMI, Overweight and Undernutrition

Trends for BMIs and prevalence of being undernourished, overweight, obese and overweight with high abdominal fat for men and women in different age, residential and education groups are in Table 1.1.¹⁷ Over the eight-year period, 1989 to 1997, average BMI for men and women increased by about 1 unit. To put things into perspective, that is about 7.4 pounds of increase in weight for a men 6 foot tall. However the overall prevalence of overweight increased from 6% to 17% in men and from 11% to 21% in women. The percent of men and women who are obese also increases dramatically. Using the waist circumference cutoffs in the previous session the percent of overweight women with increasing risks is about 9% in 1997 and the percent for men is lower, probably because the cut-off point is too high for Chinese men. There is an improvement or decrease in undernutrition for all men (by one percentage point) and women (by about 1.5 percentage point) over time.

Breaking the sample into different age groups we can see men and women in prime age (40 to 59) have the highest BMIs. The prevalence of being overweight in the younger cohort (age 20 to 39) increased from 6 to 14 percent for men and 9 to 15 percent for women in this eight-year period. Nine to 19 percent of prime age men are overweight and the percent for women in the same age is much higher (20 percent in 1989 to 25 percent in 1997). Elderly women (aged 60 and above) during 1991 to 1997 have the same risk of being overweight as prime age women and 15 to 23 percent of the elderly men are overweight. In all age groups the prevalence of obesity is higher for women than for men, so is the prevalence of being overweight with high abdominal fat. The percent of men and women being undernourished decreased the

¹⁷Pregnant women were excluded. Having high abdominal fat is defined as having waist circumference > 102 cm for male or > 88 cm for female. The sex-specific cutoffs can be used to identify increased relative risk for the development of obesity-associated risk factors in most adults with a BMI of 25 to 34.9 kg/m² (NIH 1998). In 1989 data on BMI were collected for those 20 to 45 years of age. In 1989 and 1991 data on waist circumference were not collected.

most for prime aged men and the elderly. The incidence of underweight is much higher among the elderly than the other age groups due partly to shrinking. This pattern of overweight is very similar to that found in Indonesian Living Standard Survey (Strauss et al. 2004). Although the prevalence of overweight and obesity in men and women is much lower than that of the United State, the increasing trend among all age groups is certainly of concern.

Table 1.1: Adult BMI in CHNS 89-97

	Male				Female			
	1989	1991	1993	1997	1989	1991	1993	1997
All Sample								
Number of observations	2196	3537	3329	3698	2354	3819	3597	3882
Median BMI	21.00	21.13	21.34	21.78	21.48	21.50	21.64	22.05
Mean BMI	21.23	21.60	21.88	22.34	21.74	22.10	22.14	22.66
	(2.27)	(3.06)	(3.59)	(3.55)	(2.65)	(4.36)	(3.71)	(4.10)
% Undernourished (<18.5)	8.15	9.81	8.53	7.25	9.22	10.81	9.87	7.73
	(0.27)	(0.30)	(0.28)	(0.26)	(0.29)	(0.31)	(0.30)	(0.27)
% Overweight (≥25)	6.24	10.88	12.11	17.41	11.00	16.05	16.35	20.97
	(0.24)	(0.31)	(0.33)	(0.38)	(0.31)	(0.37)	(0.37)	(0.41)
% Obese (≥30)	0.27	1.02	1.59	2.16	0.76	2.33	1.95	2.99
	(0.05)	(0.10)	(0.13)	(0.15)	(0.09)	(0.15)	(0.14)	(0.17)
% Overweight risky ^a			0.66	1.24			6.39	8.73
			(0.08)	(0.11)			(0.24)	(0.28)
Age 20-39 years								
Number of observations	1791	1771	1519	1597	1939	1965	1671	1602
Median BMI	20.96	20.97	21.16	21.57	21.36	21.23	21.29	21.51
Mean BMI	21.16	21.42	21.73	22.06	21.57	21.74	21.62	22.23
	(2.19)	(2.83)	(3.50)	(3.34)	(2.51)	(4.61)	(3.18)	(4.09)
% Undernourished (<18.5)	7.43	7.57	6.65	7.01	9.59	9.36	10.23	6.62
	(0.26)	(0.26)	(0.25)	(0.26)	(0.29)	(0.29)	(0.30)	(0.25)
% Overweight (≥25)	5.58	7.68	9.28	13.78	9.08	11.04	10.29	14.86
	(0.23)	(0.27)	(0.29)	(0.34)	(0.29)	(0.31)	(0.30)	(0.36)
% Obese (≥30)	0.17	0.68	1.58	1.69	0.41	1.07	0.72	2.06
	(0.04)	(0.08)	(0.12)	(0.13)	(0.06)	(0.10)	(0.08)	(0.14)
% Overweight risky ^a			0.59	0.56			2.21	3.75
			(0.08)	(0.07)			(0.15)	(0.19)

Table continues

The differences between urban and rural areas are striking. The prevalence of overweight or obesity in men in urban areas is more than twice of that in rural areas. In 1997 the percent overweight is 29% for urban men and 12% for rural men. The difference between women is also large (14 to 27% in urban women and 10 to

Table 1.1 (cont'd)

	Male				Female			
	1989	1991	1993	1997	1989	1991	1993	1997
Age 40-59 years								
Number of observations	405	1213	1225	1449	415	1280	1341	1542
Median BMI	21.41	21.45	21.56	22.09	22.19	22.04	22.21	22.72
Mean BMI	21.56	21.88	22.11	22.55	22.51	22.67	22.78	23.09
	(2.59)	(3.04)	(3.63)	(3.42)	(3.12)	(3.88)	(3.95)	(3.82)
% Undernourished (<18.5)	11.36	8.41	8.00	5.52	7.47	9.45	7.01	6.29
	(0.32)	(0.28)	(0.27)	(0.23)	(0.26)	(0.29)	(0.26)	(0.24)
% Overweight (≥25)	9.14	13.36	14.12	19.05	20.00	21.88	22.07	25.23
	(0.29)	(0.34)	(0.35)	(0.39)	(0.40)	(0.41)	(0.41)	(0.43)
% Obese (≥30)	0.74%	1.07	1.47	2.21	2.41	3.52	3.21	3.18
	(0.09)	(0.10)	(0.12)	(0.15)	(0.15)	(0.18)	(0.18)	(0.18)
% Overweight risky ^a			0.16	1.04			9.25	10.44
			(0.04)	(0.10)			(0.29)	(0.31)
Age ≥60 years								
Number of observations		553	585	652		574	585	738
Median BMI		21.05	21.22	21.80		21.46	21.58	22.15
Mean BMI		21.59	21.78	22.51		22.03	22.17	22.69
		(3.70)	(3.72)	(4.21)		(4.33)	(4.29)	(4.58)
% Undernourished (<18.5)		20.07	14.53	11.66		18.82	15.38	13.14
		(0.40)	(0.35)	(0.32)		(0.39)	(0.36)	(0.34)
% Overweight (≥25)		15.73	15.21	22.70		20.21	20.51	25.34
		(0.36)	(0.36)	(0.42)		(0.40)	(0.40)	(0.44)
% Obese (≥30)		1.99	1.88	3.22		4.01	2.56	4.61
		(0.14)	(0.14)	(0.18)		(0.20)	(0.16)	(0.21)
% Overweight risky ^a			1.88	3.37			11.79	15.99
			(0.14)	(0.18)			(0.32)	(0.37)
Urban Areas								
Number of observations	665	1223	1051	1148	732	1346	1151	1264
Median BMI	21.05	21.82	21.97	22.86	21.49	22.01	22.17	22.86
Mean BMI	21.36	22.21	22.58	23.32	21.85	22.61	22.76	23.28
	(2.67)	(3.47)	(4.08)	(3.94)	(2.87)	(3.81)	(3.92)	(4.38)
% Undernourished (<18.5)	10.83	9.65	8.47	5.84	10.38	9.66	7.65	6.57
	(0.31)	(0.30)	(0.28)	(0.23)	(0.31)	(0.30)	(0.27)	(0.25)
% Overweight (≥25)	9.47	17.83	20.36	28.57	14.21	22.73	23.28	26.82
	(0.29)	(0.38)	(0.40)	(0.45)	(0.35)	(0.42)	(0.42)	(0.44)
% Obese (≥30)	0.30	1.55	2.57	3.22	1.09	3.42	3.30	4.03
	(0.05)	(0.12)	(0.16)	(0.18)	(0.10)	(0.18)	(0.18)	(0.20)
% Overweight risky ^a			1.14	2.53			10.43	13.13
			(0.11)	(0.16)			(0.31)	(0.34)
Rural Areas								
Number of observations	1531	2314	2278	2550	1622	2473	2446	2618
Median BMI	20.98	20.94	21.10	21.45	21.44	21.28	21.43	21.77
Mean BMI	21.18	21.28	21.56	21.89	21.69	21.82	21.86	22.36
	(2.08)	(2.77)	(3.30)	(3.26)	(2.54)	(4.61)	(3.57)	(3.92)
% Undernourished (<18.5)	6.99	9.90	8.56	7.88	8.69	11.44	10.92	8.29
	(0.26)	(0.30)	(0.28)	(0.27)	(0.28)	(0.32)	(0.31)	(0.28)
% Overweight (≥25)	4.83	7.22	8.30	12.39	9.56	12.41	13.08	18.14
	(0.21)	(0.26)	(0.28)	(0.33)	(0.29)	(0.33)	(0.34)	(0.39)
% Obese (≥30)	0.26	0.73	1.14	1.69	0.62	1.74	1.31	2.48
	(0.05)	(0.09)	(0.11)	(0.13)	(0.08)	(0.13)	(0.11)	(0.16)
% Overweight risky ^a			0.44	0.67			4.50	6.61
			(0.07)	(0.08)			(0.21)	(0.25)

Table continues

Table 1.1 (cont'd)

	Male				Female			
	1989	1991	1993	1997	1989	1991	1993	1997
No Formal Education								
Number of observations	76	392	335	299	361	1226	1094	1034
Median BMI	21.02	21.29	21.23	21.30	21.71	21.51	21.60	21.93
Mean BMI	21.46	21.71	21.77	21.99	22.08	22.06	22.09	22.54
	(2.26)	(3.22)	(3.39)	(3.86)	(2.78)	(3.75)	(3.71)	(4.58)
% Undernourished (<18.5)	3.95	10.71	9.55	11.37	6.65	13.95	12.07	11.90
	(0.20)	(0.31)	(0.29)	(0.32)	(0.25)	(0.35)	(0.33)	(0.32)
% Overweight (≥25)	7.89	13.27	13.13	17.73	12.47	17.94	17.64	22.05
	(0.27)	(0.34)	(0.34)	(0.38)	(0.33)	(0.38)	(0.38)	(0.41)
% Obese (≥30)	0.00	1.53	1.19	1.00	1.39	3.10	2.01	3.48
	(0.00)	(0.12)	(0.11)	(0.10)	(0.12)	(0.17)	(0.14)	(0.18)
% Overweight risky ^a			0.30	0.67			8.04	11.41
			(0.05)	(0.08)			(0.27)	(0.32)
Some Primary School								
Number of observations	291	624	775	548	417	584	807	606
Median BMI	21.00	20.83	21.20	21.39	21.64	21.57	21.75	22.04
Mean BMI	20.92	21.34	21.89	21.81	21.82	22.10	22.26	22.65
	(2.04)	(3.37)	(4.01)	(3.14)	(2.61)	(3.92)	(3.79)	(3.60)
% Undernourished (<18.5)	9.62	13.94	10.58	10.40	9.35	10.79	9.05	7.10
	(0.30)	(0.35)	(0.31)	(0.31)	(0.29)	(0.31)	(0.29)	(0.26)
% Overweight (≥25)	3.09	11.22	12.90	11.13	9.59	13.18	15.74	22.44
	(0.17)	(0.32)	(0.34)	(0.31)	(0.29)	(0.34)	(0.36)	(0.42)
% Obese (≥30)	0.00	0.96	2.71	1.46	1.20	2.57	2.11	3.14
	(0.00)	(0.10)	(0.16)	(0.12)	(0.11)	(0.16)	(0.14)	(0.17)
% Overweight risky ^a			1.29	1.28			6.44	10.07
			(0.11)	(0.11)			(0.25)	(0.30)
Primary Degree								
Number of observations	501	765	589	767	484	663	458	731
Median BMI	21.06	21.01	21.37	21.54	21.63	22.02	21.95	22.55
Mean BMI	21.26	21.41	21.65	22.10	21.86	22.61	22.70	22.85
	(2.21)	(2.68)	(3.21)	(3.60)	(2.58)	(6.77)	(4.32)	(3.40)
% Undernourished (<18.5)	8.18	10.46	8.83	8.21	7.64	7.84	6.77	5.61
	(0.27)	(0.31)	(0.28)	(0.27)	(0.27)	(0.27)	(0.25)	(0.23)
% Overweight (≥25)	5.79	9.41	9.00	13.56	12.40	17.04	19.00	22.98
	(0.23)	(0.29)	(0.29)	(0.34)	(0.33)	(0.38)	(0.39)	(0.42)
% Obese (≥30)	0.60	0.65	1.02	2.35	0.62	2.26	2.40	2.87
	(0.08)	(0.08)	(0.10)	(0.15)	(0.08)	(0.15)	(0.15)	(0.17)
% Overweight risky ^a			0.51	0.78			7.21	8.89
			(0.07)	(0.09)			(0.26)	(0.28)
Lower Middle School								
Number of observations	812	1058	1042	1224	689	846	808	915
Median BMI	20.85	21.05	21.20	21.66	21.34	21.36	21.48	21.82
Mean BMI	21.18	21.56	21.82	22.26	21.64	21.88	22.01	22.62
	(2.31)	(3.05)	(3.51)	(3.39)	(2.65)	(3.34)	(3.54)	(4.48)
% Undernourished (<18.5)	8.13	8.98	7.10	5.56	10.16	8.87	9.78	6.34
	(0.27)	(0.29)	(0.26)	(0.23)	(0.30)	(0.28)	(0.30)	(0.24)
% Overweight (≥25)	6.28	8.79	11.04	16.34	11.18	14.78	14.23	18.25
	(0.24)	(0.28)	(0.31)	(0.37)	(0.32)	(0.36)	(0.35)	(0.39)
% Obese (≥30)	0.37	1.32	1.34	1.80	0.44	1.54	1.98	3.06
	(0.06)	(0.11)	(0.12)	(0.13)	(0.07)	(0.12)	(0.14)	(0.17)
% Overweight risky ^a			0.48	0.90			5.07	6.67
			(0.07)	(0.09)			(0.22)	(0.25)

Table continues

Table 1.1 (cont'd)

	Male				Female			
	1989	1991	1993	1997	1989	1991	1993	1997
Higher Middle School								
Number of observations	335	439	387	491	273	342	294	346
Median BMI	21.01	21.45	21.60	22.45	21.10	20.93	21.34	22.19
Mean BMI	21.28	21.78	22.17	22.91	21.41	21.82	21.65	22.81
	(2.35)	(2.95)	(3.81)	(3.91)	(2.58)	(3.73)	(2.71)	(4.27)
% Undernourished (<18.5)	8.06	6.38	6.98	5.70	12.82	10.82	9.18	4.91
	(0.27)	(0.24)	(0.26)	(0.23)	(0.33)	(0.31)	(0.29)	(0.22)
% Overweight (≥ 25)	7.76	12.07	13.18	24.03	9.52	14.62	15.31	19.36
	(0.27)	(0.33)	(0.34)	(0.43)	(0.29)	(0.35)	(0.36)	(0.40)
% Obese (≥ 30)	0.00	0.46	1.81	2.44	0.00	2.05	0.00	2.31
	(0.00)	(0.07)	(0.13)	(0.15)	(0.00)	(0.14)	(0.00)	(0.15)
% Overweight risky ^a			0.52	2.04			2.38	4.91
			(0.07)	(0.14)			(0.15)	(0.22)
Tech/College+								
Number of observations	181	259	201	369	130	158	136	250
Median BMI	21.50	22.14	22.21	22.88	20.81	21.09	20.93	22.14
Mean BMI	21.72	22.50	22.48	23.39	21.29	21.92	21.91	22.56
	(2.45)	(3.15)	(3.15)	(3.45)	(2.68)	(3.05)	(3.71)	(3.23)
% Undernourished (<18.5)	7.73	5.79	8.46	4.88	9.23	9.49	9.56	7.20
	(0.27)	(0.23)	(0.28)	(0.22)	(0.29)	(0.29)	(0.30)	(0.26)
% Overweight (≥ 25)	8.84	17.37	19.90	29.27	8.46	17.72	15.44	19.20
	(0.28)	(0.38)	(0.40)	(0.46)	(0.28)	(0.38)	(0.36)	(0.39)
% Obese (≥ 30)	0.00	1.16	0.50	4.61	1.54	0.63	2.94	1.60
	(0.00)	(0.11)	(0.07)	(0.21)	(0.12)	(0.08)	(0.17)	(0.13)
% Overweight risky ^a			0.50	2.71			6.62	6.80
			(0.07)	(0.16)			(0.25)	(0.25)

Note: Standard deviations are in parentheses. (a) Overweight risky or having high abdominal fat is defined as having waist circumference > 102 cm for male or >88 cm for female. Waist circumferences were not measured before 1993.

18% in rural women). Except for the year 1989 rural areas had higher prevalence of underweight than urban areas. Even though men and women in rural areas are more likely to engage in strenuous work they do not have higher BMIs. Undernutrition in the rural areas is still about 8 percent although it had been decreasing over the years. Different patterns of BMI distributions between urban and rural areas suggests different needs in care and facilities in different regions.

The second part of Table 1.1 shows the distribution of BMI over different education groups for men and women. The prevalence of overweight men is the highest among the highly educated (29%) but for women it is among those with primary or less than primary education (22%). Patterns for obesity are similar. For men the

percent underweight peaked at the group with some primary schooling (11%) and for women those without any formal education had the highest prevalence in undernutrition (12%). This suggests higher education is protective in terms of undernutrition but harmful in terms of overweight for men; and for women higher education may help to reduce both under- and overweight.

Hypertension

Overweight and obesity lead to adverse metabolic effects on blood pressure and cholesterol level. The Risk of coronary heart disease increases steadily with increasing BMI. Adults with systolic blood pressure at or above 140 mmHg or diastolic blood pressure at or above 90 mmHg, or adults taking antihypertensive agents are considered having hypertension.¹⁸ Table 1.2 summarizes the prevalence of hypertension in men and women in our sample. In the overall sample the rates increased from 8 to 23% in men and from 6 to 19% in women. There is a gap between men and women at younger age but the gap closed up as people grow older. In the 20 to 39 age group the percent of men having hypertension is twice of that of the women; in the 40 to 59 age group and the elderly the difference is about 4 percentage points. The prevalence is higher for older men and women than the younger ones. Less than 10 percent men and women have high blood pressure when they are less than 40 years old, but after age 60 half of the men and women are hypertensive.¹⁹

Rural and urban areas again have significant differences in prevalence of hypertension. The percent of men and women with high blood pressure in urban areas is almost twice as high in 1991 and 40 percent higher than that of the rural areas in 1997. Combined risks of overweight and hypertension in urban residents put more

¹⁸In 1989 the question about using antihypertensive agents was not asked. The definition in 1989 is hence based on blood pressure only. In 1991 to 97 blood pressures were taken three times for accuracy. The average of these measures are used for the definition.

¹⁹There is a dip in the prevalence of hypertension in 1993 among all age groups because of a lower diastolic average. We are not sure why this is the case.

stress on the health care facilities in those areas and the trend is increasing.

Across different education level the risks of hypertension is the highest for the least educated, 39% for men and 32% for women with no formal education. Interestingly education ceases to be protective at the college or higher level. There is steady decrease in prevalence of hypertension until people with higher middle school diploma. This pattern is the same for men and women as oppose to the different patterns in the BMI distribution across education levels for men and women.

Table 1.2: Prevalence of Adult High Blood Pressure in CHNS 1989-1997

	Male				Female			
	1989	1991	1993	1997	1989	1991	1993	1997
All Sample								
Number of observations	2193	3531	3326	3691	2334	3816	3592	3875
% High Blood Pressure	8.39 (0.28)	16.26 (0.37)	10.34 (0.30)	22.87 (0.42)	5.53 (0.23)	13.36 (0.34)	10.38 (0.31)	18.94 (0.39)
Age 20-39 years								
Number of observations	1788	1767	1518	1594	1925	1962	1668	1599
% High Blood Pressure	7.33 (0.26)	6.28 (0.24)	2.50 (0.16)	10.73 (0.31)	4.00 (0.20)	3.01 (0.17)	1.32 (0.11)	4.82 (0.21)
Age 40-59 years								
Number of observations	405	1211	1225	1447	409	1280	1341	1539
% High Blood Pressure	13.09 (0.34)	18.33 (0.39)	8.41 (0.28)	23.43 (0.42)	12.71 (0.33)	16.41 (0.37)	10.89 (0.31)	19.95 (0.40)
Age >=60 years								
Number of observations		553	583	650		574	583	737
% High Blood Pressure		43.58 (0.50)	34.82 (0.48)	51.38 (0.50)		41.99 (0.49)	35.16 (0.48)	47.49 (0.50)
Urban Areas								
Number of observations	664	1222	1049	1147	723	1346	1149	1263
% High Blood Pressure	10.54 (0.31)	23.16 (0.42)	16.11 (0.37)	28.42 (0.45)	6.09 (0.24)	19.17 (0.39)	16.01 (0.37)	23.67 (0.43)
Rural Areas								
Number of observations	1529	2309	2277	2544	1611	2470	2443	2612
% High Blood Pressure	7.46 (0.26)	12.60 (0.33)	7.69 (0.27)	20.36 (0.40)	5.28 (0.22)	10.20 (0.30)	7.74 (0.27)	16.65 (0.37)
No Formal Education								
Number of observations	76	392	334	299	357	1226	1092	1032
% High Blood Pressure	13.16 (0.34)	29.34 (0.46)	18.86 (0.39)	39.46 (0.49)	7.00 (0.26)	23.57 (0.42)	18.96 (0.39)	32.36 (0.47)
Some Primary Education								
Number of observations	291	623	774	545	414	583	805	605
% High Blood Pressure	8.25 (0.28)	17.66 (0.38)	14.99 (0.36)	26.61 (0.44)	7.49 (0.26)	12.01 (0.33)	10.31 (0.30)	18.84 (0.39)
Primary Degree								
Number of observations	501	764	589	766	479	661	458	731
% High Blood Pressure	8.38 (0.28)	16.75 (0.37)	9.00 (0.29)	25.20 (0.43)	6.47 (0.25)	11.80 (0.32)	8.30 (0.28)	17.65 (0.38)
Lower Middle School								
Number of observations	811	1056	1041	1222	685	846	807	911
% High Blood Pressure	7.52 (0.26)	11.08 (0.31)	6.44 (0.25)	17.68 (0.38)	4.09 (0.20)	4.73 (0.21)	3.10 (0.17)	10.21 (0.30)
Higher Middle School								
Number of observations	334	437	387	491	271	342	294	346
% High Blood Pressure	9.58 (0.29)	13.50 (0.34)	7.24 (0.26)	17.52 (0.38)	2.95 (0.17)	4.68 (0.21)	3.06 (0.17)	8.09 (0.27)
Tech/College								
Number of observations	180	259	201	368	128	158	136	250
% High Blood Pressure	8.33 (0.28)	17.37 (0.38)	8.46 (0.28)	23.37 (0.42)	4.69 (0.21)	10.76 (0.31)	8.09 (0.27)	14.40 (0.35)

Note: Adults with systolic blood pressure at or above 140 mmHg or diastolic blood pressure at or above 90 mmHg, or adults taking antihypertensive agents are considered having hypertension. In 1989 the question about using antihypertensive agents was not asked.

Smoking

Current and past smoking behaviors for cigarettes or pipes were asked in 1991, 93 and 97 surveys. Table 1.3 summarizes the percentage of people who at the time of the interview smoked cigarettes or pipes. The cumulative hazards of smoking depends on several factors including age at which smoking began, duration of smoking, number of cigarettes smoked per day, and the tar and nicotine content of cigarettes. Therefore current prevalence of smoking is a proxy for the cumulative hazards. Smoking causes substantially increased risk of mortality from lung cancer, heart disease, chronic respiratory diseases and second-hand smoking is also a health risk.

There is a slight decrease in prevalence of smoking in men but a slight increase in women, with the overall percentage much higher in men (66%) than in women (5%). This is generally the pattern in Indonesia and other Asia countries (Strauss et al. 2004). For men the prevalence is the highest at prime age and decreases after age 60; for women, however, the prevalence is the highest among the elderly. The difference between urban and rural men is significant (rural men with 6 to 10 percentage point higher) but not significant between urban and rural women.

The effect of education on smoking is negative and graded among women with higher educated women less likely to smoke; and among men only those with college or higher education had a slightly lower prevalence of smoking. It seems that smoking for men is a habit much more difficult to break. The decreased trend in smoking in men is true in all education categories. However, the trend in women is different. At the lower education levels (less than primary degree) the prevalence decreases in 1991 to 93 and increases in 1997; and at the higher education levels (more than middle school diploma) the prevalence increases first and then decreases.

Table 1.3: Adult Smoking Status in CHNS 1991-1997

	Male			Female		
	1991	1993	1997	1991	1993	1997
All Sample						
Number of observations	3537	3329	3689	3819	3597	3873
% Smoking	68.82%	66.60%	62.27%	4.58%	4.56%	4.73%
	(0.46)	(0.47)	(0.48)	(0.21)	(0.21)	(0.21)
Age 20-39 years						
Number of observations	1771	1519	1594	1965	1671	1600
% Smoking	69.45%	65.96%	62.17%	0.87%	1.20%	2.50%
	(0.46)	(0.47)	(0.49)	(0.09)	(0.11)	(0.16)
Age 40-59 years						
Number of observations	1213	1225	1445	1280	1341	1537
% Smoking	72.79%	71.43%	67.82%	6.88%	6.41%	4.68%
	(0.45)	(0.45)	(0.47)	(0.25)	(0.25)	(0.21)
Age >=60 years						
Number of observations	553	585	650	574	585	736
% Smoking	58.05%	58.12%	50.15%	12.20%	9.91%	9.65%
	(0.49)	(0.49)	(0.50)	(0.33)	(0.30)	(0.30)
Urban Areas						
Number of observations	1223	1051	1145	1346	1151	1261
% Smoking	64.76%	61.94%	54.93%	5.50%	5.47%	4.28%
	(0.48)	(0.49)	(0.50)	(0.23)	(0.23)	(0.20)
Rural Areas						
Number of observations	2314	2278	2544	2473	2446	2612
% Smoking	70.96%	68.74%	65.57%	4.08%	4.13%	4.94%
	(0.45)	(0.46)	(0.48)	(0.20)	(0.20)	(0.22)
No Formal Education						
Number of observations	392	335	298	1226	1094	1029
% Smoking	66.33%	65.07%	58.72%	8.08%	7.13%	7.68%
	(0.47)	(0.48)	(0.49)	(0.27)	(0.26)	(0.27)
Some Primary Education						
Number of observations	624	775	544	584	807	604
% Smoking	71.31%	67.87%	66.91%	5.31%	4.83%	6.46%
	(0.45)	(0.47)	(0.47)	(0.22)	(0.21)	(0.25)
Primary Degree						
Number of observations	765	589	767	663	458	731
% Smoking	71.50%	70.29%	64.67%	4.07%	4.80%	5.88%
	(0.45)	(0.46)	(0.48)	(0.20)	(0.21)	(0.24)
Lower Middle School						
Number of observations	1058	1042	1220	846	808	913
% Smoking	69.00%	67.08%	64.26%	1.42%	2.10%	1.86%
	(0.46)	(0.47)	(0.48)	(0.12)	(0.14)	(0.14)
Higher Middle School						
Number of observations	439	387	491	342	294	346
% Smoking	66.97%	64.08%	60.29%	0.88%	1.70%	1.16%
	(0.47)	(0.48)	(0.49)	(0.09)	(0.13)	(0.11)
Tech/College						
Number of observations	259	201	369	158	136	250
% Smoking	61.00%	55.72%	49.32%	1.90%	2.21%	0.40%
	(0.49)	(0.50)	(0.50)	(0.14)	(0.15)	(0.06)

Note: Smoking status for people who at the time of the interview smoked cigarettes or pipes was not surveyed in 1989.

Household Characteristics

Characteristics of household resources in the four years of the CHNS are summarized in Table 1.4. This period marks China's most noteworthy economic growth and expansion. From 1989 to 1997 the median real per capita income in 1988 currency increased from 424 yuan to 711 yuan and the mean went from 677 to 1039 yuan.²⁰ The increase in rural areas is bigger in percentage in rural areas (77% in mean and 84% in median) than in urban areas (30% in mean and 60% in median). The gap between urban and rural areas is decreasing. The median of real per capita income in rural areas in 1989 is 35% of the median in urban areas and the percentage increases to 40%. The mean ratios went up from 0.44 to 0.59. Although the gap is still large it is closing.

Real productive assets are discounted to the 1988 value.²¹ The mean real productive assets increased from 243 yuan in 1989 to 636 yuan in 1997. Rural households have more productive assets for engaging in agricultural activities and small businesses. The differences between rural and urban areas are getting smaller. From 1989 to 1997 productive assets increased significantly in both rural and urban areas, from 130 to 470 yuan on average in urban and from 294 to 714 yuan in rural areas. The amount of land cultivated each year also increased but urban households were involved with very little farming. The amount of land farmed in the rural areas may be a good proxy for income where there is no or little other opportunities for small businesses or off-farm labor. Combining productive assets and the amount of land farmed we have a proxy for the household resources.

²⁰Price index is based on the China Statistical Yearbook (SSB 1988-1998) for urban and rural areas at the provincial level. The total income in CHNS includes both earned and unearned income. Categories surveyed in each year are: income from wage, home gardening, household farms, farming collectives, raising livestock/poultry, household fishing, fishing collectives, household small business, welfare subsidies and income from other sources.

²¹The items included each year for productive assets are tricycles, motorcycles, tractors or walking tractors, irrigation equipment, power threshers and water pumps. Values are self-reported purchasing values in 1989 and current worths in all other years.

Individual Socioeconomic Characteristics

Since in 1989 our sample includes only those who were less than 46 years old and with higher education level than adults of all ages, when comparing the trends in education level we look at the year from 1991 to 97. The average years of education increased by about half a year. The overall prevalence of people with no formal education decreased from 22 to 18 percent. Between 91 and 97 the percent of adults with less than primary degree decreased and the percentages of all other higher levels of education increased.

There is some gender differential in education level between men and women. Men had 1.7 years more of education than women in 1991, 2 years more in 1993 and 2.1 years more in 1997. The percent of men in lower education levels (no formal education, some primary schooling and primary degree) decreased and the percent of men in other educational levels (middle school degrees, college and higher degrees) increased. For women only the percent of no formal education decreased in prevalence.

The employment rates decreased from 82% in 1991 to 77% in 1997 due partly to aging. The employment rates for men in all years were about 10 percentage points higher than those for women. There is little change in percent of married men and women over time. In our sample the increase in divorce rates is small (less than 1 percent).

Community Characteristics

Table A.4 summarizes some characteristics of the communities in 1989 to 1997. There is a dip in 1989 in the increasing trend of price of rice, wheat, beef and fish, but not for pork which had increased the most. The price of eggs went down slightly. The hospitals and clinics in 1997 were not surveyed and hence the reduced-form analysis with community characteristics in the next section used only data from 1989 to 1993.

There is improvement in clean water availability in that the percent of households

within each community with water from water factories or underground was increasing over time and the percent with water from an open well, spring or river etc. was decreasing. The percent of households with in-house-flush toilet was also increasing, suggesting improvement in sanitation environment.

Table 1.4: Basic Characteristics of Households and Individuals in CHNS 1989-97

	1989	1991	1993	1997
All Households				
Number of observations	2736	3402	3164	3512
Real PC INC Mean (yuan)	676.96	658.86	822.24	1038.55
Median	423.93	495.74	535.34	711.04
Real Productive Assets Mean (yuan)	242.73	266.11	425.37	635.60
Land Cultivated Mean (mu)	2.49	2.54	2.65	3.20
Urban Households				
Number of observations	853	1114	977	1124
Real PC INC Mean (yuan)	1101.68	1014.26	1209.49	1435.06
Median	779.43	944.38	974.84	1253.26
Real Productive Assets Mean (yuan)	130.37	146.59	376.62	469.80
Land Cultivated Mean (mu)	0.26	0.22	0.20	0.15
Rural Households				
Number of observations	1883	2288	2187	2388
Real PC INC Mean (yuan)	484.56	485.82	649.25	851.91
Median	274.96	322.91	346.56	503.13
Real Productive Assets Mean (yuan)	293.62	324.30	447.15	713.64
Land Cultivated Mean (mu)	3.50	3.67	3.75	4.63

Table continues

Table 1.4 (cont'd)

	1989	1991	1993	1997
All Individuals				
Number of observations	4550	7356	6926	7580
Mean Age	31.74	41.86	43.06	44.11
Mean Years of Education	7.32	6.01	6.04	6.58
% No Formal Education	9.60	22.00	20.63	17.59
% Some Primary	15.56	16.42	22.84	15.22
% Primary Diploma	21.65	19.41	15.12	19.76
% Lower Middle School	32.99	25.88	26.71	28.22
% Higher Middle School	13.36	10.62	9.83	11.04
% Technical/College	6.84	5.67	4.87	8.17
% Employed	95.65	82.02	80.06	76.86
% Married	84.15	82.22	82.49	81.56
% Divorced	0.42	0.69	0.53	0.75
% Widowed	0.29	5.76	6.11	6.12
Men				
Number of observations	2196	3537	3329	3698
Mean Age	31.77	42.01	43.01	43.72
Mean Years of Education	8.15	6.97	7.11	7.71
% No Formal Education	3.46	11.08	10.06	8.09
% Some Primary	13.25	17.64	23.28	14.82
% Primary Diploma	22.81	21.63	17.69	20.74
% Lower Middle School	36.98	29.91	31.30	33.10
% Higher Middle School	15.26	12.41	11.63	13.28
% Technical/College	8.24	7.32	6.04	9.98
% Employed	97.71	86.45	84.40	82.51
% Married	81.33	83.04	83.24	81.64
% Divorced	0.46	0.65	0.51	0.81
% Widowed	0.18	3.22	3.00	2.81

Table continues

Table 1.4 (cont'd)

	1989	1991	1993	1997
Women				
Number of observations	2354	3819	3597	3882
Mean Age	31.71	41.73	43.10	44.48
Mean Years of Education	6.52	5.21	5.14	5.58
% No Formal Education	15.34	32.10	30.41	26.64
% Some Primary	17.71	15.29	22.44	15.61
% Primary Diploma	20.56	17.36	12.73	18.83
% Lower Middle School	29.27	22.15	22.46	23.57
% Higher Middle School	11.60	8.96	8.17	8.91
% Technical/College	5.52	4.14	3.78	6.44
% Employed	93.73	77.92	76.05	71.49
% Married	86.79	81.46	81.79	81.48
% Divorced	0.38	0.73	0.56	0.70
% Widowed	0.38	8.12	8.98	9.27

Note: Productive assets in 1989 are calculated differently from in other years. All productive assets owned by 1989 is evaluated at last year's purchasing prices. In the other years the productive assets are the reported total value. Income and productive assets are discounted to year 1988. Land is the total amount (1 mu=667 square meters) farmed by the household in the previous year.

1.5 Reduced Form Analyses

As was briefly outlined in section two, the estimation of the reduced-form BMI demand function (1.7) uses the OLS method on data pooled over all survey years. To account for error correlations for each individual over time the individual-level cluster robust standard errors are used in all regressions. In all regressions age dummies and five-year cohort dummies are included. In a separate paper (Luo 2003a) we consider the classic identification problem in estimating age, cohort and year effects in a model. We chose the five-year cohort dummies so that the perfect linear dependency of the three variables no longer exists and the system can be identified. Surely there are many other identification strategies as noted in our paper but we chose the five-year cohorts around the early 1960s' famine and the cultural revolution period as being one meaningful strategy. In a model including only the age dummies, five-year cohort dummies and year dummies the estimated BMI age profiles for men and women are shown in Figure 1.1. Both profiles are inverse U-shaped. The age effect for women peaked around 45 to 50 and for men at a little later ages but decreasing at a slower rate.

First, the basic specification focuses on the effects of individual education and household covariates - productive assets and land, controlling for community level factors by community dummies or community year interaction dummies. This was done for the whole sample, for urban and rural areas separately, and for each age group (20-39, 40-59 and 60+) of men and women (Table 1.5 to 1.9). These stratifications enable us to find differences between different groups of people and the rural-urban stratification is consistent with the sampling process because the rural households were over-sampled. As outlined in section 2 the community dummies capture the effects of unobserved past and future community level information, assuming there is no variation over time. When such assumption seems unattainable using the community year interaction dummies partially resolves it based on a weaker assumption

that the variation into the future follows the same path in the past.

The marital status of the individual is added to the baseline model to examine whether it had incremental effects controlling for other factors. The marital status variables are not included in the basic modelling due to the concern of potential selection biases with assortative marriage behavior. In rural areas men with higher BMI may be considered having more ability to provide for the family and hence are more likely to be married. For young urban women being slim could be more attractive in both marriage market and labor market. The endogenous biases can go both ways. Although from the descriptive statistics in the previous section there is little change over time in marriage rates it is commonly perceived the divorce rates are increasing in China after the reform.

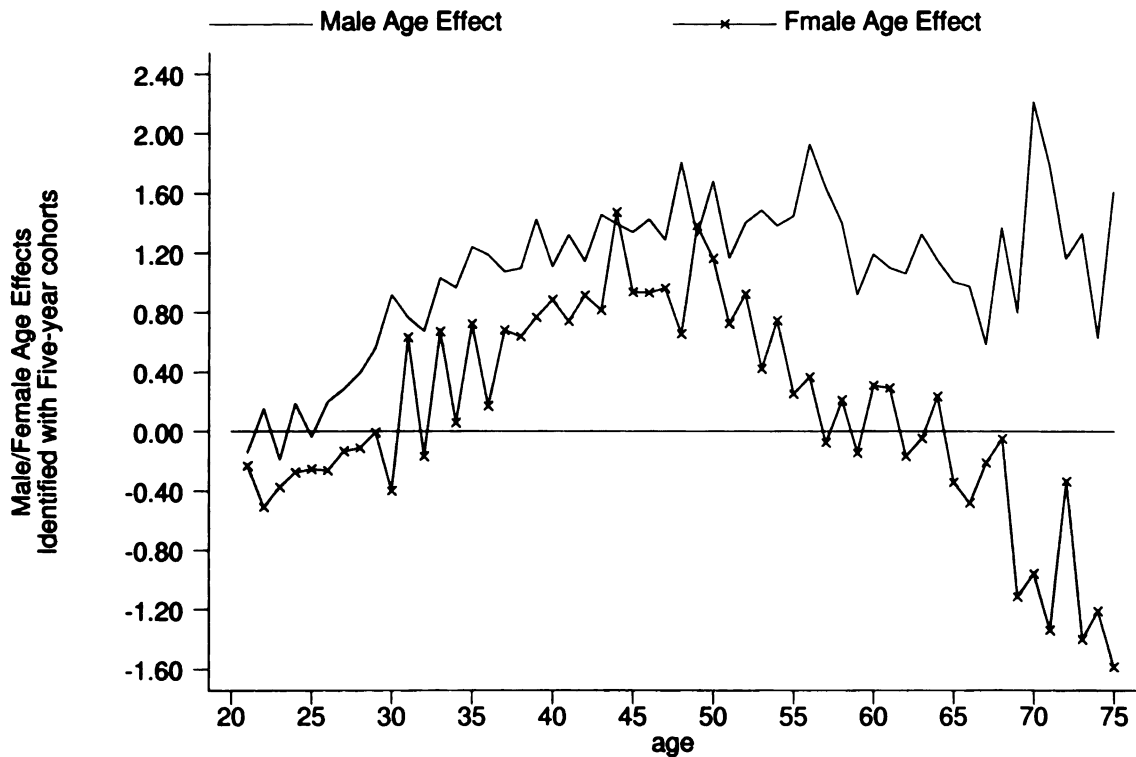


Figure 1.1: Male and Female Age Effects Identified through the specification with age dummies, five-year cohort dummies and year dummies.

Secondly, adding to the basic model the community characteristics including food prices, clinic conditions, water, toilet and sanitation conditions, the augmented model studies the joint effects of individual, household and community variables on adult BMIs. As noted in section two, these added community variables are only current prices and infrastructure. Community dummies are used to represent the unique path of past and future distributions in all excluded factors. Since the effect of food prices on BMI can be either positive or negative depending on whether income effect or substitution effect is bigger, it warrants an empirical testing to see what the actual effects are. For communities with better health facilities, water and sanitation conditions we'd expect better environment being beneficial assuming there is no selective migration to a better health environment and no program placement in response to poor health. The first assumption is not unreasonable in the early 1990s in China. Though increasingly the restriction of household registry has been relaxed, the migration from rural to urban area in China is still largely temporary and not primarily for health reasons. The second assumption may be problematic in rural areas if the government did take effort in responding to local need by investing more in health facilities or increasing the number or the training of health professionals. All community characteristic variables are treated as exogenous.

1.5.1 Education

Whole Sample

A set of dummies for the highest degree obtained with no formal education as reference is used to assess the nonlinear effects of education on BMI.²² In the overall pooled sample for women (Table 1.5 column 1) having some primary education is associated with 0.23 unit higher BMIs than those with no formal education at 0.1

²²A linear term in completed years of education, or a linear spline of years of education are also tried and found to fit less well than the set of dummies for highest degree obtained.

significant level, but the impact rises significantly when one finishes primary schooling (0.47 unit). To put things into perspective, for a 5.5 foot woman a 0.47 unit increase in BMI is equivalent to a 3-pound increase in weight assuming fixed height. Completing senior high school, technical school or college is associated with a lower BMI than having no education (-0.02 and -0.7) or lower education levels, similar to the finding of Thomas et al. (1996) for women in Côte d'Ivoire. On average a 5.5 foot woman with a college equivalent degree weighs 4 pounds lighter than an uneducated counterpart. With community dummies or with community-year interaction dummies the strong education effects remain at the same magnitude and are always jointly significant. The nonlinear effect of education for women needs further investigation. Schooling can affect health status either by raising the technical efficiency with which inputs are used or by increasing the allocative efficiency of input use leading to increased incomes (Welch 1970) or increasing women's power and influence in household decision-making (Behrman and Wolfe 1984). A reduced form analysis does not address the question of how education affects the outcome, however we do observe for whom the effect of education is the strongest.

For men education does not have any statistically significant effect for any geographic or age group, although in column 1 and 2 in Table 1.5 for men there is a graded positive effect of education before the college level. It may suggest that there is a different learning curve for men from women, or that BMIs of men are affected not by his own education but by the education of the person who prepares the meal or chooses the diet pattern. Without knowing who is the decision maker in such matters in a household we can only hypothesize these channels. The F-statistic for testing equality of coefficients between men and women confirms the significant differences.

Urban and Rural Areas

In Table 1.6, the effect of education remains mainly in the same direction for urban women except for those with some primary schooling, but none of the effects are statistically significant. The effects for men changed shapes to high school educated men with highest BMI followed by those not formally educated. The fact that education in urban areas does not affect BMI significantly for men and women indicates the most effective way of preventing overweight and obesity in urban areas may not lie in improvement on formal education.

A closer look at Table A.3 in the appendix A reviews that the patterns of BMI distribution across different education groups for women in the whole sample are the same for the patterns in the rural sample, but are different than that in the urban sample (not statistically significant with a p-value from the F-statistic being 0.28). That is why in Table 1.7 the effects of all education levels for rural women are of the same pattern as that in the overall sample. Although the raw distribution of BMI across different education levels for rural women indicates that the small number of women with college equivalent education had a higher BMI than those with no formal education, the conditional effect of college education in Table 1.7 column 1 is strong and negative. Since the prevalence of overweight for rural females is the highest among the no education group, the decreasing effects of higher education may be deemed as helpful in preventing obesity.

Comparing Table 1.6 with Table 1.7 the patterns of urban and rural male educational effects are the same for people with less than high school diploma and the effects switched signs between rural and urban men with higher educations. The urban education effects are bigger in magnitudes although none of them are significant and none of the differences between urban and rural men based on F-statistics are significant either. However, none of these patterns showed up in the overall sample regressions. The need for stratifying the sample is partly due to the clustering

sampling strategy of the data collectors.

Across Age Groups

As seen in Table 1.8 for younger generations there is a strong negative graded education effects and it starts to show at middle school level and is stronger in high school group (-0.8 compared to -0.4 in overall sample) and in the college level (-1 vs -0.7). To put it in perspective, a decrease of 1 unit in BMI for a 5.5-foot tall woman would result in a weight loss of about 6 pounds.

For the prime age women (Table 1.9) the effect of education only is significant for primary school completion and it is positive due perhaps to increased income. The women in this age group had a much higher proportion with no formal education (35%) as compared with the younger females (10%) and a higher proportion with primary schooling (21%) as compared with 17 percent in the younger generation. For all other categories of education the prime age women had a much lower percentage. The reason that the coefficients for middle school and higher education are not significant may be due to small sample sizes in those cells. Jointly the effect of education is significant at 10% level when we control for community dummies, not community-year interaction dummies.

For the elderly, education has no significant impact any more (Table 1.10), though the effect of higher education did change sign and became positive.

Some of the differences of educational effects among women in different age groups may be due to aging. As we have seen the age BMI profile for women peaks in the prime aged group. The protective effects of education wear off over the years as women age. It could also be because in the younger cohort the education effects are close to those in the developed countries if we think of education as a proxy for income. The younger generation in our sample had a much higher education level than older women (3 years more than the prime aged women and 6 years more than

the elderly). The strong and positive effects of some or complete primary education in women aged 40 to 59 are consistent with most findings in developing countries where increased income is associated with higher BMI. For the elderly this positive income effect only kicks in for those very highly educated.

For prime age men and elderly men the F-statistics for testing equality of coefficients between men and women suggest there is no systematic differences between them; and for the younger cohorts the difference is significant. Instead of the graded negative effects there is no particular pattern of the male education effects.

We do not claim our specification is complete; there can always be omitted factors in a given reduced form analysis. Generally a fixed-effect model can be used if there is more than one year of data. However, because education does not vary much over time it is difficult to identify its impact with a fixed effect model.²³

The strong nonlinear effect of education continues to exist after adding community characteristics in addition to the community dummies (see the following sections). Women with primary degrees are the ones with highest BMIs while those with technical or college degrees the lowest. The effect of education for female BMI over and above household resources and community characteristics is apparent.

Across Years

Is the effect of education getting close to that of the developed countries? Separate regressions for men and women in each round of the survey since 1991 were examined with age and community dummies included in all years. Heteroscedastic robust standard errors are in parentheses. According to the hypothesis we should see the effect of education grows more negative at each educational level as time goes by. If such was the case it could be due to increased technical efficiency. However as we can see in Table 1.11 it is not the case. If anything, the effects of education on

²³The same is true when it comes to identifying the marital premium on BMI for men.

female BMI actually decreased over time; and the effects of male education on BMI were positive and increased over the years, though not significant.

For men the effects are still close to that of a typical developing country and are actually increasingly so though not significant statistically. For both men and women there is room for improvement through bettering education.

1.5.2 Household Resources

Poorer households live in areas where fewer health facilities and practitioners are available and the living and sanitary conditions are worse. If such is the case, it is essential to control for household resources so that the estimates of community characteristics do not reflect household characteristics. Instead of using non-wage income²⁴ we control household resources using log of real productive assets discounted to 1988 value as indicators for long-run resources. Splines around the median of the positive values of real productive assets and the amount of land cultivated (mu) are included in all regressions. The coefficients for the lower and upper spline of productive assets represent the level effects, not the marginal effects. Interestingly, the effects of productive assets vary across different strata as well.

Whole Sample

For the whole sample (Table 1.5), the effect differs between men and women. For men, more productive assets are associated with higher BMI, with larger effects in the upper spline range and the joint effects are highly significant. For women, higher level of productive assets are associated with higher BMI only for those with assets less than the median; above the median the association is negative but not significant; and the joint effects are only significant at 0.1 level. The amount of land farmed was never significant statistically and had very small impact, though again

²⁴Real per capita income splines were also tried but not reported due to potential measurement errors in and endogeneity bias due to income.

men and women have opposite signs (negative for men).

The splines of real productive assets were used to capture the non-linear effects of household resources on BMI. Among those with more than median resources an increase of 100 yuan in real productive assets is associated with a 0.18 unit increase in BMI for men, so is the same increase among those with less than median resources for women. For a 6-foot man that's an increase of weight by 1.4 pound and for a 5.5-foot woman by 1.1 pound. As evident from the F-statistics for testing equality of coefficients of productive assets between men and women the difference is not significant.

The difference between different productive assets levels could be an artifact because of the high correlation between the two splines; or it could be capturing the nonlinear effects of assets nicely. The fact that they are jointly significant but not all individually suggests the former may be the case.

Urban and Rural Areas

In urban areas, the effect of productive assets and land was never significant (Table 1.6). This is due to the small proportion of urban household owning such assets and engaging in farming. Hence the results in the whole sample are mainly driven by the rural sample. The rural sample has the same pattern in productive asset effects as in the whole sample although with a larger magnitude (Table 1.7). However, the differences between urban and rural samples are not significant statistically.

Across Age Groups

For the young and prime aged men the productive assets above the median has a positive association with BMIs, and the amount of land farmed had different effects (negative for age 20 to 39 and positive for age 40 to 59). For women, the younger ones had a negative effect of higher productive assets and positive effects for lower

assets and the amount of land farmed (significant at 0.05 level), which is similar to the whole sample. For the middle aged women, different from the whole sample, higher productive assets in the above median assets range were correlated with a significantly higher BMI and the effect of land was different as well.

For the elderly men and women having productive assets above the median was associated with lower BMIs, though the effect was not significant. The effects of land again display different impacts on men and women BMI, but was similar to the whole sample.

All the above findings obtain with specifications including community or community-year interaction dummies. So the effect of productive assets is over and above the effect that was due to households living in better locations. The productive assets effects persist even after adding community level prices, water and sanitation variables.

There are several caveats in interpretation of these results. In a dynamic model BMI can affect productivity and hence productive assets in later periods. There could be selection into more strenuous vocations which pay higher wages. This would bias the coefficients upwards. Lastly measurement errors in productive assets could give rise to attenuation bias in the estimation. All these possible biases lead in different directions. Without good identifying instruments it is difficult to make any statement of the direction or magnitude of the estimates. However the association between productive assets and BMIs is strong and consistent with findings in other developing countries (Thomas et al. 1996).

1.5.3 Marital Status

As stated at the beginning of this section, adding marital status to the baseline specification demands careful interpretation of results. The fact that for women across all sample and sub-samples the coefficients of all other variables are not significantly affected could be evidence that the effects of marriage or divorce on female BMI were

not by way of other individual or household characteristics. However the significant marital effect for men did decrease the effects of education on male BMI.

Married men in all samples except for the elderly had a significantly higher BMI than those who were single, ranging from 0.3 to 1 unit of BMI. The divorced or separated men in all samples except for the prime aged had a lower BMI than the singles, though the differences were not significant in any case. The widowed men and women in the urban sample had a significantly lower BMI than the urban singles. For women, the marital status had no significant effects on BMI, except that widows in the urban areas had a much lower BMI than those who were single. In fact, the single women in almost all samples had always had higher BMI.

Empirical research in the US has consistently shown that married men have substantially higher wages than otherwise similar unmarried men (?). One commonly cited hypothesis is that marriage allows the husband to specialize in market production and the wife specialize in home production, enabling married men to acquire more specific human capital and earn higher wages. In rural areas men with higher BMI may be considered having more ability to provide for the family and hence are more likely to be married. However in China both men and women are engaged in labor market activities, with employment rates being close to each other (Table 1.4). Hence if there is assortative mating it might exist for other traits of the partner sought. For young urban women being slim could be more attractive in both marriage market and labor market; for men being muscular impresses women more. The endogenous biases can go both ways.

Potentially the selection biases can be corrected if we have proper instruments for marital status but such case is rare. Since there was little variation in marital status we cannot examine the individual fixed-effect model for assortative marriage either.

1.5.4 Community Characteristics

In the first two CHNS surveys, state and free market data were collected. But by 1997, none of the communities had separate state prices so only free market stores were visited. In all cases prices were collected for a representative basket of commodities. We use free market prices of rice, wheat, eggs, pork, beef and fish (all for the most commonly consumed types) to represent price levels in each community. Due to high correlations between rice and wheat prices, and pork and beef prices only one price from each food group was used in the model.

In 1989, 1991 and 1993, separate visits were made to obtain in-depth data in each community for every identified health service and family planning provider or facility. Information was collected concerning personnel, sources of funds, services available, prices (asked separately for insured and self-pay patients), and distance to the primary sampling units served by the facility. These were discontinued in 1997. Hence our analysis with community characteristics in addition to community dummies only include three rounds of the survey.

The types and sources of households water and toilet facility are aggregated from 40 to 80 households in each community. They represent the percent of households within a community with water from underground, etc. They are measures of quality of water and sanitation in the community. Using the aggregated measures helps avoid the selection biases due to migration.

In columns 1 and 3 of Table 1.12, the community dummies are not included, but they are in columns 2 and 4 to control for unobserved community variables that might explain where programs are located (Pitt et al. 1993). As we can see most of the significant effects of community prices and water and sanitation qualities disappear after adding community dummies. This might be evidence that community unobservables should be controlled for and/or that the variation of prices over time is not strong before 1993. Although grain ration prices were raised twice in 1991 and

1992, each time by an average of 50 percent, the free market prices were not affected as much (Ke 1999). Therefore, the food price effects on BMI can go in different directions which is why empirical studies are important.

Before controlling for community dummies the effect of urban residence on BMI was 0.3 for men and 0.5 for women, consistent with the findings that urban residents have higher BMI. However, after controlling for community unobserved heterogeneity by the community dummies, the effect of urban residence increased to 2.4 unit of BMI difference in men and 1.8 unit difference in women. For a 6-foot tall man the difference in weight between urban and rural areas, other things being equal, is 18 pounds. For a typical woman the difference in weight is 11 pounds. An urban area tends to have better quality health care infrastructure, water and sanitation conditions, as well as other unobserved characteristics. Therefore it is again clearly important to have community dummies.

The effect of the prices of rice and the price of fish were positive and significant for men and women before adding community dummies. The effects of the prices of eggs and pork for men and women were all negative and significant. Eggs and pork constitute of high calorie intakes, hence the increase in prices may have a substitution effect for other foods with lower calorie and result in lower BMI. Fish and rice, on the other hand, can be thought of as substitute for red meat and wheat flour. The increase in fish and rice prices will hence result in more consumption of less healthy foods and increased BMI.

The percentage of households in each community obtaining drinking water from more than five meters deep underground source was used as reference to three other categories of water sources, namely, from open well, from spring, river, lake, rain or snow, and from water factories. There were sign changes before and after controlling for community dummies, which indicates the estimates without could be biased. At this point we don't want to draw any conclusions of the sign of the water source

effects. The joint effects of water sources are very significant.

The percentage of households in each community with in-house with-flush toilets was used as reference to all other types of toilet facilities (Table 1.12). The percent of households with in-house no-flush toilets decreased over time (Table A.4) and since it is negatively correlated with BMI it indicates a increased BMI in such neighborhoods overtime. In a given year the higher the prevalence of household with such a facility the lower the BMI is in this community. For the other toilet facility types the negative effect is consistent with our expectation but the effects are not significant after controlling for community dummies. What's still significant after controlling for community dummies are percent of households with in-house toilet but no flush.

The percentage of households in a community with no excreta around dwelling area was used as reference to all other level of excreta. When the percent of households with very little or some excreta around dwelling area in a community increases the level of BMI decreases. The effects were strong for both men and women before adding community dummies.

Table 1.5: Determinants of Adult BMI in CHNS 1989-97: Overall

Variables	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Some primary ^a	0.042 (0.136)	0.019 (0.139)	0.007 (0.136)	-0.015 (0.139)	0.231 (0.128)*	0.219 (0.131)*	0.223 (0.128)*	0.210 (0.132)
Primary	0.044 (0.138)	0.037 (0.141)	-0.006 (0.137)	-0.012 (0.141)	0.465 (0.133)**	0.468 (0.139)**	0.458 (0.134)**	0.459 (0.139)**
Middle school	0.050 (0.139)	0.048 (0.143)	-0.008 (0.139)	-0.009 (0.143)	-0.002 (0.160)	0.014 (0.167)	-0.011 (0.161)	0.003 (0.167)
High school	0.189 (0.162)	0.212 (0.166)	0.138 (0.162)	0.161 (0.166)	-0.383 (0.197)*	-0.391 (0.207)*	-0.399 (0.201)**	-0.408 (0.212)*
Tech/College+	0.076 (0.197)	0.102 (0.202)	0.034 (0.197)	0.062 (0.202)	-0.679 (0.222)**	-0.630 (0.232)**	-0.709 (0.229)**	-0.661 (0.240)**
Log real prod asset 1 ^b	0.013 (0.018)	0.015 (0.019)	0.011 (0.018)	0.013 (0.019)	0.039 (0.023)*	0.052 (0.028)*	0.039 (0.023)*	0.052 (0.028)*
Log real prod asset 2 ^b	0.042 (0.018)**	0.038 (0.019)**	0.043 (0.018)**	0.039 (0.019)**	-0.005 (0.024)	-0.023 (0.028)	-0.005 (0.024)	-0.022 (0.028)
Land farmed	-0.001 (0.011)	-0.001 (0.010)	-0.001 (0.011)	-0.001 (0.010)	0.004 (0.009)	0.003 (0.009)	0.004 (0.009)	0.002 (0.009)
Married			0.417 (0.116)**	0.415 (0.119)**			-0.202 (0.237)	-0.194 (0.239)
Div'd Sept'd			-0.366 (0.364)	-0.336 (0.373)			-0.247 (0.451)	-0.161 (0.458)
Widowed			-0.129 (0.239)	-0.081 (0.244)			-0.489 (0.293)*	-0.493 (0.299)*

Table continues

Table 1.5 (cont'd)

	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Community dummies	Yes	No	Yes	No	Yes	No	Yes	No
Community*year dummies	No	Yes	No	Yes	No	Yes	No	Yes
No. of Observations	12760	12760	12760	12760	13652	13652	13652	13652
R-squared	0.166	0.206	0.167	0.208	0.137	0.172	0.138	0.172
White Test of homoscedasticity								
Chi-sq stat	72.07	217.10	72.54	218.11	6.60	148.10	7.05	154.82
P-value	0.0000	0.0000	0.0000	0.0000	0.0368	0.0000	0.0295	0.0000
P-value for testing coefficients equal to zero								
Education	0.8323	0.7229	0.8435	0.7403	0.0000	0.0000	0.0000	0.0000
Productive assets	0.0002	0.0011	0.0002	0.0015	0.0486	0.0877	0.0468	0.0862
Age dummies	0.0000	0.3110	0.0000	0.4365	0.0000	0.0037	0.0000	0.0068
Cohort dummies	0.0000	0.9477	0.0000	0.9376	0.0009	0.0749	0.0009	0.0743
Marital status			0.0001	0.0002			0.3764	0.3759
Community dummies	0.0000		0.0000		0.0000		0.0000	
Community*year dummies		0.0000		0.0000		0.0000		0.0000
P-value for testing equality of coefficients between men and women								
Education	0.0000	0.0000	0.0000	0.0000				
Productive assets	0.1775	0.2662	0.2690	0.1816				
Age dummies	0.0754	0.0755	0.0735	0.0712				
Cohort dummies	0.5137	0.1944	0.1347	0.4814				
Marital status			0.0566	0.0677				
Community dummies	0.0000		0.0000					
Community*year dummies		0.0000		0.0000				

Note: Also included in all regressions are age dummies and five-year-cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 and ** at 0.05 level.

a. The reference group is people with no formal education.

b. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

Table 1.6: Determinants of Adult BMI in CHNS 1989-97: Urban Areas

Variables	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Some primary	-0.091 (0.314)	-0.030 (0.333)	-0.136 (0.309)	-0.082 (0.329)	-0.019 (0.260)	-0.046 (0.267)	-0.064 (0.261)	-0.098 (0.268)
Primary	-0.209 (0.319)	-0.148 (0.337)	-0.266 (0.315)	-0.208 (0.333)	0.035 (0.283)	0.016 (0.287)	-0.037 (0.285)	-0.069 (0.289)
Middle school	-0.229 (0.311)	-0.139 (0.329)	-0.292 (0.307)	-0.205 (0.327)	0.097 (0.284)	0.084 (0.290)	0.038 (0.286)	0.013 (0.292)
High school	0.184 (0.334)	0.298 (0.350)	0.131 (0.329)	0.243 (0.346)	-0.252 (0.308)	-0.290 (0.315)	-0.324 (0.311)	-0.370 (0.317)
Tech/College+	-0.128 (0.343)	-0.023 (0.356)	-0.192 (0.340)	-0.087 (0.353)	-0.407 (0.324)	-0.391 (0.333)	-0.498 (0.328)	-0.496 (0.337)
Log real prod asset 1 ^a	0.028 (0.041)	0.033 (0.044)	0.026 (0.041)	0.030 (0.044)	-0.012 (0.038)	-0.004 (0.042)	-0.012 (0.037)	-0.004 (0.042)
Log real prod asset 2 ^a	0.026 (0.041)	0.007 (0.043)	0.030 (0.041)	0.010 (0.043)	-0.007 (0.036)	-0.018 (0.039)	-0.004 (0.036)	-0.014 (0.039)
Land farmed	-0.018 (0.071)	-0.039 (0.096)	-0.018 (0.071)	-0.040 (0.096)	-0.065 (0.059)	-0.066 (0.061)	-0.069 (0.062)	-0.069 (0.063)
Married			0.381 (0.213)*	0.411 (0.218)*			-0.413 (0.391)	-0.447 (0.361)
Div'd Sep'd			-0.384 (0.610)	-0.435 (0.644)			-0.471 (0.618)	-0.362 (0.614)
Widowed			-1.327 (0.431)**	-1.177 (0.463)**			-1.099 (0.467)**	-1.195 (0.454)**
Community dummies	Yes	No	Yes	No	Yes	No	Yes	No
Community*year dummies	No	Yes	No	Yes	No	Yes	No	Yes
No. of Observations	4087	4087	4087	4087	4493	4493	4493	4493
R-squared	0.182	0.223	0.186	0.226	0.199	0.228	0.201	0.230
White Test of homoscedasticity								
Chi-sq stat	43.13	126.88	45.32	129.67	55.32	155.56	63.18	170.36
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

Table 1.6 (cont'd)

	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
P-value for testing coefficients equal to zero								
Education	0.3522	0.3256	0.2990	0.2872	0.2814	0.3342	0.2472	0.2863
Productive assets	0.2194	0.4605	0.2036	0.4628	0.8219	0.7514	0.8616	0.8201
Age dummies	0.0000	0.5441	0.0000	0.3240	0.0000	0.0026	0.0000	0.0030
Cohort dummies	0.0001	0.7023	0.0000	0.6966	0.0163	0.0473	0.0206	0.0545
Marital status			0.0000	0.0003			0.0587	0.0342
Community dummies	0.0000		0.0000		0.0000	0.0000	0.0000	
Community*year dummies		0.0000		0.0000				0.0000
P-value for testing equality of coefficients between urban men and women								
Education	0.0939	0.0800	0.0840	0.0699				
Productive assets	0.1592	0.3878	0.1633	0.4225				
Age dummies	0.0945	0.1054	0.0598	0.0508				
Cohort dummies	0.4130	0.5014	0.4124	0.5248				
Marital status			0.0536	0.0553				
Community dummies	0.0126		0.0124					
Community*year dummies		0.0075		0.0045				
P-value for testing equality of coefficients between urban and rural men and between urban and rural women								
Education	0.6523	0.6566	0.5854	0.5978	0.2778	0.2668	0.2526	0.2362
Productive assets	0.9100	0.7107	0.9263	0.7537	0.0933	0.1352	0.1007	0.1498
Age dummies	0.1953	0.7583	0.0874	0.6078	0.1618	0.2441	0.1985	0.2715
Cohort dummies	0.1077	0.2957	0.1000	0.2993	0.1929	0.1643	0.2179	0.1870
Marital status			0.0045	0.0164			0.3127	0.2186

Note: Also included in all regressions are age dummies and five-year-cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 and ** at 0.05 level.

a. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

Table 1.7: Determinants of Adult BMI in CHNS 1989-97: Rural Areas

Variables	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Some primary	-0.051 (0.148)	-0.080 (0.148)	-0.065 (0.148)	-0.095 (0.149)	0.152 (0.150)	0.162 (0.154)	0.150 (0.151)	0.160 (0.155)
Primary	-0.075 (0.148)	-0.086 (0.149)	-0.097 (0.149)	-0.109 (0.150)	0.410 (0.150)**	0.441 (0.157)**	0.411 (0.150)**	0.442 (0.158)**
Middle school	-0.049 (0.152)	-0.070 (0.153)	-0.080 (0.152)	-0.102 (0.153)	-0.145 (0.197)	-0.121 (0.206)	-0.150 (0.200)	-0.125 (0.208)
High school	-0.001 (0.183)	-0.004 (0.187)	-0.035 (0.184)	-0.040 (0.187)	-0.303 (0.267)	-0.308 (0.286)	-0.322 (0.276)	-0.326 (0.297)
Tech/College+	0.128 (0.280)	0.137 (0.290)	0.124 (0.280)	0.133 (0.289)	-0.853 (0.352)**	-0.832 (0.369)**	-0.902 (0.373)**	-0.877 (0.391)**
Log real prod asset 1 ^a	0.009 (0.019)	0.011 (0.020)	0.008 (0.019)	0.009 (0.020)	0.059 (0.029)**	0.074 (0.035)**	0.059 (0.029)**	0.074 (0.035)**
Log real prod asset 2 ^a	0.043 (0.019)**	0.045 (0.020)**	0.043 (0.019)**	0.045 (0.020)**	-0.006 (0.031)	-0.028 (0.037)	-0.006 (0.031)	-0.027 (0.036)
Land farmed	-0.003 (0.011)	-0.003 (0.010)	-0.003 (0.011)	-0.002 (0.010)	0.003 (0.009)	0.002 (0.009)	0.002 (0.009)	0.001 (0.009)
Married			0.301 (0.134)**	0.307 (0.137)**			-0.286 (0.323)	-0.261 (0.339)
Divorced Separated			-0.641 (0.457)	-0.536 (0.464)			-0.485 (0.673)	-0.496 (0.650)
Widowed			0.237 (0.285)	0.258 (0.285)			-0.272 (0.392)	-0.237 (0.409)

Table continues

Table 1.7 (cont'd)

	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Community dummies	Yes	No	Yes	No	Yes	No	Yes	No
Community*year dummies	No	Yes	No	Yes	No	Yes	No	Yes
No. of Observations	8673	8673	8673	8673	9159	9159	9159	9159
R-squared	0.154	0.192	0.155	0.193	0.120	0.156	0.121	0.156
White Test of homoscedasticity								
Chi-sq stat	28.71	97.92	29.01	99.37	4.43	181.21	5.02	196.02
P-value	0.0000	0.0000	0.0000	0.0000	0.1091	0.0000	0.0814	0.0000
P-value for testing coefficients equal to zero								
Education	0.9676	0.9442	0.9495	0.9173	0.0006	0.0013	0.0012	0.0025
Productive assets	0.0015	0.0018	0.0018	0.0022	0.0063	0.0222	0.0062	0.0229
Age dummies	0.0000	0.1946	0.0000	0.2156	0.0000	0.4349	0.0000	0.4500
Cohort dummies	0.0000	0.5476	0.0000	0.5526	0.0154	0.2356	0.0155	0.2319
Marital status			0.0339	0.0535	0.0000		0.8361	0.8537
Community dummies	0.0000		0.0000				0.0000	
Community*year dummies		0.0000		0.0000		0.0000		0.0000
P-value for testing equality of coefficients between men and women								
Education	0.0024	0.0034	0.0032	0.0045				
Productive assets	0.3299	0.2157	0.3145	0.2100				
Age dummies	0.1547	0.3313	0.1617	0.3532				
Cohort dummies	0.2810	0.4290	0.2318	0.4114				
Marital status			0.2517	0.3877				
Community dummies	0.0011		0.0011					
Community*year dummies		0.0000		0.0000				

Note: Also included in all regressions are age dummies and five-year-cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 and ** at 0.05 level.

a. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

Table 1.8: Determinants of Adult BMI in CHNS 1989-97: Age 20 - 39

Variables	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Some primary	0.007 (0.284)	-0.035 (0.287)	-0.052 (0.284)	-0.101 (0.288)	-0.090 (0.196)	-0.023 (0.199)	-0.092 (0.197)	-0.025 (0.199)
Primary	-0.087 (0.280)	-0.114 (0.283)	-0.155 (0.281)	-0.188 (0.284)	0.013 (0.204)	0.108 (0.216)	0.014 (0.204)	0.107 (0.215)
Middle school	-0.126 (0.276)	-0.103 (0.276)	-0.199 (0.278)	-0.183 (0.278)	-0.408 (0.226)*	-0.305 (0.230)	-0.415 (0.228)*	-0.312 (0.233)
High school	0.108 (0.292)	0.142 (0.295)	0.038 (0.293)	0.065 (0.296)	-0.752 (0.254)**	-0.713 (0.272)**	-0.772 (0.264)**	-0.731 (0.283)**
Tech/College+	-0.019 (0.335)	0.019 (0.338)	-0.072 (0.335)	-0.037 (0.339)	-1.016 (0.280)**	-0.979 (0.293)**	-1.053 (0.297)**	-1.014 (0.310)**
Log real prod asset 1 ^a	0.003 (0.022)	-0.002 (0.025)	0.001 (0.022)	-0.004 (0.025)	0.036 (0.036)	0.038 (0.047)	0.036 (0.036)	0.039 (0.047)
Log real prod asset 2 ^a	0.054 (0.023)**	0.060 (0.026)**	0.055 (0.023)**	0.060 (0.026)**	-0.041 (0.037)	-0.053 (0.041)	-0.040 (0.036)	-0.052 (0.041)
Land farmed	-0.016 (0.009)*	-0.013 (0.010)	-0.015 (0.009)*	-0.012 (0.010)	0.022 (0.011)**	0.019 (0.011)*	0.020 (0.011)*	0.018 (0.011)
Married			0.308 (0.122)**	0.327 (0.126)**			-0.272 (0.285)	-0.241 (0.273)
Divorced Separated			-0.290 (0.385)	-0.321 (0.414)			-0.396 (0.568)	-0.191 (0.597)
Widowed			-0.478 (0.434)	-0.373 (0.485)			0.298 (0.947)	0.194 (1.008)

Table continues

Table 1.8 (cont'd)

	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Community dummies	Yes	No	Yes	No	Yes	No	Yes	No
Community*year dummies	No	Yes	No	Yes	No	Yes	No	Yes
No. of Observations	6518	6518	6518	6518	7007	7007	7007	7007
R-squared	0.177	0.248	0.178	0.249	0.104	0.175	0.104	0.176
White Test of homoscedasticity								
Chi-sq stat	134.38	464.96	133.50	464.14	79.23	967.75	91.00	996.39
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-value for testing coefficients equal to zero								
Education	0.5133	0.5085	0.4574	0.4638	0.0003	0.0008	0.0008	0.0017
Productive assets	0.0034	0.0052	0.0047	0.0076	0.5360	0.3412	0.5401	0.3542
Age dummies	0.0000	0.3590	0.0000	0.3333	0.0000	0.0140	0.0000	0.0132
Cohort dummies	0.0000	0.9843	0.0000	0.9872	0.0003	0.0128	0.0003	0.0122
Marital status			0.0095	0.0137			0.7134	0.8157
Community dummies	0.0000		0.0000		0.0000		0.0000	
Community*year dummies		0.0000		0.0000		0.0000		0.0000
P-value for testing equality of coefficients between men and women								
Education	0.0017	0.0013	0.0024	0.0019				
Productive assets	0.0195	0.0051	0.0228	0.0068				
Age dummies	0.3029	0.2633	0.3442	0.2747				
Cohort dummies	0.1694	0.4547	0.1092	0.4167				
Marital status			0.1312	0.1540				
Community dummies	0.0002		0.0001					
Community*year dummies		0.0000		0.0000				

Note: Also included in all regressions are age dummies and five-year-cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 and ** at 0.05 level.

a. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

Table 1.9: Determinants of Adult BMI in CHNS 1989-97: Age 40 - 59

Variables	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Some primary	-0.084 (0.195)	-0.020 (0.214)	-0.143 (0.195)	-0.080 (0.213)	0.315 (0.183)*	0.331 (0.203)	0.309 (0.184)*	0.323 (0.204)
Primary	-0.154 (0.189)	-0.132 (0.215)	-0.213 (0.188)	-0.190 (0.213)	0.502 (0.197)**	0.475 (0.219)**	0.495 (0.198)**	0.467 (0.220)**
Middle school	-0.060 (0.199)	-0.033 (0.222)	-0.134 (0.198)	-0.108 (0.220)	0.221 (0.264)	0.206 (0.274)	0.211 (0.265)	0.191 (0.275)
High school	-0.025 (0.269)	0.018 (0.299)	-0.097 (0.266)	-0.052 (0.297)	0.646 (0.422)	0.587 (0.461)	0.641 (0.422)	0.581 (0.462)
Tech/College+	0.036 (0.316)	0.161 (0.341)	-0.058 (0.315)	0.063 (0.340)	-0.154 (0.400)	0.046 (0.436)	-0.160 (0.400)	0.032 (0.437)
Log real prod asset 1 ^a	0.001 (0.030)	-0.006 (0.035)	-0.001 (0.030)	-0.008 (0.035)	0.030 (0.034)	0.022 (0.043)	0.031 (0.034)	0.021 (0.043)
Log real prod asset 2 ^a	0.066 (0.027)**	0.062 (0.031)**	0.067 (0.027)**	0.061 (0.031)**	0.077 (0.037)**	0.071 (0.046)	0.076 (0.037)**	0.071 (0.046)
Land farmed	0.018 (0.025)	0.010 (0.019)	0.016 (0.025)	0.007 (0.019)	-0.008 (0.015)	-0.008 (0.017)	-0.010 (0.015)	-0.010 (0.017)
Married			0.984 (0.303)**	1.033 (0.376)**			-0.315 (0.381)	-0.479 (0.414)
Divorced Separated			0.747 (0.735)	0.938 (0.786)			-0.049 (0.669)	-0.227 (0.797)
Widowed			0.325 (0.499)	0.345 (0.563)			-0.618 (0.489)	-0.890 (0.535)*

Table continues

Table 1.9 (cont'd)

	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Community dummies	Yes	No	Yes	No	Yes	No	Yes	No
Community*year dummies	No	Yes	No	Yes	No	Yes	No	Yes
No. of Observations	4388	4388	4388	4388	4696	4696	4696	4696
R-squared	0.206	0.281	0.209	0.284	0.219	0.292	0.219	0.293
White Test of homoscedasticity								
Chi-sq stat	16.28	128.12	16.44	130.64	36.48	202.62	36.03	202.55
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-value for testing coefficients equal to zero								
Education	0.9466	0.9266	0.9085	0.9227	0.0741	0.2724	0.0787	0.2839
Productive assets	0.0032	0.0391	0.0039	0.0502	0.0011	0.0209	0.0012	0.0218
Age dummies	0.0000	0.4636	0.0000	0.4363	0.0000	0.0015	0.0000	0.0021
Cohort dummies	0.0000	0.0976	0.0000	0.0968	0.1374	0.0238	0.1434	0.0248
Marital status			0.0069	0.0232			0.5909	0.3987
Community dummies	0.0000		0.0000		0.0000		0.0000	
Community*year dummies		0.0000		0.0000		0.0000		0.0000
P-value for testing equality of coefficients between men and women								
Education	0.1175	0.3308	0.0907	0.2833				
Productive assets	0.5480	0.6857	0.5061	0.6440				
Age dummies	0.0277	0.0109	0.0245	0.0102				
Cohort dummies	0.0021	0.0270	0.0015	0.0186				
Marital status			0.0582	0.0583				
Community dummies	0.0000		0.0000					
Community*year dummies		0.0000		0.0000				

Note: Also included in all regressions are age dummies and five-year-cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 and ** at 0.05 level.

a. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

Table 1.10: Determinants of Adult BMI in CHNS 1989-97: Age 60+

Variables	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Some primary	-0.253 (0.250)	-0.301 (0.304)	-0.256 (0.250)	-0.296 (0.303)	-0.113 (0.336)	-0.022 (0.360)	-0.123 (0.336)	-0.040 (0.360)
Primary	-0.241 (0.317)	-0.181 (0.396)	-0.253 (0.318)	-0.179 (0.396)	0.048 (0.451)	-0.370 (0.448)	0.033 (0.450)	-0.400 (0.449)
Middle school	-0.605 (0.382)	-0.389 (0.407)	-0.630 (0.386)	-0.395 (0.411)	-0.216 (0.625)	-0.311 (0.726)	-0.219 (0.632)	-0.322 (0.736)
High school	0.065 (0.579)	0.035 (0.667)	0.090 (0.578)	0.075 (0.665)	0.948 (0.925)	0.919 (1.093)	0.988 (0.920)	0.972 (1.081)
Tech/College+	-0.340 (0.498)	-0.545 (0.564)	-0.357 (0.499)	-0.551 (0.563)	0.743 (0.826)	0.602 (0.945)	0.680 (0.827)	0.520 (0.948)
Log real prod asset 1 ^a	0.081 (0.072)	0.080 (0.092)	0.079 (0.072)	0.079 (0.092)	0.123 (0.062)**	0.100 (0.066)	0.123 (0.062)**	0.100 (0.066)
Log real prod asset 2 ^a	-0.040 (0.067)	-0.064 (0.084)	-0.033 (0.067)	-0.058 (0.084)	-0.047 (0.061)	-0.032 (0.070)	-0.047 (0.062)	-0.032 (0.071)
Land farmed	-0.035 (0.022)	-0.045 (0.027)*	-0.033 (0.022)	-0.043 (0.027)	0.001 (0.031)	0.028 (0.032)	0.001 (0.031)	0.028 (0.033)
Married			-0.174 (0.588)	-0.608 (0.610)			-0.360 (0.568)	-0.472 (0.746)
Divorced Separated			-1.772 (1.052)*	-1.959 (1.328)			-0.946 (1.053)	-1.129 (1.307)
Widowed			-0.385 (0.652)	-0.796 (0.692)			-0.613 (0.585)	-0.779 (0.748)

Table continues

Table 1.10 (cont'd)

	Male: OLS				Female: OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Community dummies	Yes	No	Yes	No	Yes	No	Yes	No
Community*year dummies	No	Yes	No	Yes	No	Yes	No	Yes
No. of Observations	1854	1854	1854	1854	1949	1949	1949	1949
R-squared	0.318	0.410	0.320	0.412	0.310	0.461	0.311	0.462
White Test of homoscedasticity								
Chi-sq stat	40.01	290.69	40.48	291.87	69.59	99.62	70.12	100.05
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-value for testing coefficients equal to zero								
Education	0.6820	0.8478	0.6417	0.8358	0.8333	0.8362	0.8317	0.8148
Productive assets	0.4770	0.6805	0.4663	0.6891	0.0901	0.2370	0.0871	0.2351
Age dummies	0.2854	0.5786	0.2439	0.5718	0.6164	0.6664	0.6292	0.6527
Cohort dummies	0.0029	0.8936	0.0035	0.9082	0.3240	0.5504	0.3541	0.5394
Marital status			0.2845	0.4900			0.5864	0.5247
Community dummies	0.0000		0.0000		0.0000		0.0000	
Community*year dummies		0.0000		0.0000		0.0000		0.0000
P-value for testing equality of coefficients between men and women								
Education	0.8955	0.8607	0.9034	0.8712				
Productive assets	0.8419	0.7940	0.8528	0.8308				
Age dummies	0.0146	0.3236	0.0183	0.3729				
Cohort dummies	0.5578	0.9328	0.5177	0.9258				
Marital status			0.8590	0.9577				
Community dummies	0.0000		0.0000					
Community*year dummies		0.0000		0.0000				

Note: Also included in all regressions are age dummies and five-year-cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 and ** at 0.05 level.

a. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

Table 1.11: Adult BMI by CHNS Round 1991, 93, and 97: Overall

Variables	Male: OLS			Female: OLS		
	1991	1993	1997	1991	1993	1997
Some primary ^a	-0.065 (0.206)	0.208 (0.220)	-0.095 (0.269)	0.232 (0.248)	0.380 (0.190)**	0.015 (0.231)
Primary	-0.087 (0.206)	-0.145 (0.235)	0.224 (0.269)	0.721 (0.216)**	0.561 (0.254)**	0.186 (0.234)
Middle school	0.013 (0.211)	-0.075 (0.222)	0.223 (0.274)	-0.215 (0.371)	0.223 (0.231)	-0.111 (0.273)
High school	0.068 (0.240)	0.290 (0.272)	0.383 (0.306)	-0.560 (0.443)	-0.246 (0.266)	-0.558 (0.383)
Tech/College+	0.076 (0.305)	-0.178 (0.323)	0.330 (0.343)	-1.146 (0.482)**	-0.514 (0.411)	-0.604 (0.363)*
Log real prod asset 1 ^b	-0.002 (0.041)	0.019 (0.038)	0.020 (0.029)	0.122 (0.094)	0.041 (0.041)	0.041 (0.036)
Log real prod asset 2 ^b	0.044 (0.043)	0.046 (0.033)	0.039 (0.032)	-0.122 (0.087)	-0.011 (0.038)	0.019 (0.040)
Land farmed	-0.008 (0.022)	0.039 (0.025)	-0.025 (0.011)**	0.008 (0.027)	0.009 (0.015)	-0.003 (0.013)
Married	0.327 (0.169)*	0.379 (0.221)*	0.645 (0.228)**	-0.470 (0.671)	-0.020 (0.289)	-0.034 (0.336)
Divorced Separated	-0.258 (0.683)	-1.006 (0.872)	-0.489 (0.510)	-0.202 (0.821)	-0.032 (0.665)	-0.240 (0.973)
Widowed	-0.156 (0.334)	-0.161 (0.394)	0.268 (0.433)	-0.479 (0.697)	-0.338 (0.422)	-0.655 (0.398)*

Table continues

Table 1.11 (cont'd)

	Male: OLS		Female: OLS	
	1991	1993	1991	1993
Community dummies	Yes	Yes	Yes	Yes
No. of Observations	3537	3329	3698	3882
R-squared	0.220	0.176	0.217	0.197
P-value for testing coefficients equal to zero				
Education	0.9576	0.2574	0.2794	0.0018
Productive assets	0.2946	0.0511	0.0556	0.5053
Age dummies	0.0000	0.0006	0.0000	0.0000
Marital status	0.0725	0.0362	0.0030	0.7807
Community dummies	0.0000	0.0000	0.0000	0.0000
P-value for testing equality of coefficients between men and women				
Education	0.0055	0.0094	0.0837	
Productive assets	0.1577	0.4496	0.8983	
Age dummies	0.2489	0.0544	0.6108	
Marital status	0.3889	0.3746	0.2645	
Community dummies	0.0279	0.0003	0.0014	

Note: Also included in all regressions are age dummies. Heteroscedastic robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 and ** at 0.05 level.

a. The reference group is people with no formal education.

b. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

Table 1.12: Determinants of Adult BMI in CHNS 1989-93 with Community Characteristics: Overall

Variables	Male OLS				Female OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Some primary	-0.078 (0.155)	0.093 (0.149)	-0.121 (0.154)	0.062 (0.148)	0.241 (0.138)*	0.328 (0.146)**	0.233 (0.139)*	0.324 (0.147)**
Primary	-0.118 (0.152)	-0.027 (0.148)	-0.181 (0.151)	-0.068 (0.147)	0.585 (0.165)**	0.571 (0.153)**	0.579 (0.166)**	0.569 (0.153)**
Middle school	-0.017 (0.155)	-0.006 (0.150)	-0.089 (0.155)	-0.051 (0.150)	0.148 (0.154)	0.039 (0.189)	0.139 (0.155)	0.034 (0.190)
High school	0.126 (0.182)	0.131 (0.176)	0.064 (0.181)	0.091 (0.175)	-0.220 (0.183)	-0.321 (0.222)	-0.235 (0.186)	-0.336 (0.230)
Tech/College+	0.080 (0.218)	0.047 (0.218)	0.026 (0.217)	0.006 (0.217)	-0.519 (0.249)**	-0.744 (0.276)**	-0.545 (0.254)**	-0.772 (0.288)**
Log real prod asset 1 ^a	0.032 (0.022)	0.007 (0.024)	0.030 (0.022)	0.004 (0.024)	0.080 (0.036)**	0.038 (0.036)	0.080 (0.036)**	0.038 (0.036)
Log real prod asset 2 ^a	0.020 (0.023)	0.040 (0.023)*	0.021 (0.023)	0.042 (0.023)*	-0.034 (0.034)	-0.022 (0.035)	-0.034 (0.034)	-0.021 (0.035)
Land farmed	-0.005 (0.016)	0.022 (0.019)	-0.005 (0.016)	0.022 (0.019)	-0.001 (0.012)	0.014 (0.012)	-0.001 (0.012)	0.013 (0.012)
Married			0.485 (0.125)**	0.291 (0.123)**			-0.171 (0.294)	-0.247 (0.305)
Divorced Separated			-0.622 (0.506)	-0.357 (0.501)			-0.052 (0.546)	-0.137 (0.539)
Widowed			-0.285 (0.278)	-0.301 (0.272)			-0.550 (0.398)	-0.372 (0.386)

Table continues

Table 1.12 (cont'd)

	Male OLS				Female OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Urban residence	0.303 (0.141)**	2.359 (0.864)**	0.318 (0.140)**	2.419 (0.867)**	0.483** (0.144)**	1.782 (0.981)*	0.475 (0.144)**	1.797 (0.983)*
Price of rice	0.277 (0.118)**	0.057 (0.121)	0.267 (0.117)**	0.056 (0.120)	0.354 (0.146)**	-0.009 (0.152)	0.353 (0.146)**	-0.005 (0.152)
Price of eggs	-0.070 (0.015)**	-0.009 (0.016)	-0.068 (0.015)**	-0.009 (0.016)	-0.072 (0.018)**	-0.008 (0.018)	-0.071 (0.018)**	-0.008 (0.018)
Price of pork	-0.111 (0.023)**	-0.013 (0.025)	-0.108 (0.023)**	-0.012 (0.025)	-0.155 (0.025)**	-0.063 (0.032)*	-0.154 (0.025)**	-0.063 (0.032)*
Price of fish	0.042 (0.016)**	0.004 (0.019)	0.043 (0.016)**	0.006 (0.019)	0.051 (0.017)**	-0.009 (0.022)	0.050 (0.018)**	-0.008 (0.022)
Clinic # physicians	-18.901 (14.177)	-4.517 (14.438)	-18.123 (14.194)	-4.143 (14.450)	0.235 (15.891)	1.765 (17.110)	0.283 (15.878)	1.728 (17.086)
Clinic # nurses	-27.062 (61.809)	34.998 (59.647)	-25.679 (61.899)	36.456 (59.649)	12.505 (68.754)	31.549 (72.975)	12.885 (68.587)	30.630 (72.931)
Clinic # beds	11.920 (19.882)	-3.652 (18.836)	11.328 (19.925)	-4.436 (18.834)	-5.720 (21.787)	-14.285 (22.733)	-5.842 (21.739)	-14.036 (22.726)
Water Source 2 ^c	-0.697 (0.141)**	0.257 (0.216)	-0.704 (0.141)**	0.267 (0.215)	-0.966 (0.173)**	0.045 (0.205)	-0.968 (0.173)**	0.042 (0.205)
Water Source 3	-0.683 (0.160)**	1.062 (0.246)**	-0.680 (0.160)**	1.060 (0.247)**	-1.058 (0.214)**	0.316 (0.272)	-1.056 (0.216)**	0.309 (0.273)
Water Source 4	-0.504 (0.131)**	0.727 (0.214)**	-0.516 (0.131)**	0.721 (0.214)**	-0.630 (0.140)**	0.174 (0.228)	-0.628 (0.141)**	0.172 (0.228)

Table continues

Table 1.12 (cont'd)

	Male OLS				Female OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
In house no flush	-0.993 (0.293)**	-1.351 (0.683)**	-0.981 (0.291)**	-1.351 (0.682)**	-0.380 (0.303)	-0.940 (0.589)	-0.373 (0.304)	-0.928 (0.594)
Outside toilets	-0.332 (0.210)	-0.474 (0.514)	-0.327 (0.209)	-0.449 (0.515)	0.047 (0.213)	-0.091 (0.485)	0.046 (0.213)	-0.081 (0.489)
Open pit cement or earth	-0.758 (0.217)**	-0.116 (0.513)	-0.770 (0.217)**	-0.092 (0.514)	-0.390 (0.212)*	-0.234 (0.497)	-0.385 (0.212)*	-0.223 (0.500)
No toilets	-2.164 (0.634)**	-0.619 (0.795)	-2.133 (0.633)**	-0.581 (0.794)	-1.920 (0.674)**	-0.434 (0.758)	-1.884 (0.673)**	-0.441 (0.756)
Very little excreta	-0.603 (0.168)**	-0.092 (0.179)	-0.601 (0.168)**	-0.094 (0.179)	-0.517 (0.242)**	-0.237 (0.281)	-0.516 (0.241)**	-0.242 (0.282)
Some excreta	-0.650 (0.165)**	0.250 (0.173)	-0.636 (0.165)**	0.238 (0.173)	-0.921 (0.206)**	0.196 (0.203)	-0.919 (0.205)**	0.197 (0.203)
Much excreta	0.472 (0.567)	-0.035 (0.771)	0.506 (0.571)	-0.069 (0.771)	1.046 (0.727)	0.381 (0.800)	1.028 (0.729)	0.384 (0.800)
Community Dummies	No	Yes	No	Yes	No	Yes	No	Yes
Number of Obs	8806	8806	8806	8806	9482	9482	9482	9482
R-squared	0.074	0.166	0.077	0.167	0.075	0.135	0.075	0.135
White Test of homoscedasticity								
Chi-sq stat	54.11	84.70	53.01	83.56	2.52	7.19	2.71	7.81
P-value	0.0000	0.0000	0.0000	0.0000	0.2835	0.0275	0.2577	0.0201

Table continues

Table 1.12 (cont'd)

	Male OLS				Female OLS			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
P-value for testing coefficients equal to zero								
Education	0.5534	0.7650	0.4801	0.7329	0.0000	0.0000	0.0000	0.0000
Prod Assets	0.0138	0.0171	0.0178	0.0187	0.0217	0.5128	0.0233	0.5091
Age dummies	0.0009	0.0030	0.0064	0.0080	0.0000	0.0007	0.0000	0.0007
Cohort dummies	0.0229	0.1130	0.0132	0.0920	0.1439	0.1683	0.1464	0.1671
Marital Status			0.0000	0.0062			0.4764	0.7920
Prices	0.0000	0.8017	0.0000	0.7890	0.0000	0.4425	0.0000	0.4431
Water source	0.0000	0.0000	0.0000	0.0000	0.0000	0.6968	0.0000	0.7109
Toilet type	0.0002	0.0687	0.0002	0.0692	0.0090	0.2697	0.0101	0.2900
Excreta	0.0000	0.3723	0.0001	0.4070	0.0000	0.4206	0.0000	0.4182
Clinic characteristics	0.5234	0.5851	0.5553	0.6071	0.8999	0.8638	0.8982	0.8677
Community dummies		0.0000		0.0000		0.0000		0.0000
P-value for testing equality of coefficients between men and women								
Education	0.0000	0.0000	0.0000	0.0001				
Prod Assets	0.2771	0.2329	0.2645	0.2304				
Age dummies	0.0000	0.0281	0.0000	0.0167				
Cohort dummies	0.0391	0.0490	0.0313	0.0406				
Marital Status			0.0354	0.1416				
Prices	0.0689	0.3276	0.0445	0.2848				
Water source	0.4204	0.9934	0.2613	0.9057				
Toilet type	0.1297	0.0107	0.1706	0.0068				
Excreta	0.9391	0.9895	0.9274	0.9881				
Community dummies		0.0015		0.0019				

Note: Also included in the models are age dummies and five-year-cohort dummies. Model (1) and (3) do not control for community dummies; model (2) and (4) do. Person level cluster robust standard errors are in parentheses. * indicates the coefficients are statistically significant at 0.1 level and ** at 0.05 level.

a. The real productive assets splines are round the median of positive productive assets values deflated to 1988. Log real prod asset 1 is the spline less than the median and 2 is the spline greater than or equal to the median.

c. The percent of households with the following water sources: (1) underground, (2) open well, (3) spring, river lake, rain or snow, and (4) water factory. Omitted is the first source.

1.6 Dynamic Conditional BMI Demand

The estimation of equation (1.9), the dynamic conditional BMI demand function as outlined earlier, has both advantages and disadvantages. On the one hand we do not need to include all past variables based on the assumption that lagged BMI is a summary statistic for all past information. On the other hand, the estimates from such relationship are only partial effects of the variables we are interested in.

We study the determinants of BMI in 1993 conditional on BMI in 1991 and the determinants of BMI in 1997 conditional on BMI in 1993 controlling for age dummies, and community dummies when community characteristics are not added explicitly.²⁵ The lagged BMI is treated as endogenous and the model is estimated by OLS, 2SLS and GMM methods. In 2SLS and GMM estimations the endogenous BMI is instrumented with previous year's log of real productive assets splines, land cultivated, as well as interactions of real productive assets and land with community free market prices of rice, eggs, pork and fish, and water, sanitation and clinic characteristics in the previous year. The first stage regressions are included in appendix B. There is strong evidence of heteroscedasticity in the dynamic models. Under such condition the 2SLS is consistent and asymptotically correct inference can be made using the Huber-White "sandwich" robust variance-covariance matrix. However, the GMM estimator is more efficient, although with many overidentifying restrictions the small sample property of the GMM estimator can be poor (p.204 Wooldridge 2002). In particular, Wald tests tend to over-reject the null hypotheses (Baum et al. 2003). The GMM and OLS estimators are included in the paper for comparisons. The specification of interaction between lagged BMI with age groups is also tried and no particular patterns were found.

²⁵The estimation for BMI in 1997 conditional on BMI in 1991 was also carried out but not reported here because in 1997 some community characters are not available.

1.6.1 With Community Dummies

In Table 1.13 the regressions are for the dependent variable adult BMIs in 1993 conditional on BMIs in 1991 and education levels and household resources as in the baseline model. The IVs are the above stated variables in 1989. We conduct several specification tests. The first stage F tests for the identifying instruments and the overidentification tests²⁶ suggested the IVs were valid instruments (see appendix B for the first stage regressions). All first stage F tests are very significant, signaling the correlation between the identifying IVs and the lagged BMI is not weak. The overidentification χ_2 tests are heteroscedasticity-robust version based on Wooldridge (2002 p. 123). Since each year the question on land cultivated is based on last year's value, here we chose to exclude it for the overidentification test. All tests can not reject the null hypothesis that all other instruments are uncorrelated with the error terms. The Hausman test between the OLS and the 2SLS estimates of the lagged BMI for men suggested a significant difference at 10 percent level but cannot confirm such difference for women.²⁷ All three methods give positive and significant effects of the lagged BMI on the current BMI. The estimated effect of the BMI in 1991 on the BMI in 1993 in OLS is around 0.6 and in 2SLS and GMM about 0.8 for men, and around 0.6 in OLS and around 0.7 in 2SLS and GMM for women.

Controlling for lagged BMI the effects of education and household resources in 1991 were no longer significant in both OLS and 2SLS estimates, except that in GMM estimation the primary diploma in 1991 showed a strong negative effect for men. The male marriage premium in BMI become insignificant.

The conditional age and community effects are significant for women and the age

²⁶The choices of instrumental variables for the overidentification test is arbitrary. The land cultivated is assumed to be uncorrelated with the error terms and the overidentification test is based on the null hypothesis that all other identifying IVs are uncorrelated with the error terms.

²⁷The Hausman test between OLS and 2SLS estimate of the lagged BMI effect was based on the assumption of homoscedasticity, which was clearly violated in the model. It is not surprising that the test has such low power.

effect conditional on past BMI for men is not jointly significant.

In Table 1.14 the BMI in 1997 conditional on the BMI in 1993 is estimated. The lagged BMI is instrumented by the same set of variables from 1991.²⁸ The first stage F tests for the identifying instruments showed signs of weak instruments. The overidentification tests passed. The Hausman test between the OLS and 2SLS estimates of the lagged BMI cannot reject the null hypothesis that they were the same. Both 2SLS and GMM estimates were higher than the OLS estimate for men, but it's not the case for women.

The partial effects of some primary education for men were negative and significant in all estimates. The strong nonlinear effect of productive assets between the higher and lower levels is much larger than those estimated in the reduced form regressions. The effects of productive assets are of opposite signs than those in Table 1.10 for both men and women in the lower half of the assets spectrum. The marriage premiums for men no longer exist.

The fact that we can reject BMI being exogenous in the first conditional regression (BMI 93 on BMI 91) but not the second (BMI 97 on BMI 93) suggests that although BMI is a cumulative measure of health, values from a farther past does not serve as a summary statistic as well as values from a more recent past. The conditional dynamic model is hinged on the assumption that lagged BMI is a sufficient statistic.

1.6.2 With Community Characteristics

In Table 1.15 and 1.16, instead of controlling for community level information with dummy variables we use changes in community characteristics between current and lagged period to examine the effects explicitly. Similar to the dynamic models

²⁸I have also tried to instrument the BMI in 1993 with log of the real productive assets and land in 1989 and interactions of these variables with community free market prices of rice, wheat, eggs, pork, beef and fish, as well as water, sanitation and clinic characteristics in 1989, however, there is strong evidence of the weak IV problem.

with community dummies, the first stage F-tests and the overidentification tests all validate the instruments. The Hausman test for BMI 91 in the conditional regression for BMI 93 rejects the null hypothesis that BMI 91 is exogenous; whereas in the conditional model of BMI 97 on BMI 93 we can not reject the OLS and 2SLS are the same.

Oddly, the effects of changes in prices of rice, eggs and pork in two conditional models for men and women are of exact opposite signs. In our sample all these price changes over time are positive. With an increasing in the price of rice, for example, we would expect to see BMI to increase if substitution effect is stronger than income effect. However, in our sample the increase in wheat price is actually bigger than the increase in rice price from 1991 to 93. Therefore there is no strong substitution effect and the coefficient for rice price changes are negative.

Also significant are some water and toilet type characters in 2SLS results and again of opposite signs in some cases. For women only the change in in-house no-flush toilet percent and no toilet percent changes is negatively associated with BMI. Overall the age dummies and water source dummies are significant for men in table 15. Although the conditional BMI specification works better with shorter lagged period, the changes in community level information during such short period may not be large enough to correctly identify any effect.

Table 1.13: Determinants of Adults BMI in 1993 Conditional on BMI in 1991: Overall

Variables	Male					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 91	0.573 (0.061)**	0.843 (0.191)**	0.803 (0.088)**	0.571 (0.061)**	0.837 (0.192)**	0.802 (0.087)**
Some primary 91	-0.100 (0.218)	-0.091 (0.226)	-0.355 (0.227)	-0.134 (0.216)	-0.117 (0.225)	-0.404 (0.228)*
Primary 91	-0.225 (0.226)	-0.204 (0.229)	-0.495 (0.220)**	-0.255 (0.225)	-0.228 (0.229)	-0.534 (0.221)**
Middle school 91	-0.252 (0.215)	-0.267 (0.221)	-0.384 (0.239)	-0.293 (0.216)	-0.299 (0.222)	-0.443 (0.242)*
High school 91	-0.083 (0.253)	-0.161 (0.267)	-0.256 (0.281)	-0.121 (0.251)	-0.189 (0.263)	-0.315 (0.283)
Tech/College+ 91	0.060 (0.294)	0.065 (0.297)	-0.343 (0.466)	0.025 (0.292)	0.037 (0.295)	-0.419 (0.474)
Log R prod assets 91 1	0.006 (0.037)	0.015 (0.037)	0.046 (0.029)	0.002 (0.037)	0.012 (0.037)	0.041 (0.029)
Log R prod assets 91 2	0.044 (0.042)	0.026 (0.041)	-0.021 (0.032)	0.045 (0.042)	0.027 (0.041)	-0.020 (0.031)
Land 91	0.005 (0.025)	-0.001 (0.026)	-0.016 (0.019)	0.003 (0.025)	-0.003 (0.026)	-0.018 (0.019)
Married 91				0.376 (0.242)	0.294 (0.266)	0.317 (0.199)
Divorced Separated 91				0.108 (0.626)	0.164 (0.618)	-0.065 (0.583)
Widowed 91				-0.203 (0.408)	-0.191 (0.423)	-0.270 (0.372)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	2477	2477	2477	2477	2477	2477
R-squared	0.370	0.330	0.237	0.370	0.332	0.226
First stage for BMI 91						
R-squared		0.258			0.259	
F-statistic for Identifying IVs		2.82			2.70	
P-value		0.0000			0.0000	
Test of homoscedasticity						
Chi-sq stat	118.22	89.44		116.86	88.17	
P-value	0.0000	0.0000		0.0000	0.0000	
Hausman Test between OLS and 2SLS: BMI 91						
Chi-square statistics		3.18			3.07	
P-value		0.0747			0.0798	
Overidentification test						
Chi-square statistics dof		Chi(56)			Chi(56)	
Chi-square statistics		56.52			56.96	
P-value		0.4555			0.4390	
P-value for testing coefficients equal to zero						
Education	0.7620	0.7608	0.2342	0.7064	0.7172	0.1882
Assets	0.3171	0.4992	0.2415	0.3565	0.5390	0.3186
Age dummies	0.4321	0.4164	0.3632	0.5246	0.5394	0.4484
Marital Status				0.1890	0.4746	0.1266
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

Table 1.13 (cont'd)

Variables	Female					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 91	0.552 (0.041)**	0.644 (0.106)**	0.657 (0.066)**	0.552 (0.041)**	0.641 (0.106)**	0.650 (0.066)**
Some primary 91	0.189 (0.181)	0.163 (0.185)	0.214 (0.172)	0.188 (0.181)	0.163 (0.185)	0.214 (0.172)
Primary 91	0.276 (0.199)	0.222 (0.210)	0.234 (0.187)	0.271 (0.199)	0.218 (0.210)	0.232 (0.186)
Middle school 91	0.051 (0.205)	0.044 (0.205)	0.064 (0.181)	0.047 (0.205)	0.040 (0.205)	0.057 (0.181)
High school 91	-0.306 (0.256)	-0.306 (0.260)	-0.158 (0.401)	-0.305 (0.256)	-0.306 (0.259)	-0.171 (0.402)
Tech/College+ 91	0.343 (0.462)	0.432 (0.477)	0.510 (0.850)	0.358 (0.469)	0.443 (0.483)	0.487 (0.854)
Log R prod assets 91 1	0.031 (0.039)	0.026 (0.040)	0.030 (0.032)	0.031 (0.039)	0.026 (0.040)	0.030 (0.032)
Log R prod assets 91 2	-0.019 (0.037)	-0.014 (0.037)	-0.014 (0.031)	-0.018 (0.037)	-0.013 (0.037)	-0.013 (0.031)
Land 91	0.014 (0.030)	0.015 (0.030)	-0.020 (0.025)	0.013 (0.030)	0.014 (0.030)	-0.021 (0.025)
Married 91				0.153 (0.410)	0.139 (0.422)	0.152 (0.359)
Divorced Separated 91				0.685 (0.638)	0.644 (0.627)	0.681 (0.565)
Widowed 91				-0.146 (0.482)	-0.162 (0.489)	-0.125 (0.415)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	2701	2701	2701	2701	2701	2701
R-squared	0.442	0.435	0.403	0.442	0.436	0.409
First stage for BMI 91						
R-squared		0.238			0.238	
F-statistic for Identifying IVs		3.25			3.35	
P-value		0.0000			0.0000	
Test of homoscedasticity						
Chi-sq stat	113.30	110.84		113.71	110.62	
P-value	0.0000	0.0000		0.0000	0.0000	
Hausman Test between OLS and 2SLS: BMI 91						
Chi-square statistics		0.69			0.65	
P-value		0.4052			0.4206	
Overidentification test						
Chi-square statistics dof		Chi(56)			Chi(56)	
Chi-square statistics		39.51			41.54	
P-value		0.9534			0.9252	
P-value for testing coefficients equal to zero						
Education	0.2554	0.3416	0.2949	0.2555	0.3374	0.2818
Assets	0.7173	0.7921	0.5794	0.7307	0.8008	0.5922
Age dummies	0.0003	0.0013	0.0001	0.0003	0.0011	0.0001
Marital Status				0.4684	0.4672	0.3977
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Also included in all regressions are age dummies. In 2SLS BMI 91 is instrumented with real productive assets in 89, land cultivated and interactions of real productive assets and land with community free market prices of rice, eggs, pork and fish, as well as water, sanitation and clinic characteristics in 89. Huber/White robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 level; ** at 0.05.

Table 1.14: Determinants of Adults BMI in 1997 Conditional on BMI in 1993: Overall

Variables	Male					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 93	0.391 (0.066)**	0.436 (0.144)**	0.640 (0.091)**	0.389 (0.067)**	0.436 (0.143)**	0.638 (0.091)**
Some primary 93	-0.476 (0.262)*	-0.488 (0.261)*	-0.594 (0.230)**	-0.494 (0.261)*	-0.507 (0.259)*	-0.597 (0.228)**
Primary 93	0.128 (0.326)	0.134 (0.327)	-0.034 (0.259)	0.103 (0.326)	0.111 (0.328)	-0.037 (0.259)
Middle school 93	0.290 (0.356)	0.295 (0.357)	0.104 (0.293)	0.253 (0.354)	0.259 (0.357)	0.089 (0.291)
High school 93	0.208 (0.365)	0.200 (0.364)	-0.037 (0.298)	0.182 (0.364)	0.175 (0.363)	-0.052 (0.297)
Tech/College+ 93	0.514 (0.464)	0.504 (0.456)	0.651 (0.783)	0.476 (0.464)	0.468 (0.454)	0.688 (0.787)
Log R prod assets 93 1	-0.077 (0.043)*	-0.075 (0.043)*	-0.056 (0.034)	-0.079 (0.043)*	-0.077 (0.043)*	-0.058 (0.034)*
Log R prod assets 93 2	0.092 (0.040)**	0.089 (0.039)**	0.075 (0.032)**	0.093 (0.040)**	0.090 (0.040)**	0.075 (0.031)**
Land 93	0.005 (0.011)	0.004 (0.011)	-0.006 (0.010)	0.004 (0.011)	0.004 (0.011)	-0.006 (0.010)
Married 93				0.302 (0.351)	0.297 (0.351)	0.155 (0.304)
Divorced Separated 93				-0.288 (0.854)	-0.198 (0.857)	0.382 (0.800)
Widowed 93				-0.329 (0.500)	-0.296 (0.498)	0.076 (0.455)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	1815	1815	1815	1815	1815	1815
R-squared	0.385	0.383	0.306	0.386	0.384	0.302
First stage for BMI 93						
R-squared		0.259			0.261	
F-statistic for Identifying IVs		1.28			1.14	
P-value		0.0807			0.2296	
Test of homoscedasticity						
Chi-sq stat	70.55	40.69		70.89	41.68	
P-value	0.0000	0.0000		0.0000	0.0000	
Hausman Test between OLS and 2SLS: BMI 93						
Chi-square statistics		0.24			0.26	
P-value		0.6242			0.6096	
Overidentification test						
Chi-square statistics dof		Chi(56)			Chi(56)	
Chi-square statistics		65.03			63.71	
P-value		0.1911			0.2235	
P-value for testing coefficients equal to zero						
Education	0.0056	0.0070	0.0003	0.0066	0.0080	0.0003
Assets	0.0663	0.0765	0.0605	0.0649	0.0758	0.0561
Age dummies	0.0000	0.0000	0.0047	0.0000	0.0000	0.0084
Marital Status				0.2590	0.3811	0.9290
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

Table 1.14 (cont'd)

Variables	Female					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 93	0.481 (0.075)**	0.399 (0.155)**	0.624 (0.099)**	0.483 (0.075)**	0.393 (0.154)**	0.592 (0.092)**
Some primary 93	0.196 (0.292)	0.237 (0.299)	0.024 (0.190)	0.165 (0.287)	0.210 (0.295)	0.021 (0.189)
Primary 93	0.047 (0.316)	0.082 (0.316)	-0.081 (0.243)	0.017 (0.314)	0.056 (0.313)	-0.075 (0.243)
Middle school 93	0.001 (0.351)	-0.002 (0.355)	0.071 (0.250)	-0.043 (0.347)	-0.045 (0.351)	0.055 (0.250)
High school 93	-0.336 (0.443)	-0.383 (0.459)	-0.086 (0.354)	-0.350 (0.440)	-0.400 (0.456)	-0.104 (0.351)
Tech/College+ 93	-1.023 (0.628)	-1.052 (0.609)*	-0.848 (0.534)	-1.004 (0.631)	-1.031 (0.610)*	-0.840 (0.535)
Log R prod assets 93 1	-0.053 (0.048)	-0.047 (0.049)	-0.023 (0.040)	-0.054 (0.048)	-0.047 (0.049)	-0.027 (0.040)
Log R prod assets 93 2	0.048 (0.038)	0.050 (0.039)	0.035 (0.034)	0.047 (0.038)	0.049 (0.038)	0.038 (0.034)
Land 93	0.005 (0.015)	0.003 (0.015)	-0.005 (0.013)	0.005 (0.014)	0.003 (0.014)	-0.006 (0.012)
Married 93				0.361 (0.428)	0.408 (0.459)	0.302 (0.408)
Divorced Separated 93				0.315 (0.846)	0.249 (0.897)	0.750 (0.761)
Widowed 93				-0.755 (0.570)	-0.680 (0.583)	-0.530 (0.489)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	1948	1948	1948	1948	1948	1948
R-squared	0.352	0.348	0.3219	0.355	0.350	0.330
First stage for BMI 93						
R-squared		0.246			0.247	
F-statistic for Identifying IVs		1.29			1.20	
P-value		0.0767			0.1437	
Test of homoscedasticity						
Chi-sq stat	100.69	40.43		102.23	43.82	
P-value	0.0000	0.0000		0.0000	0.0000	
Hausman Test between OLS and 2SLS: BMI 93						
Chi-square statistics		0.31			0.37	
P-value		0.5769			0.5437	
Overidentification test						
Chi-square statistics dof		Chi(56)			Chi(56)	
Chi-square statistics		46.91			48.07	
P-value		0.8015			0.7656	
P-value for testing coefficients equal to zero						
Education	0.4125	0.3383	0.6858	0.4786	0.3865	0.7065
Assets	0.4247	0.4240	0.5843	0.4180	0.4265	0.5223
Age dummies	0.0000	0.0000	0.0000	0.0000	0.0006	0.0000
Marital Status				0.0322	0.0303	0.0167
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.6330

Note: Also included in all regressions are age dummies. In 2SLS BMI 93 is instrumented with real productive assets in 1991, land cultivated and interactions of real productive assets and land with community free market prices of rice, eggs, pork and fish, as well as water, sanitation and clinic characteristics in 1991. Huber/White robust standard errors are in parentheses. * indicates the coefficient is statistically significant at 0.1 level; ** at 0.05.

Table 1.15: Determinants of 1993 BMI Conditional on 1991 BMI with Changes in Community Characteristics

Variables	Male					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 91	0.610 (0.062)**	0.881 (0.111)**	0.883 (0.049)**	0.607 (0.062)**	0.874 (0.113)**	0.876 (0.050)**
Some primary 91	-0.027 (0.214)	0.079 (0.227)	0.053 (0.193)	-0.068 (0.214)	0.050 (0.229)	0.024 (0.193)
Primary 91	0.011 (0.204)	0.099 (0.206)	0.034 (0.166)	-0.028 (0.203)	0.071 (0.206)	0.009 (0.165)
Middle school 91	-0.039 (0.191)	-0.004 (0.199)	0.105 (0.176)	-0.094 (0.192)	-0.042 (0.201)	0.060 (0.177)
High school 91	0.128 (0.231)	0.129 (0.237)	0.247 (0.212)	0.078 (0.229)	0.096 (0.234)	0.222 (0.209)
Tech/College+ 91	0.365 (0.267)	0.344 (0.272)	0.387 (0.249)	0.326 (0.265)	0.319 (0.269)	0.367 (0.247)
Log real prod asset 91 1	0.004 (0.026)	0.013 (0.026)	0.016 (0.021)	0.003 (0.026)	0.012 (0.026)	0.016 (0.021)
Log real prod asset 91 2	0.065 (0.036)*	0.050 (0.032)	0.025 (0.022)	0.063 (0.036)*	0.048 (0.032)	0.024 (0.022)
Land farmed in 90	-0.040 (0.022)*	-0.033 (0.021)	-0.015 (0.016)	-0.042 (0.021)*	-0.034 (0.021)	-0.015 (0.015)
Married 91				0.516 (0.249)**	0.358 (0.272)	0.350 (0.212)*
Divorced or Separated 91				0.173 (0.637)	0.240 (0.617)	0.318 (0.582)
Widowed 91				-0.040 (0.390)	-0.038 (0.408)	-0.026 (0.366)
Urban residence	0.402 (0.179)**	0.216 (0.200)	0.314 (0.151)**	0.402 (0.179)**	0.219 (0.200)	0.323 (0.151)**
Δ Price of rice	-0.010 (0.004)**	-0.007 (0.003)**	-0.004 (0.003)*	-0.009 (0.004)**	-0.006 (0.003)**	-0.004 (0.003)
Δ Price of eggs	0.001 (0.018)	0.005 (0.020)	-0.001 (0.015)	0 (0.018)	0.004 (0.020)	-0.002 (0.015)
Δ Price of pork	-0.017 (0.014)	-0.010 (0.014)	-0.011 (0.011)	-0.017 (0.014)	-0.010 (0.014)	-0.011 (0.011)
Δ Price of fish	-0.027 (0.027)	-0.013 (0.027)	0.003 (0.022)	-0.023 (0.027)	-0.011 (0.027)	0.004 (0.021)
Δ Clinic physicians	-16.016 (16.580)	-10.186 (16.154)	-10.965 (13.762)	-16.356 (16.629)	-10.469 (16.147)	-10.156 (13.730)
Δ Clinic nurses	38.155 (56.472)	43.631 (53.733)	44.652 (43.525)	40.777 (56.516)	45.346 (53.851)	48.568 (43.498)
Δ Clinic beds	-7.229 (17.994)	-9.625 (17.143)	-8.745 (14.216)	-7.468 (17.982)	-9.753 (17.126)	-9.694 (14.181)
Δ Water source 2	0.386 (0.341)	0.142 (0.373)	0.277 (0.300)	0.377 (0.340)	0.139 (0.372)	0.292 (0.298)
Δ Water source 3	0.545 (0.428)	0.766 (0.451)*	0.717 (0.360)**	0.533 (0.433)	0.754 (0.456)*	0.714 (0.363)**
Δ Water source 4	1.862 (0.464)**	1.634 (0.475)**	1.183 (0.374)**	1.910 (0.465)**	1.671 (0.478)**	1.234 (0.372)**

table continues

Table 1.15 (cont'd)

	Male					
	OLS	2SLS	GMM	OLS	2SLS	GMM
Δ In house no flush	-1.098 (1.567)	-1.042 (1.555)	-1.294 (0.996)	-1.051 (1.577)	-1.011 (1.560)	-1.270 (0.993)
Δ Outside toilets	1.054 (0.734)	1.337 (0.846)	0.793 (0.655)	1.050 (0.734)	1.327 (0.846)	0.808 (0.656)
Δ Open pit	1.520 (0.698)**	1.798 (0.796)**	0.995 (0.628)	1.507 (0.699)**	1.783 (0.797)**	0.993 (0.628)
Δ No toilets	1.020 (1.316)	1.087 (1.343)	0.425 (1.121)	0.922 (1.322)	1.017 (1.351)	0.339 (1.122)
Δ Very little excreta	-0.274 (0.309)	-0.389 (0.351)	-0.207 (0.246)	-0.275 (0.309)	-0.388 (0.351)	-0.219 (0.246)
Δ Some excreta	-0.181 (0.288)	-0.067 (0.322)	0.225 (0.235)	-0.203 (0.288)	-0.086 (0.321)	0.212 (0.234)
Δ Much excreta	0.503 (1.068)	-0.539 (1.366)	0.985 (0.923)	0.573 (1.060)	-0.466 (1.361)	1.056 (0.911)
Community Dummies	No	No	No	No	No	No
Number of obs	2436	2436	2436	2436	2436	2436
R-squared	0.315	0.269	0.259	0.317	0.272	0.262
First stage for BMI 91						
R-squared		0.133			0.135	
F-statistic for Identifying IVs		4.64			4.96	
P-value		0.0000			0.0000	
Test of homoscedasticity						
Chi-sq stat	101.58	37.72		98.99	34.77	
P-value	0.0000	0.0000		0.0000	0.0000	
Hausman Test between OLS and 2SLS for BMI 91						
Chi-square statistics		7.35			7.04	
P-value		0.0067			0.0080	
Overidentification test for a third instrument						
Chi-square statistics dof		chi(56)			chi(56)	
Chi-square statistics		67.52			64.50	
P-value		0.1392			0.2038	
P-value for testing coefficients equal to zero						
Education	0.6972	0.8070	0.6175	0.6892	0.8129	0.6309
Productive Assets	0.1115	0.1455	0.1349	0.1432	0.1766	0.1381
Age dummies	0.0366	0.1213	0.0383	0.0240	0.0955	0.0139
Marital Status				0.0620	0.4210	0.2337
Prices	0.0174	0.1649	0.4128	0.0242	0.1917	0.4412
Water source	0.0006	0.0035	0.0118	0.0004	0.0027	0.0080
Toilet type	0.0820	0.0569	0.1951	0.0944	0.0631	0.2004
Excreta	0.7909	0.6854	0.2650	0.7629	0.7049	0.2441
Clinic chars	0.2669	0.4081	0.1607	0.2258	0.3747	0.1156

Table continues

Table 1.15 (cont'd)

Variables	Female					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 91	0.595 (0.041)**	0.837 (0.074)**	0.871 (0.043)**	0.595 (0.041)**	0.837 (0.074)**	0.868 (0.043)**
Some primary 91	0.104 (0.170)	0.090 (0.182)	0.001 (0.163)	0.105 (0.170)	0.092 (0.182)	-0.007 (0.163)
Primary 91	0.308 (0.196)	0.166 (0.216)	-0.068 (0.158)	0.302 (0.196)	0.160 (0.215)	-0.065 (0.158)
Middle school 91	0.150 (0.191)	0.123 (0.198)	0.106 (0.169)	0.141 (0.190)	0.114 (0.198)	0.095 (0.169)
High school 91	-0.240 (0.236)	-0.259 (0.257)	-0.151 (0.235)	-0.234 (0.234)	-0.252 (0.256)	-0.157 (0.234)
Tech/College+ 91	0.277 (0.323)	0.402 (0.323)	0.364 (0.296)	0.297 (0.327)	0.422 (0.328)	0.398 (0.300)
Log real prod asset 91 1	0.018 (0.033)	0.004 (0.033)	0.026 (0.026)	0.018 (0.034)	0.004 (0.033)	0.026 (0.026)
Log real prod asset 91 2	0.002 (0.033)	0.004 (0.033)	-0.016 (0.027)	0.001 (0.033)	0.003 (0.033)	-0.015 (0.026)
Land farmed in 90	0.005 (0.025)	0.020 (0.026)	0.008 (0.020)	0.005 (0.025)	0.020 (0.026)	0.007 (0.020)
Married 91				0.381 (0.418)	0.371 (0.459)	0.374 (0.399)
Divorced or Separated 91				0.743 (0.629)	0.599 (0.649)	0.744 (0.599)
Widowed 91				-0.065 (0.499)	-0.010 (0.532)	0.120 (0.472)
Urban residence	0.487 (0.155)**	0.266 (0.174)	0.204 (0.148)	0.482 (0.155)**	0.263 (0.174)	0.197 (0.147)
Δ Price of rice	0.003 (0.005)	0.002 (0.004)	0.004 (0.003)	0.002 (0.005)	0.002 (0.004)	0.005 (0.003)
Δ Price of eggs	-0.017 (0.019)	-0.022 (0.019)	-0.024 (0.016)	-0.017 (0.019)	-0.022 (0.020)	-0.026 (0.016)*
Δ Price of pork	-0.011 (0.012)	-0.003 (0.011)	-0.001 (0.009)	-0.011 (0.011)	-0.003 (0.011)	-0.001 (0.009)
Δ Price of fish	-0.049 (0.027)*	-0.041 (0.029)	-0.048 (0.024)**	-0.048 (0.027)*	-0.040 (0.029)	-0.046 (0.024)**
Δ Clinic physicians	-18.546 (20.299)	-4.112 (22.970)	3.699 (17.435)	-19.062 (20.409)	-4.577 (23.112)	4.049 (17.514)
Δ Clinic nurses	-81.505 (95.088)	-34.404 (106.567)	-2.262 (82.692)	-84.451 (95.471)	-36.525 (106.880)	-3.803 (82.521)
Δ Clinic beds	0.945 (25.309)	-13.465 (29.868)	-12.685 (23.102)	1.943 (25.442)	-12.713 (29.989)	-11.324 (23.089)
Δ Water source 2	-0.311 (0.326)	-0.191 (0.332)	0.008 (0.276)	-0.316 (0.326)	-0.196 (0.332)	-0.018 (0.277)
Δ Water source 3	-0.308 (0.525)	0.181 (0.539)	0.407 (0.432)	-0.300 (0.529)	0.187 (0.545)	0.357 (0.436)
Δ Water source 4	-0.404 (0.510)	-0.412 (0.518)	0.156 (0.412)	-0.411 (0.511)	-0.419 (0.519)	0.106 (0.416)

table continues

Table 1.15 (cont'd)

	Female					
	OLS	2SLS	GMM	OLS	2SLS	GMM
Δ In house no flush	-0.262 (1.019)	-0.298 (1.036)	-0.279 (0.893)	-0.261 (1.017)	-0.296 (1.035)	-0.215 (0.889)
Δ Outside toilets	-0.076 (0.610)	0.202 (0.701)	0.334 (0.588)	-0.110 (0.613)	0.173 (0.704)	0.276 (0.592)
Δ Open pit	-0.186 (0.693)	-0.076 (0.784)	0.206 (0.643)	-0.227 (0.698)	-0.112 (0.789)	0.137 (0.648)
Δ No toilets	0.464 (1.170)	-0.014 (1.265)	-0.660 (1.053)	0.410 (1.166)	-0.063 (1.265)	-0.810 (1.053)
Δ Very little excreta	-0.225 (0.287)	-0.237 (0.310)	-0.331 (0.254)	-0.236 (0.285)	-0.247 (0.308)	-0.331 (0.253)
Δ Some excreta	0.832 (0.327)**	0.929 (0.350)**	0.606 (0.268)**	0.804 (0.325)**	0.904 (0.348)**	0.594 (0.266)**
Δ Much excreta	0.656 (1.101)	-0.522 (1.084)	-0.413 (0.954)	0.616 (1.105)	-0.558 (1.086)	-0.385 (0.954)
Community Dummies	No	No	No	No	No	No
Number of obs	2638	2638	2638	2638	2638	2638
R-squared	0.391	0.341	0.318	0.392	0.342	0.320
First stage for BMI 91						
R-squared		0.124			0.125	
F-statistic for Identifying IVs		6.93			7.44	
P-value		0.0000			0.0000	
Test homoscedasticity						
Chi-sq stat	98.67	40.02		99.30	40.15	
P-value	0.0000	0.0000		0.0000	0.0000	
Hausman Test between OLS and 2SLS for BMI 91						
Chi-square statistics		11.61			11.60	
P-value		0.0007			0.0007	
Overidentification test for a third instrument						
Chi-square statistics dof		chi(56)			chi(56)	
Chi-square statistics		58.48			56.89	
P-value		0.3846			0.4418	
P-value for testing coefficients equal to zero						
Education	0.3296	0.5643	0.6960	0.3514	0.5734	0.6665
Productive Assets	0.7171	0.9567	0.6077	0.7332	0.9634	0.5975
Age dummies	0.0027	0.0292	0.0196	0.0050	0.0299	0.0204
Marital Status				0.2493	0.4163	0.4329
Prices	0.0779	0.2752	0.0681	0.0768	0.2768	0.0582
Water source	0.7857	0.6972	0.8194	0.7807	0.6827	0.8629
Toilet type	0.8560	0.6224	0.4429	0.8436	0.6021	0.4613
Excreta	0.0093	0.0087	0.0164	0.0114	0.0097	0.0176
Clinic chars	0.5767	0.7177	0.8846	0.5730	0.7190	0.9074

Note: Also included in all regressions are age dummies for 1993. BMI 91 is instrumented by real productive assets and land cultivated in 1989 and interactions of these variables with free market prices for rice, eggs, pork and fish, clinic characteristics, water and sanitation variables in 1989. Huber/White robust standard errors are in parentheses.

Table 1.16: Determinants of 1997 BMI Conditional on 1993 BMI with Changes in Community Characteristics

Variables	Male					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 93	0.459 (0.070)**	0.485 (0.154)**	0.774 (0.079)**	0.455 (0.070)**	0.483 (0.153)**	0.776 (0.080)**
Some primary 93	-0.572 (0.249)**	-0.571 (0.248)**	-0.610 (0.228)**	-0.592 (0.247)**	-0.590 (0.246)**	-0.617 (0.226)**
Primary 93	-0.088 (0.329)	-0.081 (0.331)	-0.244 (0.271)	-0.114 (0.328)	-0.106 (0.331)	-0.236 (0.270)
Middle school 93	0.300 (0.337)	0.303 (0.338)	0.212 (0.291)	0.258 (0.335)	0.263 (0.338)	0.209 (0.291)
High school 93	0.159 (0.333)	0.153 (0.333)	-0.132 (0.288)	0.138 (0.331)	0.132 (0.331)	-0.133 (0.286)
Tech/College+ 93	0.500 (0.364)	0.491 (0.362)	0.280 (0.321)	0.464 (0.362)	0.456 (0.359)	0.272 (0.320)
Log real prod asset 93 1	0.021 (0.033)	0.021 (0.033)	0.028 (0.029)	0.020 (0.033)	0.020 (0.033)	0.027 (0.029)
Log real prod asset 93 2	0.030 (0.033)	0.029 (0.032)	0.010 (0.028)	0.031 (0.033)	0.030 (0.032)	0.011 (0.028)
Land farmed in 92	-0.008 (0.012)	-0.007 (0.012)	-0.002 (0.008)	-0.008 (0.012)	-0.007 (0.012)	-0.002 (0.008)
Married 93				0.271 (0.310)	0.256 (0.326)	-0.107 (0.265)
Divd Septd 93				-1.131 (0.959)	-1.067 (0.989)	-0.277 (0.916)
Widowed 93				-0.563 (0.465)	-0.547 (0.462)	-0.348 (0.432)
Urban residence	1.137 (0.224)**	1.101 (0.287)**	0.789 (0.212)**	1.128 (0.224)**	1.089 (0.286)**	0.772 (0.213)**
△ Price of rice	0.015 (0.006)**	0.014 (0.006)**	0.006 (0.005)	0.014 (0.006)**	0.014 (0.006)**	0.006 (0.005)
△ Price of eggs	-0.007 (0.020)	-0.008 (0.022)	-0.034 (0.016)**	-0.006 (0.020)	-0.008 (0.022)	-0.033 (0.016)**
△ Price of pork	0.002 (0.011)	0.002 (0.012)	0.015 (0.010)	0.002 (0.011)	0.003 (0.012)	0.014 (0.010)
△ Price of fish	-0.007 (0.003)**	-0.007 (0.003)**	-0.007 (0.003)**	-0.007 (0.003)**	-0.007 (0.003)**	-0.007 (0.003)**
△ Water source 2	0.589 (0.356)*	0.579 (0.355)	0.439 (0.265)*	0.596 (0.355)*	0.585 (0.354)*	0.463 (0.262)*
△ Water source 3	-0.326 (0.363)	-0.332 (0.360)	-0.100 (0.313)	-0.348 (0.361)	-0.354 (0.358)	-0.093 (0.310)
△ Water source 4	-0.016 (0.360)	-0.006 (0.366)	0.286 (0.323)	-0.006 (0.362)	0.005 (0.367)	0.309 (0.324)

table continue

Table 1.16 (cont'd)

	Male					
	OLS	2SLS	GMM	OLS	2SLS	GMM
△ In house no flush	-0.527 (0.588)	-0.508 (0.578)	-0.289 (0.485)	-0.560 (0.581)	-0.538 (0.572)	-0.296 (0.482)
△ Outside toilets	0.455 (0.501)	0.418 (0.543)	-0.268 (0.432)	0.465 (0.505)	0.426 (0.545)	-0.266 (0.435)
△ Open pit cement or earth	0.791 (0.445)*	0.758 (0.476)	0.149 (0.386)	0.793 (0.447)*	0.758 (0.476)	0.150 (0.388)
△ No toilets	0.452 (1.381)	0.437 (1.371)	0.977 (1.052)	0.542 (1.389)	0.522 (1.379)	0.955 (1.056)
△ Very little excreta	0.137 (0.459)	0.153 (0.466)	0.309 (0.359)	0.123 (0.459)	0.140 (0.467)	0.291 (0.360)
△ Some excreta	-0.581 (0.458)	-0.551 (0.487)	0.006 (0.367)	-0.526 (0.455)	-0.497 (0.478)	0.022 (0.364)
△ Much excreta	1.592 (1.014)	1.621 (1.016)	1.580 (0.802)**	1.502 (1.012)	1.538 (1.018)	1.522 (0.797)*
Community Dummies	No	No	No	No	No	No
Number of obs	1777	1777	1777	1777	1777	1777
R-squared	0.295	0.294	0.201	0.297	0.296	0.1995
First stage for BMI 93						
R-squared		0.141			0.146	
F-statistic for Identifying IVs		2.12			2.30	
P-value		0.0000			0.0000	
Test of homoscedasticity						
Chi-sq stat	95.80	18.77		94.28	19.03	
P-value	0.0000	0.0001		0.0000	0.0001	
Hausman Test between OLS and 2SLS for BMI 93						
Chi-square statistics		0.09			0.10	
P-value		0.7626			0.7480	
Overidentification test for a third instrument						
Chi-square statistics dof		chi(56)			chi(56)	
Chi-square statistics		63.79			70.04	
P-value		0.2216			0.0983	
P-value for testing coefficients equal to zero						
Education	0.0005	0.0005	0.0002	0.0006	0.0006	0.0002
Productive Assets	0.1989	0.2046	0.2854	0.2001	0.2052	0.2788
Age dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Marital Status				0.0550	0.1652	0.8811
Prices	0.0168	0.0217	0.0128	0.0202	0.0248	0.0137
Water source	0.2088	0.2111	0.2800	0.1809	0.1841	0.2330
Toilet type	0.0875	0.1140	0.3555	0.0844	0.1100	0.3625
Excreta	0.2323	0.2434	0.2478	0.2906	0.2948	0.2692

Table continues

Table 1.16 (cont'd)

Variables	Female					
	OLS	2SLS	GMM	OLS	2SLS	GMM
BMI 93	0.532 (0.077)**	0.526 (0.114)**	0.670 (0.062)**	0.532 (0.076)**	0.522 (0.114)**	0.638 (0.065)**
Some primary 93	0.116 (0.256)	0.119 (0.262)	0.118 (0.191)	0.086 (0.252)	0.089 (0.259)	0.094 (0.188)
Primary 93	0.250 (0.273)	0.252 (0.269)	0.298 (0.228)	0.223 (0.271)	0.227 (0.267)	0.263 (0.227)
Middle school 93	0.237 (0.292)	0.237 (0.292)	0.334 (0.235)	0.194 (0.290)	0.195 (0.289)	0.284 (0.232)
High school 93	-0.070 (0.418)	-0.072 (0.432)	0.359 (0.353)	-0.102 (0.416)	-0.107 (0.430)	0.290 (0.351)
Tech/College+ 93	-0.649 (0.477)	-0.650 (0.477)	-0.443 (0.445)	-0.658 (0.480)	-0.660 (0.479)	-0.472 (0.447)
Log real prod asset 93 1	0.046 (0.039)	0.046 (0.038)	0.053 (0.032)*	0.048 (0.038)	0.049 (0.038)	0.054 (0.032)*
Log real prod asset 93 2	-0.004 (0.031)	-0.003 (0.032)	-0.013 (0.027)	-0.008 (0.031)	-0.007 (0.032)	-0.013 (0.027)
Land farmed in 92	0.000 (0.013)	0.000 (0.012)	-0.011 (0.010)	0.000 (0.012)	0.000 (0.012)	-0.011 (0.009)
Married 93				0.101 (0.429)	0.103 (0.434)	0.133 (0.408)
Divd Septd 93				-0.569 (0.836)	-0.574 (0.841)	-0.209 (0.736)
Widowed 93				-0.980 (0.560)*	-0.979 (0.562)*	-0.735 (0.506)
Urban residence	0.660 (0.237)**	0.665 (0.241)**	0.325 (0.185)*	0.647 (0.237)**	0.655 (0.241)**	0.375 (0.186)**
△ Price of rice	-0.019 (0.013)	-0.019 (0.013)	-0.012 (0.010)	-0.019 (0.013)	-0.019 (0.013)	-0.013 (0.010)
△ Price of eggs	0.032 (0.018)*	0.032 (0.018)*	0.016 (0.015)	0.032 (0.018)*	0.032 (0.018)*	0.016 (0.015)
△ Price of pork	0.006 (0.019)	0.006 (0.019)	-0.001 (0.012)	0.007 (0.019)	0.007 (0.019)	0.002 (0.012)
△ Price of fish	-0.004 (0.002)	-0.004 (0.002)	-0.003 (0.002)	-0.004 (0.003)	-0.004 (0.003)	-0.003 (0.002)
△ Water source 2	-0.061 (0.307)	-0.058 (0.317)	-0.102 (0.251)	-0.025 (0.305)	-0.022 (0.315)	-0.064 (0.250)
△ Water source 3	-0.035 (0.478)	-0.027 (0.515)	-0.117 (0.397)	-0.119 (0.475)	-0.105 (0.513)	-0.142 (0.398)
△ Water source 4	0.463 (0.418)	0.464 (0.419)	0.503 (0.344)	0.510 (0.416)	0.511 (0.417)	0.569 (0.345)*

table continues

Table 1.16 (cont'd)

	Female					
	OLS	2SLS	GMM	OLS	2SLS	GMM
△ In house no flush	-1.375 (0.655)**	-1.376 (0.652)**	-1.378 (0.568)**	-1.362 (0.653)**	-1.363 (0.650)**	-1.385 (0.566)**
△ Outside toilets	-0.005 (0.487)	-0.016 (0.554)	-0.085 (0.454)	-0.033 (0.489)	-0.050 (0.556)	-0.137 (0.458)
△ Open pit	0.040 (0.482)	0.032 (0.522)	-0.057 (0.403)	-0.007 (0.485)	-0.020 (0.525)	-0.137 (0.409)
△ No toilets	-2.689 (1.135)**	-2.700 (1.170)**	-2.224 (1.030)**	-2.604 (1.153)**	-2.621 (1.189)**	-2.179 (1.051)**
△ Very little excreta	0.105 (0.379)	0.104 (0.379)	0.049 (0.301)	0.102 (0.378)	0.100 (0.378)	0.109 (0.303)
△ Some excreta	0.556 (0.463)	0.558 (0.464)	0.487 (0.376)	0.602 (0.465)	0.605 (0.468)	0.565 (0.382)
△ Much excreta	-0.421 (1.284)	-0.448 (1.308)	-0.006 (0.974)	-0.366 (1.276)	-0.409 (1.303)	-0.090 (0.984)
Community Dummies	No	No	No	No	No	No
Number of obs	1901	1901	1901	1901	1901	1901
R-squared	0.292	0.292	0.2709	0.295	0.295	0.2809
First stage for BMI 93						
R-squared		0.138			0.139	
F-stat for Identifying IVs		2.96			2.90	
P-value		0.0000			0.0000	
Test homoscedasticity						
Chi-sq stat	122.75	20.93		122.68	22.80	
P-value	0.0000	0.0000		0.0000	0.0000	
Hausman Test between OLS and 2SLS for BMI 93						
Chi-square statistics		0.00			0.01	
P-value		0.9495			0.9199	
Overidentification test for a third instrument						
Chi-square statistics dof		chi(56)			chi(56)	
Chi-square statistics		69.29			68.77	
P-value		0.1094			0.1175	
P-value for testing coefficients equal to zero						
Education	0.5374	0.5141	0.4293	0.5975	0.5732	0.5297
Productive Assets	0.3614	0.3212	0.2018	0.3583	0.3139	0.1855
Age dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Marital Status				0.0345	0.0338	0.0408
Prices	0.0847	0.1158	0.3383	0.1033	0.1351	0.3374
Water source	0.6190	0.6198	0.3297	0.5620	0.5653	0.2620
Toilet type	0.0805	0.0799	0.0721	0.1075	0.1057	0.0848
Excreta	0.6704	0.6716	0.6305	0.6275	0.6270	0.5331

Note: Also included in all regressions are age dummies for 1993. BMI 91 is instrumented by real productive assets and land cultivated in 1989 and interactions of these variables with free market prices for rice, eggs, pork and fish, clinic characteristics, water and sanitation variables in 1989. Huber/White robust standard errors are in parentheses.

1.7 Concluding Remarks

Using an unique longitudinal data set of adult Chinese in the 1990s we estimate the reduced form and the dynamic BMI demand function for men and women in an era of transition in China. This paper fills in the gap of previous research that lacks clear theoretical framework and becomes the first paper on Chinese adult BMI estimations with emphasis on individual as well as community characteristics.

The key findings of this paper are for women the effect of education is very strong and inversely U-shaped, even after controlling for community characteristics. The education effect is stronger in rural areas than in urban areas and for younger generations than for older generations. Adult men in China had higher levels of education than women, however their educational effects on BMI were not significant. If the BMI of man is a proxy for both health condition and human capital, the protective effects of education in health (negative) and the enhancing effects of education in human capital (positive) might very well cancel each other out. For women it seems appropriate that BMI plays only the role of a proxy for health condition because a woman's human capital does not depend strongly on her stature. Hence the effect of education on BMI reflects its effect on health condition only. Over time the protective effects of education on women dissolve and the enhancing effects for men increase (though not significant conditional on household and community variables). There is room for improvement for both men and women through increasing education quantity and quality.

Assets, prices and environmental health conditions are all important determinants of adult BMIs in China and these factors affect men and women in different age groups and across regions differently. The effect of productive assets differs between men and women. For men, more productive assets are associated with higher BMI, with larger effects in the upper spline range and the joint effects are highly significant. For women, higher level of productive assets are associated with higher

BMI only for those with assets less than the median; above the median the association is negative but not significant. Eggs and pork constitute of high calorie intakes, hence the increase in prices may have a substitution effect for other foods with lower calorie and result in lower BMI. Fish and rice, on the other hand, can be thought of as substitute for red meat and wheat flour. The increase in fish and rice prices will hence result in more consumption of less healthy foods and increased BMI.

When the amount of information at the community level is plenty it is critical to choose proper number of variables to include in any model, avoiding both the multicollinearity problem and a fully saturated model. The current food prices affect BMI through intakes depending on which foods have higher prices; and the changes in food prices over time affect BMI depending on which foods have higher changes in prices. There is a trade-off between including community dummies or not when we use current community characteristics in the model. The advantage is community dummies capture unobserved heterogeneity at the aggregate level. The disadvantage is any effect of community variables will have to be identified through the variation of them over time and in a short run the change may be too little for good identification.

In the short term dynamic BMI demand function it seems to be quite reasonable to assume the lagged BMI is a good summary statistic of all past information for the individual. The estimated effect of lagged BMI in 2SLS is bigger than that in OLS estimation. Controlling for lagged BMI the effects of education and household resources in 1991 were no longer significant in both OLS and 2SLS estimates, except that in GMM estimation the primary diploma in 1991 showed strong negative effect for men. The male marriage premium in BMI became insignificant. The conditional age and community effects are significant for women and the age effect conditional on past BMI for men is not jointly significant. With an increase in the price of rice, for example, we would expect to see BMI to increase if substitution (for wheat) effect is stronger than income effect. However, in our sample the increase in wheat price

is actually bigger than the increase in rice price from 1991 to 1993. Therefore there is no strong substitution effect and the coefficient for rice price changes are negative. Although the conditional BMI specification works better with shorter lagged period, the changes in community level information during such short period may not be large enough to correctly identify any effect.

There are some limitations in this paper. First, using the productive asset in the reduced model as a proxy for income and treating it as exogenous is not without criticism. In a dynamic model BMI can affect productivity and therefore the holding of productive assets in later periods. If there was selection into more strenuous vocations with higher wages, the results will be biased upwards. Attenuation bias could also rise with imperfect measures of productive assets and land.

Secondly the dynamic BMI demand function hinges on the assumption that one lagged BMI is a sufficient statistic for all past information. When the time gap between the dependent and the conditional BMI is too large this assumption is unattainable.

Finally, the model is based on the assumption of perfect insights. Ignoring uncertainty is not innocuous when there are risks other than idiosyncratic shocks. Uncertainty would be important if overweight increases risk of mortality and morbidity and people are well aware of it. The length of individual life time and the household size are all assumed to be exogenous whereas in reality both are potentially endogenous.

With the above caveats in mind this paper provides careful exposition of the socioeconomic determinants of adult BMI in China in the 1990s and finds strong association between education and BMI in women and in rural areas. The knowledge of this relationship can assist public policies to identify target groups for improving their health status. As the Chinese economy undergoes the rapid structural transition it is extremely important to find factors that can make such transition as smooth as possible. Increased overweight and obesity in adults imposes high burden on the

health care system, thus obesity awareness education (even general education level) and prevention measures should pay off in the long run.

Chapter 2: Adult Chinese

Macronutrient Consumption and Socioeconomic Determinants in the Early 1990s

2.1 Introduction

Alongside its rapid economic development and social and cultural transitions in the 1990s China also experienced shifts in diet, physical activity and leading causes of death (see Ge et al. 1992 and references in Chapter one). The dietary pattern is moving toward one in which the proportion of energy intake from fat increases each year. The Chinese Recommended Dietary Allowance (CRDA) of 2,400 Kcal was established by the Chinese Nutrition Society in 1981 (Chen 1990, Ge et al. 1991) for adult male aged 18 to 40 and undertaking very light activity. According to the Food and Agriculture Organization (FAO) estimates from the Food Balance Sheet (FBS)¹ in the 1980s the CRDA has been reached, per capita daily protein and fat intakes rose dramatically in the 1980s and 1990s, and both the percent of protein and percent of fat from vegetable products are decreasing and the percents of protein and fat from animal products are increasing (Figure 2.1). The overall calorie supply, for instance, has increased from 1,953 in 1966 to 2,766 kcal in 1996 per person per day. The per capita supply of meat has quadrupled, growing from 77 kcal per person per

¹The FAO collected officially reported data on the production, trade and utilization of agricultural commodities for countries worldwide. These databases are used to set up FBSs which provide essential information on a country's food system (Heilig 1999). How reliable these data are is discussed in the next Section.

day in the mid-1960s to 320 kcal per person per day in the mid-1990s.² According to a household survey of eight provinces in China during the eight-year period of 1989 to 1997 the intake of cereals decreased considerably by 127 grams per person per day; the intake of vegetables decreased by 32 grams per person per day; and the intake of animal foods increased by 46.7 and 36.8 grams per person per day for urban and rural residents (Du et al. 2002). During the rapid economic growth after reforms more health risks are resulted from the deteriorating dieting habits (Guo, Popkin and Zhai 1999, Guo et al. 2000). It is projected that diet-related chronic diseases, such as obesity and coronary heart disease, will present a huge health care burden for China in the near future (Popkin, Paeratakul, Zhai and Ge 1995a).

Similar trends in nutrition transition are found in many low-income countries (Popkin 2002, Drewnowski and Popkin 1997, Guo 1998) but no universal conclusions concerning the roles of socioeconomic determinants on the dietary transitions exit. Increased income expands a household's budget constraint hence resulting in higher quantity and quality of food consumption. Urbanization and increasing exposure to western life styles may change tastes and food choices. Improved education raises the allocative efficiency of food consumption, especially during times of technological change. Higher education is also associated with increased income and investment in human capital. Changes in prices of foods and other community characteristics also influence food consumption through price and/or income effects. In Luo (2003b) we present evidence that many individual, household and community socioeconomic determinants are significantly related with adult body mass index (BMI). It is conceivable these factors should also be related with food consumption, a key input into

²The balance sheets can be only as good as rudimentary information for food availabilities. In theory, food balance sheets should take into account the complete food chain, from the field to supply in the shops. They are an accounting system that balances production, exports, imports, overall domestic supply, use as feed or seeds, waste, inventory change, and direct use for human consumption. Unfortunately, accurate data are not always available, especially for the amount of waste. Missing data have to be estimated so that the sheet can be balanced. Such estimations inevitably introduce inaccuracies. However, the FAO time series on food availability provides useful and valuable information on general trends of dietary patterns.

BMI. Food consumption quantity in this paper refers to individual daily intakes of calories, fats and proteins and food consumption quality is measured by percent of calories from fat, from protein and from carbohydrates. Nutritional status, referring to a person's physical wellbeing as a result of ingestion, absorption and utilization of nutrients, and measured by BMI in our previous study, is not the focus of this paper.

In nutrition literature, diet can be described in terms of its chemical composition, for example, its nutrient content, or alternatively, in terms of foods or food groups. The advantage of representing diet as the total intake of a nutrient rather than using the contribution of only one food at a time is that such information can be directly related to existing knowledge of biology (Willett 1998). The advantage of studying foods or food groups in economic analysis is that the own and cross price elasticities are well defined and easy to interpret. However, there is huge geographic variation in food consumption preference in China. The northern part (north of Yangzi River) of the country produces and consumes more wheat products, while rice cultivation and consumption dominate in the south. The average daily per capita consumption of rice, wheat and other cereals was 25 g, 340 g and 215 g, respectively in the northern provinces and the southern counterparts of these staples consumed was 401 g, 72 g and 42 g in 1982 (Ge et al. 1991). In addition, since nutrient intakes have been viewed as indicators of a person's wellbeing, they are used in this paper to find their determinants rather than finding the determinants for several key foods.

There are generally two categories of nutrient determinant studies: the demand for nutrient intakes and the production of nutritional status as a biological indicator (Behrman and Deolalikar 1988), the latter being beyond the scope of this paper and only lightly touched for completeness of documentation. In the first category, prices, income, endowment and many economic and social factors are presumed to determine the behavioral process. In this paper we focus on testing the effect of socioeconomic factors on demands of total calorie, protein and fat intakes and the percent of energy

from protein, fat and carbohydrates. We use the China Health and Nutrition Survey (CHNS) data in 1989, 1991 and 1993 for adults at or over 20 years of age who were studied in our previous chapter on BMI.

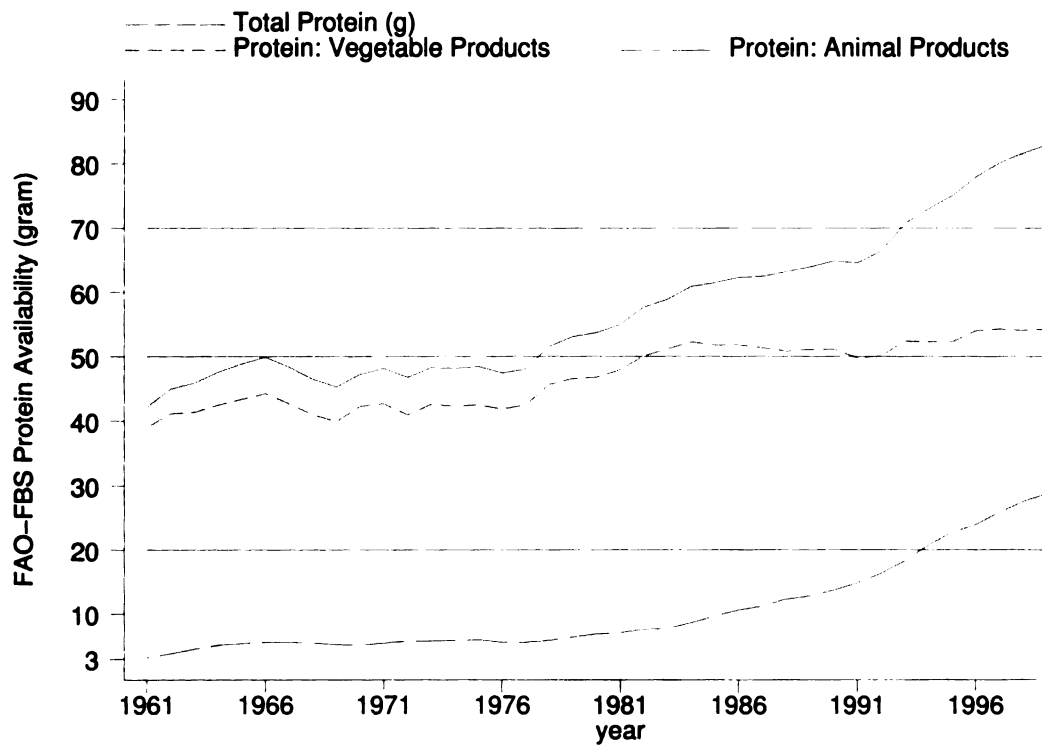
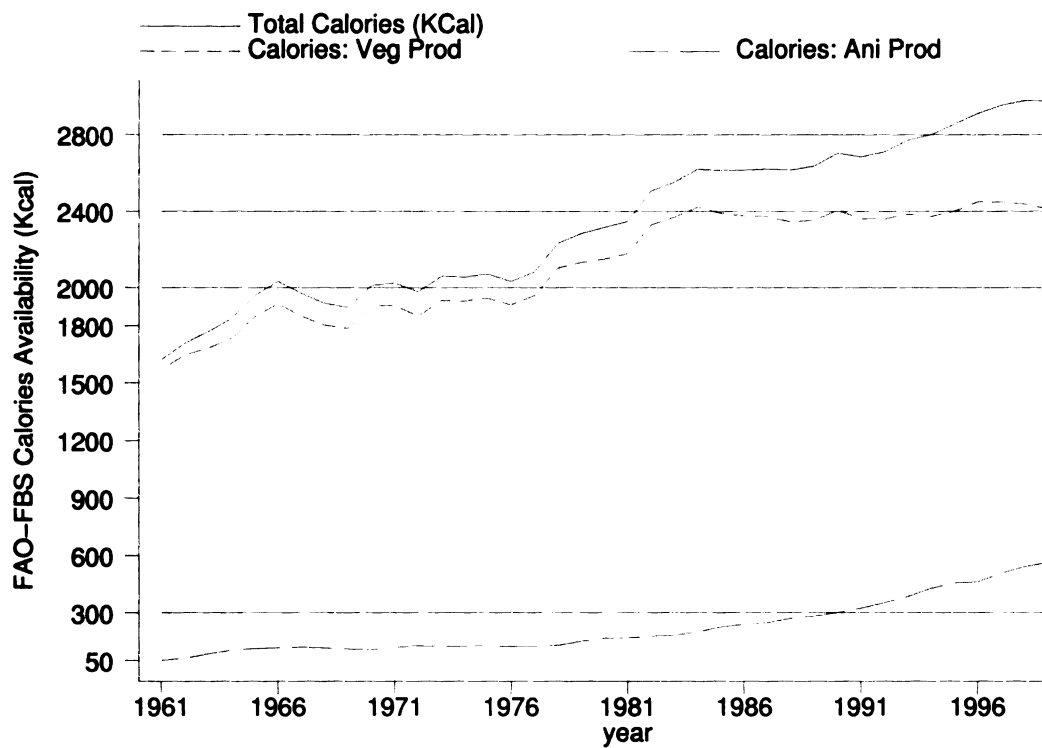
For reasons stated in the earlier chapter we do not focus our attention on the association between income and the above outcomes but instead use productive assets as measure of household resources. In the overall sample education does not have significant impact on calorie intakes but does affect percent of calories from fat, from protein and from carbohydrates differently in different region and at different age. The effect of productive assets is nonlinear and in inverted U-shape for male calorie, fat and protein intakes; whereas for women more productive assets are associated with more fat and protein intakes and more percents of calories from fat and from protein. In rural areas the effect of productive assets is stronger than that in the urban areas. At different ages the effects of assets are different. For the elderly men calorie intakes are related with productive assets in U-shape rather than the inverted U-shape seen for younger men. The effects of productive assets on fat and protein intakes are significant for younger men and women but not for the elderly. Prices of foods, community water and sanitation conditions are also studied. The effects of prices on calorie, fat and protein intakes and the quality of diet measures can go in either direction. The prices of rice and eggs are negatively associated with calorie, fat, protein intakes and percents of calories from fat and protein, and are positively related to percent of calories from carbohydrates. The price of pork is positively associated with all intakes but the percent of calories from carbohydrates. The price of edible oils is negatively associated with calories, fat and percent of calories from fat in the overall sample for men and women but is positively related to the percent of calories from protein for women. Improvements in sanitation are associated with more energy and protein intakes in urban areas.

The basic estimates in this paper use pooled OLS method and focus on individ-

ual and household characteristics controlling for community year interactions. The augmented models focus on the effects of prices, water and sanitation characteristics at community levels. The pooled estimates imply the effects of prices and other determinants over time do not change, which may be hard to maintain. During times of transition, it is important to examine the effects of changing preferences of food consumption caused by changes in the socioeconomic factors. By exploring the longitudinal nature of the data set we are able to find out the socioeconomic determinants of the change of people's food consumption. Hence the changes in all outcomes between 1989 and 1991 and the changes between 1991 and 1993 are pooled together and the OLS method was used in the basic and augmented models again to find out the effects of the changes in prices and other community characteristics on the changes in calorie, fat and protein intakes and the changes in percent of calories from fat, protein and carbohydrates. The changes in prices and community characteristics may not bear the same effect on the changes of the nutrient consumption.

Finally for completeness of documentation, a simple health production function, as measured by weight in 1993 (or BMI in 1993), is estimated using lagged weight, height (or lagged BMI) and current food intakes as inputs. The two stage least squares (2SLS) estimates suggest adult weights between a short period of two years can be modelled as a random walk process.

The paper is organized as follows. Next section reviews the literature in nutrient demand studies, followed by description of the data and discussion of the econometric issues in the reduced form analysis of nutrient intakes. Section IV and V present the basic and augmented models for the levels and changes in outcomes of interest. The penultimate section includes some discussion of the production function analysis. The final section concludes.



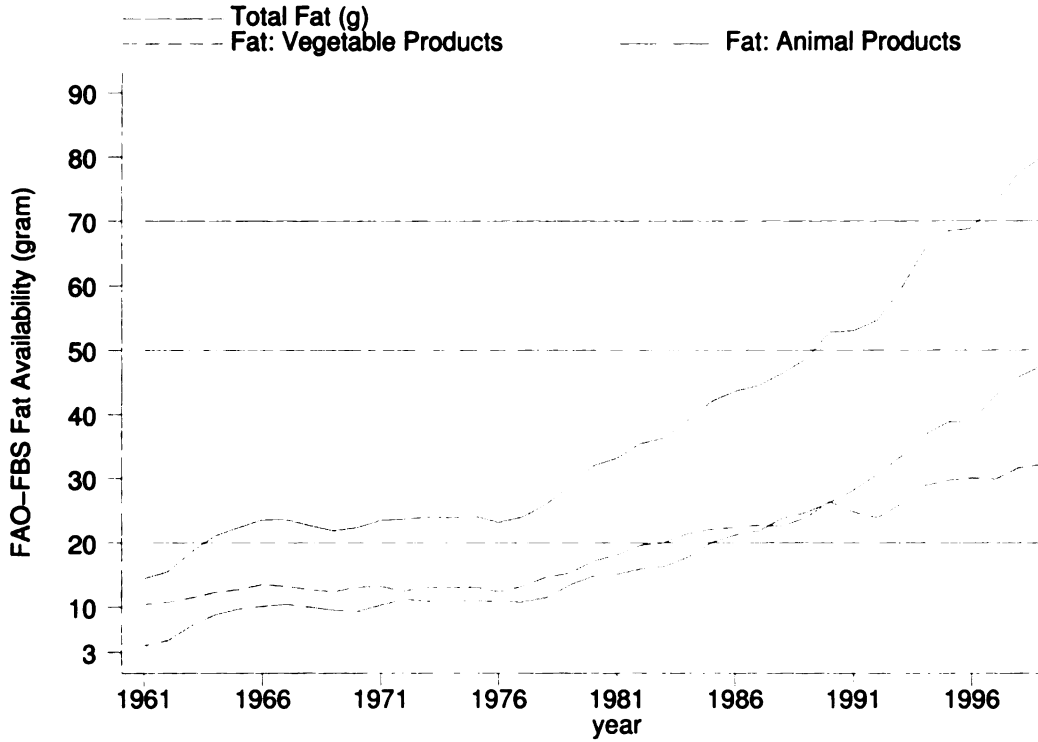


Figure 2.1: FAO Food Balance Sheet Estimates of calories, protein and fat availabilities and percent of protein and fat from vegetable and animal products in 1961-2001.

2.2 Literature Review

2.2.1 Reduced Form Demands

Generally the reduced-form demand function for foods can be solved recursively from a dynamic household maximization of intertemporally separable utility under certainty (Luo 2003*b*). The demand equation for consumption goods x , health or nutritional status h , leisure l , and health inputs m can all be written as

$$(x, h, l, m)_i^t = g_{x,h,l}^t(p_x, p_m, w_i, y_i, \theta_i, \theta_h, z_c; v_i, \xi_h) \quad (2.1)$$

where $p_x = (p_T^x, p_{T-1}^x, \dots, p_0^x)$ are past and future prices for x , $p_m = (p_T^m, p_{T-1}^m, \dots, p_0^m)$ are prices for m , $w_i = (w_{iT}, w_{iT-1}, \dots, w_{i0})$ are wage rates from period 0 to T , $y = (y_{iT}, y_{iT-1}, \dots, y_{i0})$ are household non-wage income, and v_i , ξ_h are unobserved individual and household heterogeneity in input demands over the life time, such as genetic traits, household preferences and environmental factors, are known to family members and are assumed to be distributed independently of socioeconomic determinants of interest. Vectors of exogenous individual (θ_i), household (θ_h), and community (z_c) characteristics are observed, such as education levels, household resources and community infrastructures. In equation (2.1) all the determinants, including individual, household and community characteristics (θ_i, θ_h, z_c) and prices, except for wage, are assumed to be orthogonal to v_{it} , ξ_h . Clearly in a nutritional status production function analysis all nutrient intakes and other health inputs, and leisure should be treated as endogenous. Studies that do not clearly distinguish the two types of analyses are bound to result in bias and fall short of empirical usefulness in policy implications (Schultz 1984). In reviewing the literature we put all studies in perspective of the above framework and discusses findings.

2.2.2 Descriptive Studies

Piazza (1983) estimated China's per capita daily nutrient availability in energy, fat and protein for the period of 1950 to 1982 based on yearly food balance sheets (FBS). FBS estimate the quantities of food commodities available for direct human consumption as the differences between domestic supply and all non-human food end uses. Wang et al. (1993) improved upon the Piazza underestimates in lean years and overestimates in bumper harvest years by including food-stock changes in the FBS. Both studies show rapid nutritional improvements. The total per capital daily energy availability rose from 1,614 Kcal in 1950 to 2,526 Kcal in 1981 (Piazza 1983); and the counterparts estimated in Wang et al. (1993) are 1,752 Kcal in 1950 to 2,506 Kcal in

1981. The level of per capita daily energy availability peaked in 1984 at 2,863 Kcal and started to decrease to 2,652 in 1991 (Wang et al. 1993). In addition to energy level increases, the percentage shares of energy and protein from animal products more than doubled during 1979-1991. Percent energy from animal product increased from 5.1% in 1979 to 11.6% in 1991 and animal protein increased from 8.4% to 20.5%.

Estimating per capita food availability from aggregated balance sheets can be noisy and erroneous. To estimate food available for human consumption, one needs to incorporate measurements of production, trade, use as feed or seeds, waste and stock changes. When much food consumption is for on-farm consumption and does not pass through commercial channels, underreporting of production is a characteristic of most developing reporting agricultural system (Poleman 1981). Consistency checks between food production, consumption and trade data in India and Pakistan indicate that output of major food grain was underestimated by about 25 percent in the 1950s and 1960s (Evenson and Pray 1994). However, in China underreporting may not be the problem due to exaggeration in official reports and there remains much uncertainty on crop waste after the harvest and the magnitude of grain stocks (Smil 1981). Overestimates may result by assuming low waste, minimal animal feed and high grain milling rates. In addition, average per capita nutrient availability derived from an aggregate macroeconomic balance sheet provides general trends in consumption but is not opportune for studies at microeconomic level.

Availability is not intake either (Dowler and Seo 1985). Typically after constructing food balance available for consumption and converting food quantities into nutrient content using standard food composition tables, the average per person availability is calculated where each person is weighted differently according to age and gender. The weights may "have little to do with the metabolic processes involved." (Srinivasan 1992). Accurately measuring nutrient intakes is complex and difficult. Household food inventory method for evaluating household food intakes and individ-

ual 24-hour recall method are becoming increasingly available in developing countries. The household inventory method suffers from several potential biases as food availability in the national level, and leakages (wastage, food prepared for guests and given away, and meals eaten away from home) are not captured or systematically related to income (Strauss and Thomas 1998). Substantial intra-household variation is not taken into account either (Behrman and Deolalikar 1988). The 24-hour recall method corrects the above errors but may suffer from recall errors. Srinivasan (1994) does not think the 24-hour recall estimates could be termed “habitual or long-term intakes.” Bouis (1994) compared food demand patterns from food expenditures (availability) and 24-hour recall surveys (intakes) in Kenya and Philippine and found that the relationship between calorie and income could be overstated using the former method.

The CHNS is one of the rare household surveys that use both individual 24-hour recall and household inventory methods over three random consecutive days to estimate nutrient consumptions.³ The 1991 Food Composition Table (FCT 1991) for China was utilized to calculate nutrient values for the dietary data.⁴ The fundamental

³As documented by the CHNS research team: all adults aged 20 to 45 in 1989 and all household members in 1991 and subsequent surveys provided individual data on dietary intake, with children’s intakes reported by mothers. The 3 consecutive days during which detailed household food consumption data were collected were randomly allocated from Monday to Sunday and are almost equally balanced across the 7 days of the week for each sampling unit. Household food consumption was determined by examining changes in inventory from the beginning to the end of each day, in combination with a weighing and measurement technique. Chinese balances with a maximum limit of 15 kilograms and a minimum of 20 grams were used. All processed foods (including edible oils and salt) remaining after the last meal before initiation of the survey were weighed and recorded. All purchases, home production, and processed snack foods were recorded. Whenever foods were brought into the household unit, they were weighed and preparation waste (e.g., spoiled rice, discarded cooked meals fed to pets or animals) was estimated when weighing was not possible. At the end of the survey, all remaining foods were again weighed and recorded. The number of household members and visitors were recorded at each meal.

Each individual’s average daily dietary intake, calculated from the household survey, was compared with his or her dietary intake based on 24-hour recall data. Where significant discrepancies were found, the household and the individual in question were revisited and asked about their food consumption in order to resolve these discrepancies.

All field workers were trained nutritionists who are otherwise professionally engaged in nutrition work in their own counties and who have participated in other national surveys. Almost all interviewers were graduates of post-secondary schools; many had four-year degrees. In addition, 3 days of specific training in the collection of dietary data were provided for this survey.

⁴This FCT represents a significant advance over the earlier China FCT both for higher quality chemical analyses and for improved techniques of developing average nutrient values for foods whose

assumption of such a calculation is that the nutrient content of a specific food is approximately constant. Such an assumption may be acceptable for certain foods, such as carrots, but is not without contention for others, such as selenium (Willett 1998). Uncertainty regarding the constancy of the nutrient content of food can be and has been checked in nutritional studies but can not be avoided in household surveys whether the inventory method or the 24-hour recall method is used.

Zhai et al. (1996) compares the individual 24-hour recall method and the household inventory weighing method using the 1991 CHNS data for all individuals including adults and children. They separate the sample into three groups of households – those with guests for meals, those with away from home food consumptions, and those with neither. The 1991 Chinese FCT consisting of raw food items and their nutrient contents was used and a procedure to modify the 24-hour recall measure to include edible oils and other condiments was developed. With the modification for oils and condiments, the authors found only a 74 kcal difference between the two methods for daily calorie intake for a reference man weighing 65 kg, age 18-45 and undertaking light physical activities, with household inventory method having higher estimates. The relative differences were larger for protein (3%) and fat (5%). For households with guests and away from home consumptions the ratios are 3-5% for protein and 4-12% for fat (household inventory method gave higher estimates in both outcomes). Without the modification the difference is much larger. The study shows considerable agreement of the two methods and the importance of allocating household cooking oil used to each individual. Not surprisingly, the household inventory and modified 24-hour recall measures of energy intake in 1991 were lower than that in Wang et al. (1993), and the weighing inventory method estimation of fat intake is much higher (62 vs 54 per person per day). As acknowledged by the authors, allocation of oil among household members was based on weights defined by the Chinese

nutrient value varies over the country in a geographic context. The UNC group has worked with the Institute of Nutrition and Food Hygiene to update and improve this FCT.

recommended dietary allowances for energy, rather than based on each individual's proportion of the total household food consumption. An effort was made by the research group in later studies to allocate total oil consumption to all members based on their individual consumption (Guo, Popkin, Mroz and Zhai 1999), although they did not report detailed comparison of the results from the two modification methods.

2.2.3 Income and Nutrients

There is a considerable volume of research in studying the income or expenditure elasticities of calories (Behrman and Deolalikar 1988, Bouis 1994, Strauss and Thomas 1995, Subramanian and Deaton 1996, and Behrman et al. 1997) and less so for protein and other macronutrients. A number of studies focus on the role of income in household food demand in China (Bhandari 1998 and Guo 1998). This paper is not aimed at establishing such relationships for two reasons. Income measures in household surveys are fraught with measurement errors which result in attenuation bias in estimating income elasticities. Secondly, when unobserved factors affecting nutrient intakes (such as body metabolism rates) also correlate with income levels (such as through piece rate income determined by body strength) then income should be treated as endogenous. Otherwise there will be endogeneity bias. However proper instruments dealing with both problems are extremely difficult to find and verify. For completeness of documentation we review a few of such studies for China plagued with one or two of the problems and studies in which researchers strive for a solution albeit imperfect. For other countries in nutrient and income relationships see comprehensive reviews in Behrman and Deolalikar (1988) and Strauss and Thomas (1998) and references within.

Using a sample of adults 18 to 50 years of age from the 1991 CHNS data Ma and Popkin (1995) analyzes three nutrients: fat intake, calories and percentage of calories from fat. The authors considered income, age, sex, education, smoking,

drinking alcohol, intensity of labor activity, family size and region of residence as explanatory variables. The authors also tried including height, weight, number of children under the age of seven years, and job types and excluded them based on 10% statistical significance cutoff. The authors estimated three statistical models: a linear relationship between income and outcomes controlling for other covariates, a model with linear and interactions of income and other covariates, and a switching model regressing all other covariates on outcomes based on different income cut-offs in defining high vs low income families. In the first model the authors found a positive income effect for fat intakes and percent calorie from fat but a negative income effect for calorie intake. In the interaction model the interaction terms are not significant. In the switching model, where the sample is broken at 85% of income level, the two regimens, high vs low income families, have significantly different coefficients for all covariates and the income effects of fat and calorie intakes have different signs for high vs low income families. The results of this paper is hard to interpret. First, it includes choice variables that are clearly part of the biological nutrition production process, such as smoking, drinking alcohol, intensity of labor activity. These variables should be considered endogenous. The study is thus not a reduced form demand function study and it is not treating factors in a production function as endogenous either. Secondly, the switching point, i.e., the income cutoff point used to define high vs low income families, was found using a trial and error approach based on statistical significant differences on coefficients; and the model did not control for potential self-selection bias. Despite the painstaking effort and the correct envision of the nonlinear income effect and behavioral differentials between the higher and lower income families, all main parameters of interest are biased and the direction of the bias is unknown.

In Guo et al. (2000) the researchers present evidence of remarkable shifts in how Chinese diets varied with income over 1989 and 1993, and are interested in finding

the nonlinear effects of income on six groups of key food consumption (rice, wheat flour, coarse grains, pork, eggs and edible oils), and dietary fat intake. Using the 1989, 1991 and 1993 waves of the CHNS data for adults aged 20 to 45, a two-stage estimation method was to correct for measurement error and endogeneity bias of household income, while also controlling for a set of food prices, age, gender, education level, household size, place of residence and region. The authors only mentioned that the instruments in the first stage for per capita household income included “community information, family background variables, and household business and asset measures” but did not give specific variables. The authors did not provide statistical testing on the validity of the instruments they used. Individual level random-effect models were used for each key food and for dietary fat intake to control for within household unobservables. The model estimated nonlinear income effects through income, income square, income and year interaction, and income square and year interaction and the authors presented variations in elasticities through graphs. The major findings are: income effects for low-fat, high-fiber foods (wheat flour products, rice and coarse grain) fell from 1989 to 1993. Higher-fat foods such as pork, edible oils and eggs became more responsive to income levels. The quantity of fat in the diets increased significantly and seemed to be rising rapidly with increases in income. The authors believe these changes are “an important deterioration in the healthiness of the Chinese diets that could burgeon as the Chinese economy continues its expansion.” This paper is the only one studying Chinese food consumption treating income as endogenous and the results are striking. Simple income changes alone can not explain the complex shifts in Chinese adult dietary patterns in the late 1980s to the mid-1990s.

For the elderly in China, the CHNS 91 and 93 were used (Stookey et al. 2000) to estimate the determinants of nutrition intakes. The dependent variables included energy, fat, protein intakes, consumption of rice, high-fat red meat, eggs, and plant

oils. The explanatory variables were age groups, sex, income tertiles, and rural-urban residence. Increasing income was significantly associated with greater energy, fat and protein intakes. Rice consumption declined with older age and consumption of plant oils, high-fat red meat and eggs increased by income tertiles.

There is a list of studies focusing on the relationship between income and dietary transitions (Guo, Popkin and Zhai 1999 and many similar studies in the references) that generally suffer from the same biases as the above two examples. Generally these studies point to the relationship that higher income levels, particularly in urban areas, are associated with consumption of a diet higher in fat.

Because the authors in Guo et al. (2000) were mostly interested in estimating income elasticities of various foods and fat intake they did not present results on any other socioeconomic determinants. These factors are of interest in our analysis. In this paper, instead of studying the income effect of nutrient intakes, at individual and household levels we focus on the effects of education and productive assets.

2.2.4 Price and Nutrients

As is seen in equation (2.1) food consumption is a function of prices of all goods. For farm households in the special case in which household and farm-allocative decisions are separable, production inputs prices affect household consumption only via profits and can be replaced by farm profits or full income. Strauss (1984) shows the own-price effects on calories consumption in Sierra Leone are negative when profits are allowed to vary, although they could also be positive through income effect for net sellers. Pitt and Rosenzweig (1985) estimates food price effects on aggregate household consumption of foods converted into nine nutrients based on a probability sample of 2,847 farm households in Indonesia. Only one of the four prices was found to be unambiguously and negatively associated with the aggregate consumption of all nutrients. Knowledge of how changes in food prices affect individual level nutrient

intakes is crucial to effective policy making. Here we review one paper on the effects of prices on macronutrients in China (chapter 7 in Guo 1998).

Using the 1989, 1991 and 1993 samples of CHNS of adults aged 20-45 years, Guo, Popkin, Mroz and Zhai (1999) studies the own- and cross-effects of six prices (rice, wheat flour, coarse grain, pork, eggs and edible oils) on the consumption of six food groups and on overall energy, protein, fat intakes, the proportion of the recommended daily allowances for energy and protein, and the percentage of calories from fat. The sample is stratified by income and the price elasticities were estimated in two parts – the probability of consuming certain food items and the quantity consumed given it is nonzero, similar to Pitt (1983) in an effort to incorporating corner solution at zero consumption. Age, gender, education, household size, urban residence and regions are included as control variables in addition to prices and income. Free market prices were measured at the community level and when the goods were not sold in the free market prices from the ration system were used. The measurement of intakes is modified on account of the cooking oils based on proportion of individual meat and vegetable consumption in total household consumption. In both the model for the likelihood of consuming specific foods and the model for the quantity consumed, the same set of socio-demographic factors are used. The authors found the likelihood of consuming edible oils and rice would decrease by 16 to 20 percent respectively for 10% increase in real prices of edible oils and rice and price changes for animal protein foods had a large effect on reducing fat intake. Increases in the prices of pork, eggs and edible oils are predicted to lower fat intake. Only increases in pork prices led to reduced protein intakes. Besides the stratification of samples based on endogenous incomes and inclusion of household size as exogenous control variable, there may be severe bias due to omitted variables at the community level. Besides prices the authors did not include other community characteristics such as water and sanitation factors. Aggregate policies, not only on prices but also on program such as

food subsidies, could also affect food consumption. Ignoring these factors produces downward bias in standard errors of parameters for prices and unsigned biases in the parameters themselves. When time-invariant factors are not of primary interest and more than one year of data are available, the fixed-effect model will remove the time-persistent unobserved heterogeneity bias. We include community dummy variables representing unobserved heterogeneity and find the effects of several food prices changed signs. Since prices were measured at community level, not estimated from household unit costs, the issue of endogeneity of prices due to quality selection is avoided (Deaton 1988), although the prices could be measured with error, especially due to substituting free market prices with ration prices.

In our study, specific foods or food groups are not the focus. However, better understanding of the impacts of food prices on total calories and the composition of energy intakes has important policy implications. It provides guidance toward shaping the healthiness of possible targeting populations and reducing the risk factors related to quality of diets.

2.2.5 Education and Nutrients

The interpretation of the effect of education on labor market outcomes is debated over how much of it actually represents the benefit of human capital accumulation and how much from unobserved individual (ability) and family background (for a comprehensive review see Strauss and Thomas (1995)). Similar arguments can be made in studying the effect of education on nutrient intakes. Improved education raises the allocative efficiency of food consumption, especially at times of technological change. Higher education is also associated with increased income and investment in human capital. In addition, schooling can affect preference or tastes for food composition, although the effect may be small. More recently the causal effect of education on earnings has been estimated through supply-side institutional features such as com-

pulsory schooling laws, differences in distance to schools as instrumental variables and it is believed that such IV estimates reveals underlying heterogeneity in the returns to human capital investments (Card 2001). In a reduced-form demand analysis we are not concerned with estimating the causal effect of education and can not distinguish the different interpretations of the effects of education.

Bhandari (1998) estimated the association between education and food consumption in China using the 1991 CHNS. The author studied the likelihood of consuming 22 food groups and the effects of male and female household head education on macronutrient intakes of calorie, protein and fat. The explanatory variables considered exogenous were income, age, gender, urban and province of residence, physical activity levels, and five interactions. There was no significant effect of either male or female education on caloric intake, a significantly negative effect of male education on protein intake and a positive female education effect on fat intake. Since all of the above mentioned limitations apply for this study, the results are to be interpreted with caution. Another disadvantage of the study is the author used only one wave of the data in which the effects of community year interactions and the changes in prices, water and sanitation conditions can not be effectively controlled.

Since education partially reflects household resources it is necessary in our basic and augmented analyses to control for household resources and local infrastructures.

2.3 Data and Econometric Issues

2.3.1 The Data

Three rounds of the CHNS in 1989, 1991 and 1993 are used in this study. The focus is on adults at or above 20 years of age and those who were in the study of determinants of BMI in an earlier paper. For detail of the sample derivation see Luo (2003*b*). The 1997 nutrition data are not publicly available yet. For some individual,

household and community characteristics see Table 1. The percentage of women with no formal education is much higher than the percent of men (about 30% vs 10%). Since in 1989 our sample includes only those who were less than 46 years old and had higher average education level than the average education of adults of all ages, when comparing the trends in education level we look at the year from 1991 to 97 (Luo 2003*b*). The average years of education increased by about half a year from 1991 to 1997. The overall prevalence of people with no formal education decreased from 22 to 18 percent. Between 91 and 97 the percent of adults with less than primary degree decreased and the percentages of all other higher levels of education increased. There is some gender differential in education levels between men and women. Men had 1.7 years more of education than women in 1991, 2 years more in 1993 and 2.1 years more in 1997. The percent of men in lower education levels decreased (no formal education by 3%, some primary schooling by 3%, and primary degree by 1%), and the percent of men in higher education levels increased (for middle school degrees by 3%, high school by 1%, and college and above education by 2%), whereas for women the increase in education happened in all levels and there was a significant decrease of percent of women with no formal education.

A person is employed if she is a regular worker, contract worker, or a temporary worker in state and collective enterprises and three-source invested enterprises.⁵ The employment rates of men are also higher than that of women. This period marks China's most noteworthy economic growth and expansion. There is a significant increase in household real productive assets between 1989 and 1993, all values discounted to the 1988 yuan. The gap between rural and urban productive assets is closing, with urban residents owning less such assets. There is little change in the amount of household cultivated land.

⁵It also includes those employed by individual private businesses, and the self-employed in various kinds of household sideline productions, small retail businesses, handicrafts, etc. These can be paid or unpaid occupations. Farmers engaging in agricultural labor and soldiers are considered as having a job, but students' vacation work and housewives' house work are not considered as being jobs.

The community survey section includes information on infrastructure, services, populations in the village/neighborhood, percentage of land with poor quality, daily wages for unskilled farmers and construction workers, percent of work force engaged in agriculture/working out of town for more than one month, and hospital and clinic infrastructure and personnel. In each community, state ration coupon, retail and free market prices were collected in stores for the most commonly consumed items of rice, wheat, egg, pork, beef and fish. Since free market prices reflect the value of each produce better, they are included in the reduced-form analysis with community characteristics. As is seen in Table 2.1, the prices of the six food groups in 1989 seem to be too high,⁶ causing a dip in the generally increasing trends. This is possibly because the price data were “cleaned” by the CHNS team before being put on the web site.⁷ In our paper whenever certain foods are not usually consumed in some community the relative median prices between two similar food groups are used to help impute the missing values. For example, when the rice price in one urban community is missing it is imputed by the product of the ratio of provincial urban median prices of rice and wheat and its own price of wheat. There are only a small number of communities with such imputed price information.⁸

The types and sources of households water sources, toilet facilities and sanitation (excreta) condition around the living areas are aggregated from the household level to derive the percent of households within each community having certain resources. For example, in Table 1 the average of percent of households in each community drawing

⁶According to Ge et al. (1992) in May 1991, the price of grain and edible oil on ration was readjusted by a large margin for the first time since the mid-1960s. The price of grain was raised by 70 percent and the price of edible oil almost doubled. Although to lessen the impact of the rise in costs on the population's living standard, the government provided a subsidy to urban workers, the free market prices in 1991 were expected to be much higher than that in 1989.

⁷This is possibly done by replacing missing values with the mean of the non-missing data as was done when the CHNS research team was cleaning the income data. This imputation was not done for other years (because the original data from other years included the coding for missing values such as -999 but no such codes exist for 1989 data) and thus resulted in the high 1989 prices.

⁸For example, in 1991, free market prices of rice in 20 out of 189 communities are replaced this way, and only one community's free market prices of pork was missing and replaced this way via ratio of prices of pork and beef.

water from an open well in 1989 is 17% and it decreased to 12% in 1993.

The three consecutive days during which detailed household food consumption data were collected were randomly allocated from Monday to Sunday and are almost equally balanced across the 7 days of the week for each sampling unit. Individual dietary intake for the same 3 consecutive days was surveyed for all children aged 1 to 6 and all adults aged 20 to 45 in 1989, and for all individuals in later years.

The 1991 Food Composition Table for China was utilized to calculate nutrient values for the dietary data. This FCT represents a significant advance over the earlier China FCT both for higher quality chemical analyses and for improved techniques of developing average nutrient values for foods whose nutrient value varies over the country in a geographic context.

The outcomes of interest are calorie, fat, protein and carbohydrate intakes per person per day and percent of calorie from fat, protein and carbohydrates. The 1991 Chinese Food Composition Table (FCT) was used to convert all 22 categories of foods into macronutrients. The number of different foods in the FCT is 636 (for a list of all foods see Bhandari 1998). The individual 24-hour recall measure of intakes from all food groups is used without modification for household cooking oils (Table 2.2). The energy content of fat, protein and carbohydrates is calculated based on the formulae that one gram of fat yield 9 kcal of energy; one gram of protein and carbohydrates each yield 4 kcal of energy (Whitney and Rolfes 1996). Compared with results shown in Zhai et al. (1996) for a "reference man" our estimate of average daily calorie and protein intakes in 1991 for all men (2,795 g and 102 g respectively) are actually higher than the modified measure (2,351 and 70 g) and the household inventory measure (2,425 and 71 g). However, the fat intake for men is much lower (52 g) than the household inventory measure (62 g) and a little lower than the modified (54 g) measure. Compared to the average per capita daily nutrient availability measures in Wang et al. (1993) the calorie (2,652 g) and fat (54 g) measures are close to our

estimates in 1991, but the protein intake (70 g) is closer to that in Zhai et al. (1996).

2.3.2 Measurement Errors

The modification of cooking oil and condiments affect the estimates of fat intake significantly. Hence our estimates of fat intake serve as a lower bound for the actual fat intake. The other measurement problems might include the following: (1) the non-edible part of each food is assumed to be excluded in the recall measure; (2) wastes in food are not included; and (3) recall bias may not be random. If low-income households waste less than high-income households then nutrient intakes will be measured with upward bias and the bias is higher for higher income groups. However, if higher educated people consume more processed food or have better technology in preparing food then the bias may go the other way around. If recall bias is systematically related with gender, age, income or education as in cases that women, younger people, and/or persons with higher income or higher education report intakes more accurately the biases can be upward or downward.

One advantage of using the 24-hour recall measure is that away-from-home consumption is included and the downward bias in the weighing method is avoided. If people don't know ingredients of cooked foods eaten away from home, the measurement error will increase.

Measurement errors in the dependent variables can be included as one composite error terms in the regression analysis. But if any of the above systematic errors are correlated with explanatory variables the results will be biased. Even when the measurement errors in the dependent variables are uncorrelated with the regressors and resulting in unbiased estimates of coefficients, the estimated standard errors will be larger than the standard errors from models without measurement errors. In such cases statistically significant relationships are harder to establish. To minimize

Table 2.1: Individual, Household and Community Characteristics in CHNS 89-93

Male	Individual			Overall	Household		
	1989	1991	1993		1989	1991	1993
N	2074	3413	3208		N	2617	3356
Age	31.85	42.27	43.26	Real Prod Assets (yuan)	239.53	270.14	426.16
Urban	0.31	0.35	0.32	Land (mu)	2.47	2.50	2.66
No Education	0.04	0.11	0.10	Rural Areas	1989	1991	1993
Some Primary	0.13	0.17	0.24		1801	2248	2152
Primary Degree	0.22	0.22	0.18	Real Prod Assets (yuan)	288.84	330.64	451.19
Middle School	0.37	0.30	0.31	Land (mu)	3.49	3.64	3.78
High School	0.16	0.12	0.11	Urban Areas	1989	1991	1993
Tech/College+	0.09	0.07	0.06		815	1108	969
Employed	0.98	0.86	0.84	Real Prod Assets (yuan)	125.29	148.17	383.50
				Land (mu)	0.21	0.22	0.20
					Community		
					1989	1991	1993
Female	Individual						
	1989	1991	1993				
N	2242	3752	3476		N	180	189
Age	31.82	41.80	43.40	Free Market Price of Rice	1.55	1.07	1.25
Urban	0.31	0.36	0.32	Wheat	1.38	1.12	1.31
No Education	0.16	0.32	0.31	Pork	5.74	5.65	7.10
Some Primary	0.17	0.15	0.22	Beef	7.57	7.17	9.25
Primary Degree	0.20	0.17	0.13	Edible Oil	5.73	5.11	5.98
Middle School	0.29	0.22	0.22	Eggs	5.98	3.95	4.50
High School	0.12	0.09	0.08	Water Source Open well	0.17	0.14	0.12
Tech/College+	0.06	0.04	0.04	River, etc	0.08	0.07	0.07
Employed	0.94	0.78	0.76	Water Factory	0.35	0.41	0.42

Note: There are 18,165 observations across all three years (4,316 in 1989; 7,165 in 1991 and 6,684 in 1993). In 1989 individuals over 45 years of age are excluded for they were not interviewed for dietary information. Real productive assets are discounted to 1988 value. The items included are tricycles, motorcycles, tractors or walking tractors, irrigation equipment, power threshers and water pumps. Values are self-reported purchasing values in 1989 and current worths in all other years.

Table 2.2: Average Daily Calorie, Fat, Protein and Carbohydrates Intakes: Patterns and Trends in CHNS 89-93

Male	Calorie (Kcal)			Fat (g)			Protein (g)			Carbohydrates (g)		
	1989	1991	1993	1989	1991	1993	1989	1991	1993	1989	1991	1993
Overall	2816.94	2794.99	2688.56	47.65	51.76	53.86	101.18	102.06	100.44	482.57	462.36	434.09
Urban	2598.40	2577.65	2500.23	55.43	59.57	63.10	101.84	103.34	101.92	410.51	392.87	368.06
Rural	2912.94	2911.93	2775.72	44.23	47.56	49.58	100.89	101.37	99.75	514.23	499.75	464.66
Age 20-39	2815.52	2913.59	2802.60	47.57	53.17	55.92	101.75	106.31	104.92	482.25	486.19	457.07
Age 40-59	2823.10	2811.20	2727.54	47.99	51.62	53.30	98.74	100.86	101.26	483.99	464.77	441.14
Age 60+		2394.46	2322.74		47.71	49.88		91.55	87.54		383.71	362.05
No Education	3133.97	2618.60	2602.30	41.56	40.86	42.34	108.18	91.27	89.60	576.47	460.10	449.20
Some Primary	3001.04	2786.70	2657.82	43.22	47.62	51.52	101.46	97.74	96.40	538.34	471.13	431.77
Primary Degree	2846.73	2847.16	2726.77	45.95	50.35	49.14	98.53	101.72	98.64	493.32	479.15	455.86
Middle School	2808.16	2854.06	2735.91	47.12	52.29	57.46	101.46	104.48	104.52	483.29	471.68	437.13
High School	2721.10	2801.24	2720.45	52.28	59.36	59.85	102.69	107.12	107.11	444.89	442.88	421.43
Tech/College+	2539.20	2671.56	2536.19	55.15	67.47	66.58	100.70	111.26	106.21	396.16	388.52	361.26

Female	Calorie (Kcal)			Fat (g)			Protein (g)			Carbohydrates (g)		
	1989	1991	1993	1989	1991	1993	1989	1991	1993	1989	1991	1993
Overall	2441.09	2403.25	2313.70	40.57	44.31	45.87	88.94	89.48	88.23	427.91	408.72	385.06
Urban	2214.54	2214.52	2149.33	46.41	51.63	54.51	89.26	90.22	89.52	357.74	343.44	323.73
Rural	2542.66	2507.13	2390.81	37.95	40.28	41.81	88.79	89.07	87.62	459.37	444.65	413.84
Age 20-39	2439.61	2509.81	2405.95	41.36	46.52	47.50	89.51	93.41	92.01	425.49	426.99	400.89
Age 40-59	2447.86	2431.95	2349.42	36.98	43.58	45.99	86.34	89.05	88.67	438.93	417.21	393.03
Age 60+		1977.00	1974.98		38.39	41.06		77.11	76.73		327.66	322.68
No Education	2687.51	2344.65	2275.60	33.95	37.54	37.27	89.25	83.38	81.92	503.91	415.31	400.99
Some Primary	2475.20	2467.40	2378.96	32.12	41.43	45.14	84.13	88.59	88.36	460.59	431.49	403.12
Primary Degree	2510.96	2532.06	2302.78	42.48	46.55	46.28	91.34	93.70	89.10	437.60	431.61	380.06
Middle School	2360.76	2406.16	2330.88	43.72	47.66	50.51	89.08	92.78	91.92	401.04	399.26	375.39
High School	2290.17	2307.26	2322.62	46.98	53.88	59.72	91.13	92.11	97.07	373.93	360.67	347.69
Tech/College+	2140.36	2266.48	2156.97	48.25	58.75	62.45	88.93	98.54	95.62	335.67	332.41	301.37

Note: The daily intakes of energy, fat, protein and carbohydrates are based on the 1991 China Food Composition Table (FCT). All food items are included. The major food categories are: alcoholic beverages, cereals, condiments, dairy products, egg and egg products, fish, fresh beans, fruits, leafy vegetables, legumes, meats, melons, milk substitute and infant foods, mushrooms, nightshades, nuts and seeds, shrimp and crab, pickled vegetable, roots and stems, poultry, and shellfish. For the number of observations within each cell, median distributions and standard deviations for the mean see appendix A.

random measurement errors, daily nutrient intakes are averages over three consecutive days whenever available.

Random measurement errors in the explanatory variables cause the estimates to be biased toward zero. Incomes in developing countries have been found to be particularly error-ridden. That is one of the reasons we prefer not to use income as a main regressor. Instead we use real productive assets and land farmed in calibrating household resources.⁹

2.3.3 Patterns and Trends

Tables 2.2 and 2.3 report the general patterns and trends of average outcomes. For the number of observations, median and standard errors of the mean in each stratum see Appendix E. In Table 2.2 there is a decreasing trend in caloric intake for men in urban and rural areas (98 kcal and 137 kcal decrease respectively) and women (66 kcal in urban and 152 kcal decrease in rural areas) from 1989 to 1993. There is an increase in calories in 1991 for more educated men and women and for younger adults less than 40 years of age, and it is followed by a decrease across all subgroups in 1993.

For men there is a general increasing trend in fat intake (in grams) except for a relatively small dip for the non-educated and highest educated groups. For women the increase is more pronounced, with the highly educated (high school diploma or better) seeing the biggest increase. These women are also those with largest increases in BMI (1). As some studies show a doubling of edible oil in this period (Guo, Popkin, Mroz and Zhai 1999), the absolute level of fat intakes should be higher than those reported in table 2. Our estimates provide a lower bound for the level and trends of

⁹Productive assets may be estimated with errors too. The items included are tricycles, motorcycles, tractors or walking tractors, irrigation equipment, power threshers and water pumps. Real productive assets are discounted to 1988 value. Values are self-reported purchasing values in 1989 and current worths in all other years. There are items excluded from this calculation because they are not surveyed in all waves of the CHNS.

fat intake. Overall protein intakes have been stable, but there is a decreasing trend among the least educated and an increasing trend among the highly educated. It is evident that education is associated with better quality of nutrients but whether it is through increased human capital or through shifted tastes is unknown. There is cross the board decrease in carbohydrate intake. Simple carbohydrates are proxies for lower quality diet.

In Table 2.3, columns 1 to 3, there is an increase in the percent of calories from fat for men and women in all subgroups. Column 4 to 6 show increasing trends in percent of calories from protein except for least educated men. There is a steady decrease in percent calories from carbohydrates in all subpopulation strata. Alongside economic development we can see decreased carbohydrates and increased fat intakes as signs of closing the gap in eating patterns between China and developed countries. The departure from the traditional low-fat diet in China has also been viewed as a sign of deteriorating consumption pattern (Popkin 1999, Guo, Popkin, Mroz and Zhai 1999).

Urban residents consume less energy and carbohydrates, but more fat in all years (Table 2.2). Calorie and carbohydrate intakes decline as education levels go up, and for fat the reverse is generally true. There is no significant difference in protein intake between regions. There is a increase in protein intakes among higher educated people if the education level is less finely cut at the high end of the distribution. As people age there is an decrease in all nutrient intakes. The relationships between percent calorie intake from fat, protein and carbohydrates and residence, aging and education levels are more stable the levels of intakes. Although fat intake decreases as people grow old the percent of calorie from fat increases for men. The percents of calorie from fat and protein for urban residents and higher educated men and women are higher than those of rural residents and less educated.

Table 2.3: Average Percent of Calorie from Fat, Protein and Carbohydrates: Patterns and Trends in CHNS 89-93

Male	% Fat			% Protein			% Carbohydrates		
	1989	1991	1993	1989	1991	1993	1989	1991	1993
Overall	14.61	16.15	17.45	14.37	14.70	15.04	69.42	66.97	65.38
Urban	18.27	20.01	21.86	15.63	16.06	16.33	64.49	61.98	59.93
Rural	13.00	14.08	15.41	13.82	13.98	14.44	71.59	69.65	67.91
Age 20-39	14.61	15.88	17.43	14.45	14.69	15.04	69.38	67.57	65.97
Age 40-59	14.61	16.02	17.04	14.03	14.46	14.98	69.58	66.95	65.50
Age 60+		17.25	18.36		15.28	15.17		65.16	63.69
No Education	11.26	13.66	14.18	13.65	14.07	13.90	74.54	70.87	69.70
Some Primary	12.74	14.85	16.84	13.56	14.16	14.60	72.16	68.53	65.95
Primary Degree	13.60	15.50	15.66	13.75	14.39	14.58	70.76	67.98	67.78
Middle School	14.41	16.03	18.34	14.47	14.74	15.33	69.70	66.87	64.57
High School	16.48	18.26	19.37	15.06	15.32	15.91	66.63	64.38	62.62
Tech/College+	18.93	21.96	22.45	15.80	16.77	16.88	63.53	59.03	58.23
Female	% Fat			% Protein			% Carbohydrates		
	1989	1991	1993	1989	1991	1993	1989	1991	1993
Overall	14.39	16.01	17.28	14.57	14.98	14.60	70.70	68.52	67.06
Urban	18.09	20.02	21.92	16.02	16.33	16.70	65.50	62.99	61.11
Rural	12.74	13.80	15.10	13.92	14.24	14.67	73.03	71.57	69.85
Age 20-39	14.69	16.10	17.34	14.66	14.98	15.39	70.36	68.54	66.99
Age 40-59	13.07	15.55	16.96	14.17	14.72	15.15	72.24	69.17	67.53
Age 60+		16.72	17.87		15.52	15.50		67.03	66.18
No Education	10.47	13.94	14.32	13.20	14.37	14.44	75.98	71.13	70.83
Some Primary	11.48	14.67	16.63	13.70	14.45	14.94	74.54	70.36	68.13
Primary Degree	14.31	15.81	17.50	14.47	14.81	15.50	70.74	68.93	66.59
Middle School	16.10	17.19	18.95	15.07	15.47	15.84	68.55	66.96	64.92
High School	17.30	20.18	21.89	15.68	16.03	16.85	66.74	63.28	61.00
Tech/College+	19.53	22.34	24.97	16.46	17.32	17.79	63.60	59.75	56.92

Note: The energy content of fat, protein and carbohydrates is calculated based on the formulae that one gram of fat yields 9 kcal of energy; one gram of protein and carbohydrates each yields 4 kcal of energy. All food items are included. For number of observations, median distribution and standard deviation of the mean see appendix A.

2.3.4 Estimation Methods

As was discussed in Chapter 1, the estimation of equation (2.1) for the reduced-form macronutrient intakes uses OLS on data pooled over three years. To account for heteroscedasticity and correlation of errors for each individual over time individual-level robust standard errors are used in all regressions. In all regressions individual-year age dummies and five-year cohort dummies are included. In a separate paper (Luo (2003a)) we consider the identification problem in estimating age, cohort and year effects in a model. We chose the five-year cohort dummies so that the perfect linear dependency of the three variables no longer exists and the system can be identified. Surely there are many other identification strategies as noted in Luo (2003a) but we chose five-year cohorts around the early 1960s' famine and the cultural revolution period as being one meaningful strategy. Individual-year age dummies are preferred to using age as a continuous variable because the dummies capture the nonlinear effects of age better.

Modelling Levels

Two sets of regressors are of interest at the individual and household levels and at community levels. First, the basic specification focuses on the effects of individual education and household covariates - productive assets and land, controlling for community level factors by community-year interaction dummies. Education is categorized as no formal education (the reference group), some primary education, primary degree, middle school degree, high school degree, and technical institute or college level or above. We control household resources using log of real productive assets discounted to 1988 value as indicators for long-run resources. Splines around the median of the positive values of real productive assets and the amount of land cultivated (mu) are included in all regressions. The coefficients for the lower and upper spline of productive assets represent the level effects, not the marginal effects.

This basic model is estimated for the whole sample, urban and rural areas separately, and for each age group (20-39, 40-59 and 60+) for men and women calorie (Table 2.4), fat (Table 2.5), protein intake (Table 2.6) and percent calorie from fat, protein and carbohydrates (Tables 2.7 to 2.9). These stratifications enable us to find differences between different groups of people. The rural-urban stratification is consistent with the presumption that rural households produce and consume at least part of their produce and urban households are primarily consumers. The community dummies capture the effects of unobserved time-invariant community level information. For factors such as prices, time-invariance seems unlikely. Using community-year interaction dummies allow the effects to differ by year, as prices, future price expectations and other community factors change over time.

Secondly, we add to the basic model community characteristics such as prices, water, toilet and sanitation conditions replacing community-time interaction dummies. The augmented model studies the joint effects of individual, household and community variables on all outcomes (Tables 2.10 to 2.12). In this specification we use community dummies instead of community-year interaction dummies to capture community level unobserved heterogeneity. It is important to control for such heterogeneity because infrastructure, such as the quality of water and sanitation conditions, is likely to depend on community resources and may be non-randomly placed according to unobserved community attributes (Pitt et al. 1993).

Modelling Changes

Estimating equation (2.1) with pooled OLS implies the effects of all socioeconomic determinants are fixed over time, whereas the theory tells us we should, for example, include all past and future prices in the model to allow different effects of these variables. The constant effect assumption can be easily violated during the time of rapid transition in China. Finding out the socioeconomic determinants of the

changes of people's food consumption has important policy implication. Exploring the longitudinal nature of the data set, for all outcomes, the changes between 1989 and 1991 and the changes between 1991 and 1993 are pooled together and the OLS method was used in the basic and augmented models again to find out the effects of the changes in prices and other community characteristics on the changes in calorie, fat and protein intakes and the changes in percent of calories from fat, protein and carbohydrates.

When the outcomes are changes two sets of regressors are again of interest. First, the basic model, includes education categories, real productive assets and land cultivated in base year and community dummies. Secondly, the augmented model, adds changes in food prices and community water and sanitation characteristics as explanatory variables in addition to the individual and household characteristics in the basic model. Not only do the baseline levels of prices and community characteristics affect the changes of nutrient intakes, but the changes in prices and other local factors may also have an impact. Conceivably people respond to food price changes by substituting cheaper foods when the prices of substitution foods decrease or complementing and consuming more food when the prices of complement foods decrease.

Modelling Production Function

As outlined in Chapter 1 adequate to the application of BMI, the health status at the beginning of period $t + 1$ is assumed to be influenced by health status at t , health inputs in period t , m_{it} , such as the balance between energy expenditure (basic metabolism, working or exercising) and intakes, medical care, or illness spell, which does not bring utility directly. More leisure time may bring feelings of well-being and sound state of mind, which can be positive influence of one's physique. Individual characteristics, θ_{it} , such as age¹⁰, gender and community or environmental

¹⁰The solution to the maximization problem is a time-invariant policy function that determine the control variable (x_{it}, l_{it}, m_{it}) and the state variable h_{it} . It is conceivable as a person grows old

characteristics, z_{ct} , may also have a direct impact on health outcomes:

$$h_{it+1} = f_i(h_{it}, m_{it}, l_{it}, \theta_{it}, z_{ct}; v_{it}) \quad (2.2)$$

where v_{it} are unobservable individual, household, and community factors that affect member i 's health. It is assumed that $\partial f_i / \partial h_{it} > 0$, and $\partial f_i / \partial m_{it} > 0$. It is necessary that $\partial f_i / \partial h_{it} < 1$. Equation (2.2) is not a general form of health production function in that it assumes that h_{it} is a sufficient statistic that summarizes the effect of all past inputs and choices in periods $1, \dots, t-1$ and there is no direct lagged effects from them on h_{it+1} . A more general health production function would allow all past inputs to be included.

To estimate a health production function several foreseeable problems are considered (Behrman and Deolalikar 1988, Strauss and Thomas 1995, Cebu-Study-Team 1992, Wolpin 1997, Sickles and Taubman 1997). First the choice variables of nutrient intake could be related to the unobserved disturbance term in the health production. The disturbance may include unanticipated health shocks, omitted inputs and individual heterogeneity. The correlation between nutrient intake and the error term suggests the OLS estimates of the equation (2.2) are biased. We measured h by body mass index in Chapter 1 and estimated the reduced-form health demand function. Here simple health production functions, as measured by logarithm of weight in 1993 and logarithm of BMI in 1993, are estimated using lagged weight, height (or lagged BMI), physical activity levels, and current food intakes as inputs with instrumental variables. The identifying instruments used are education, real productive assets, land cultivated and community characteristics including prices, water and sanitation conditions in 1991.

The biological process of producing body weight and BMI is necessarily a balance

the effect of inputs on health (body size) varies. To incorporate such effect we can assume a health production function that varies with age at time t , a_{it} :

$$h_{it} = f_i(h_{it-1}, m_{it}, l_{it}, \theta_{it}, z_{ct}; v_{it}) + g_{it}(a_{it})m_{it}$$

between energy inflows and energy expenditure by metabolic processes and physical activities aimed at maintaining good health and daily living (Srinivasan 1992). In CHNS a set of energy expenditure questions related to occupations that have been asked as part of the nutrition data collection since 1989.¹¹ I categorize the energy expenditure into two groups with light and very light activity as reference comparing to moderate, heavy and very heavy activity levels, and instrument them with the above set of IVs.

There are other inputs such as m_{it}, l_{it} that are omitted from the simple analysis. Omitted variable bias in OLS can be upward or downward depending on the correlations between intakes and the omitted variables if we assume $\partial f_i / \partial m_{it} > 0$ and $\partial f_i / \partial l_{it} > 0$. When genetic endowments of an individual are excluded and are positively related with nutrient intake for better-endowed persons are more likely to utilize more nutrients, the estimated effects of nutrient intake on health are likely to be overstated. The nutrient intakes and physical activities are quite possibly measured with error. The fat intake in this paper serves as lower bound of the true value because we do not modify the oil content of cooked meals. The random measurement error leads to attenuation bias in OLS regressions. All determinants in the reduced-form demand function are used as identifying instruments in the production function analysis. They serve to correct both the omitted variable biases and the attenuation bias, although to which end they could do a better job is unknown.

¹¹Very light physical activity refers to working in a sitting position, e.g., accountant, office worker, electrical appliances repairer, watch repairer. Light physical activity normally refers to working in a standing position, e.g., shop assistant, laboratory technician, teacher. Moderate physical activity, e.g., student, driver, electrician, metal worker, salesman. Heavy physical activity, e.g., farmer, dancer, steel worker, athlete. Very heavy physical activity, e.g., loader, logger, miner, stonecutter.

2.4 Results: Determinants of Levels

2.4.1 Basic Models

The basic results for total calories, fat and protein intakes, and the percent of calories from fat, from protein and from carbohydrates are summarized in Tables 2.4 to 2.9. All analyses are done separately for men and women in the whole sample, from rural and urban areas and for different age groups. Community-year interactions, individual-year age dummies and five-year cohort dummies are also included.

Education does not have significant impacts on calorie intakes but does affect the percent of calories from fat, from protein and from carbohydrates differently in different region and at different ages. The effect of productive assets is nonlinear and has an inverted U-shape for male calorie, fat and protein intakes. In rural areas the effect of productive assets is stronger than that in the urban areas. At different ages the effects of education and assets are different.

Total Calories (Kcal)

Results for daily calorie intake (kcal) of men and women in different strata on the basic specification are shown in Table 2.4. The effects of education for men is not significant in any stratum; while the effect of female education is nonlinear and significant for the whole sample (joint F-test has p-value 0.03) and for the elderly (age 60 and above, p-value=0.02). Overall, as female education increases the level of calorie intakes increases first and then decreases; however for the elderly the reverse is true. For the elderly those with the high school education consume 897 more Kcal of calories than those without any education.

The effect of productive assets is not significant for women but is significant for rural and younger men at 0.01 level and for the elderly at 0.1 level. Below the median productive assets are positively correlated with calorie intake and above the

median the correlation is negative for rural and younger men. For the elderly the relationship is reversed. This may be because for the young and rural men at lower level of productive assets they complement labor input and hence increase calorie intake whereas at higher level of productive assets they substitute for labor inputs and hence decrease calorie demand. For the elderly the joint effect of the splines is significant at 10% level and only the higher level spline is individually significant at 10% level. As a person gets older he may need more productive assets to complement given labor input and hence result in higher demand for calorie.

The effect of land is only positively associated with calorie intakes for elderly females. Community dummies are all jointly significant. The age profiles between men and women for the urban and rural samples are different (p -value=0.004 for urban sample and 0.008 for rural sample). Testing for differences in parameters between urban and rural men and women only turns to be significant for age dummies in regressions for men.

Fat Intakes (g)

Results for daily fat intakes (g) are summarized in Table 2.5. For men the effect of education dummy coefficients are individually significant but not jointly so. For women the effect of education is strongest in the whole and urban samples. For all sub-samples the highest educated men and women always have the highest fat intake (except for prime aged men and elderly women). Higher educated people are more likely to be engaged in more sedentary jobs and require less calories. However they consume more energy-rich fatty foods. This reflects the role of education as a proxy for income since consumption of animal products and income levels are positively associated as shown in Du et al. (2002) and Guo, Popkin and Zhai (1999).

The relationship between productive assets and fat intakes for men is the same as the relationships in the calorie intake regressions. For women in the overall sample

the effect of productive assets is positive and for women in the rural and younger aged sample the effects are the opposites than that found in men. Land farmed is negatively correlated with fat intake for rural and younger men. Growing grain products and raising animals are substitute uses of land. When more land is farmed there is less access to animal products and hence less fat intakes. The difference in parameters between men and women is significant for productive assets in rural and young cohort and the effects of community dummies differ between men and women when the sample is stratified by age. This may indicate clustering of similar-age people in some community. There is no significant difference between urban and rural men and women in education and productive assets effects.

Protein Intakes (g)

Results for daily protein intake (g) in Table 2.6 are similar to the results for fat intakes. Now, the effect of land on female protein intake is negative for the rural and prime aged samples. The effects of education, productive assets, age and cohort dummies between urban and rural residents are not statistically different. This is consistent with what we find in the descriptive measures in previous section. The positive female education effects for the overall sample and the urban sample are statistically significant at 0.05 and 0.1 levels respectively. This is consistent with what Behrman and Wolfe (1989) finds for women in Nicaragua using sibling data and fixed- and random-effect models in comparison with the standard estimates. The authors argue for the support of the effect of women's schooling in increasing nutrient intakes rather than reflecting the effects of unobserved childhood background variables.

Percent Calories From Fat

The effects of education on the percent calories from fat (Table 2.7) is significantly positive for men and women in all strata, except for elderly men. Recall only some

of these effects are significant in the model for the level of fat intakes in Table 5. If the body has an excess of energy-yielding nutrients (fat, protein and carbohydrates), it rearranges them into carbohydrates and fat storage compounds to be drawn upon between meals and overnight. Health recommendations urge people to limit fat intake to 30 percent of total calories required. Energy yielding nutrients also provide the raw materials for building the body's tissues and regulating its many activities. When the percent energy from fat is too high (not just the level of fat intakes) and energy expenditure is low, the body will gain weight (Whitney and Rolfes 1996). Higher educated people tend to have higher percent calorie from fat. The magnitude of the effect of education is different for different groups of people. Compared with the uneducated the effect of high school education is the strongest for the younger men and for the elderly women. However in Chapter 1 we find that in the overall pooled sample for women having some primary education is associated with 0.23 unit higher BMIs than those with no formal education at 0.1 significant level, but the impact rises significantly when one finishes primary schooling (0.47 unit). To put things into perspective, for a 5.5 foot woman a 0.47 unit increase in BMI is equivalent to a 3-pound increase in weight assuming fixed height. Completing senior high school, technical school or college is associated with a lower BMI than having no education (-0.02 and -0.7) or lower education levels. On average a 5.5 foot woman with a college equivalent degree weighs 4 pounds lighter than an uneducated counterpart.

The positive and increasing effects of productive assets on the percent of calories from fat is significant for women in all samples but the elderly sample, and for men in all samples but the urban sample. Again, land effects are generally negative except for elderly females. The difference in cohort effects between urban and rural men is significant, maybe indicating unobserved childhood background measures because the rural-urban differential in availability of foods was bigger in earlier times and the famine in the 1960s hit rural areas harder than the urban areas.

Percent Calories From Protein

Protein is a vital structural substance in all body cells. Meat is a good source of protein, as are milk, eggs, legumes, and many grains and vegetables. When the percent of calories from proteins is high and the total protein intakes are low, more protein is utilized to create energy and less is metabolized as protein to provide the vital supply to build cells.

The effects of education and productive assets on the percent of calories from protein (Table 2.8) is more pronounced than those in the level of protein intake. This suggests the percent calories from protein measures somewhat different aspect of nutrient intakes than the level of protein intakes alone. Or the level measure may be too skewed or with too much measurement errors that result in insignificant parameters.

Beside the significant education effects in all samples except for the elderly female, of particular interest is the difference between the effects of education in urban and rural men ($p\text{-value}=0.052$). The highest educated men in urban areas consume 1.26 percent more calorie from fat than those least educated, whereas men in the same education category in the rural areas consume about 0.4 percent more calorie from protein than those without any formal education. The male education effects among different age groups are stronger for the elderly and the female education effects are stronger for the younger ones.

For both men and women higher productive assets are associated with higher percent of calories from protein. Recall for men the productive assets are negatively associated with protein intakes and for women the association is positive. Economically better-off women may choose higher quality foods.

Percent Calories From Carbohydrates

Carbohydrates are converted by the body into glucose for immediate energy and into glycogen for reserve energy. Together glucose and glycogen provide about half of all the energy human nerves, muscles and other body tissues use. Higher education is related with lower carbohydrate intake for both men and women in all strata (Table 2.9). It could be because higher educated people are engaged in less labor intensive work that require less calories, or because they have more energy from fat and protein due to income effects.

The land effects are positive for exactly the opposite reason for negative land effect on fat intake. The differences between urban and rural population is significant for education and cohort effects among men.

2.4.2 Augmented Models

In addition to the above education and household resource effects we are interested in the effects of different food prices and community water and sanitation conditions on the demand for nutrients. For total calorie, fat and protein intakes we take the logarithm of the original level and the prices used are free market prices and are also in logarithm. For the percents of calories from fat, protein and carbohydrates regressions, the outcomes and prices are in original levels. Tables 2.10 to 2.12 summarize these results for overall, urban and rural samples. The food prices chosen in this paper are prices of rice, pork, eggs and edible oils. As is seen in Table 2.1, there are other food prices available in the survey. However, they can not be all included due to high correlations between them. For example, the price of rice and the price of wheat, the price of pork and the price of beef are highly correlated. When variables with high correlations are included in the same model the estimates are inaccurate. Community dummies, individual-year age dummies and five-year cohort dummies are also included.

All Outcomes

The positive effects of urban residency on calories for men and women are consistent with the findings that urban residents are more likely to consume and consume more of selected food groups and are more likely to be overweight (Popkin 1999).

The elasticities of the price of rice on total calorie, fat and protein intakes, and the percents of calories from fat and protein are negative in all sample strata; and the elasticity of the price of rice is positive for the percent of calories from carbohydrates. These results suggest that with an increase in rice price people consume less energy-rich fat and protein and substitute for cheaper foods with high carbohydrates.

The effects of the price of pork are the exact opposite of the effects of the price of rice. The effects of the prices of eggs are in the same directions as in the effects of the price of rice. The effects of the price of edible oils are similar except that it has a positive effect on the percent of calories from protein.

Generally there are two effects of increased food prices. For net producers of foods increases in prices drive up income and hence demand for more or better food consumption. For net buyers of foods increases in prices generally tighten the budget set and reduce the demand for normal goods. However, the fact that the coefficients in urban (Table 2.11) and rural (Table 2.12) areas are close to each other indicates that possibly no rural households are net producers or that income effects are small. In addition since the outcomes are not specific foods consumed but macronutrients from all foods the two effects of prices are more difficult to disentangle.

The types and sources of households water sources, toilet facility and sanitation (excreta) condition around the living areas are aggregated from the household level to derive the percent of households within each community having certain resources. Compared with communities with underground water sources (Table 2.10) a 1% increase in percentage of households in a community with rain, snow or river as water sources is associated with significant increase in calorie intake, protein and percents

of calories from protein and a decrease in percent calorie from carbohydrates for both men and women. Higher percentages of household using water from snow, rain or river indicates a poorer neighborhood. The effect on calories makes sense. There is evidently some other mechanism through which the water condition of the community is positively related with protein intake

Compared with individuals living in clean environments men from communities with a higher percentage of households with little excreta around the living quarters consume more calories, fat and protein (for women only the protein effect is strong and positive). Men whose living environment has some excreta consume less calories, fat, and the percent of calories from fat, but have more percents of calories from protein and carbohydrates. The effects of excreta are stronger in urban areas than those in rural areas. These results suggest there may be some income effects proxied by these environmental factors.

The effects of individual and household characteristics on the percent of calories from fat, protein and carbohydrates in all strata are strong and exhibit the same properties as in the basic regressions.

Table 2.4: Daily Calorie Intakes (Kcal) in CHNS 89, 91, 93 From All Food Groups

Variables	Male					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	2.073 (35.832)	-2.730 (64.603)	-7.718 (43.356)	53.687 (93.206)	-59.653 (62.780)	-2.116 (70.670)
Primary	-1.254 (37.474)	34.627 (65.083)	-23.464 (45.608)	31.184 (92.298)	-41.878 (64.047)	-44.806 (88.419)
Middle school	-42.834 (37.256)	-58.158 (64.867)	-41.865 (45.352)	-32.030 (88.842)	-83.212 (65.812)	-44.148 (85.817)
High school	-66.048 (42.370)	-38.713 (69.665)	-96.446 (53.725)*	-55.265 (92.030)	-50.869 (81.757)	8.316 (152.778)
Tech/College+	-57.371 (49.208)	-55.810 (71.138)	-31.530 (82.732)	-59.995 (100.486)	-73.466 (82.486)	115.590 (130.409)
Log real prod assets 1	23.097 (5.774)**	11.031 (8.954)	30.555 (7.453)**	31.129 (8.517)**	22.996 (11.371)**	-6.554 (17.266)
Log real prod assets 2	-13.192 (5.557)**	0.545 (8.832)	-20.890 (7.119)**	-16.494 (8.005)**	-19.654 (10.642)*	29.992 (16.515)*
Land	4.261 (2.878)	56.608 (45.737)	4.156 (2.859)	2.240 (4.243)	5.810 (5.013)	3.788 (7.205)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	8695	2842	5853	4681	2732	1121
R-squared	0.331	0.281	0.334	0.347	0.406	0.503
P-value for testing coefficients equal to zero						
Education	0.2040	0.3963	0.3826	0.1118	0.8450	0.8510
Assets	0.0003	0.3083	0.0002	0.0007	0.1095	0.0762
Age dummies	0.0000	0.0000	0.0002	0.9302	0.1083	0.0111
Cohort dummies	0.2001	0.7844	0.2694	0.9592	0.0436	0.4543
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.9789	0.7150	0.7375	0.9354	0.8875	0.0810
Assets	0.0158	0.7434	0.0106	0.0167	0.1555	0.1184
Age dummies	0.0008	0.0040	0.0082	0.4308	0.2620	0.0122
Cohort dummies	0.0214	0.1177	0.4415	0.3834	0.2019	0.0626
Community dummies	0.0000	0.5938	0.0000	0.0000	0.0000	0.0000

Table continues

Table 2.4 (cont'd)

Variables	Female					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-14.187 (21.477)	-7.236 (42.620)	-25.473 (25.368)	1.183 (37.256)	-31.098 (35.013)	-89.883 (57.689)
Primary	1.796 (24.473)	-38.340 (46.176)	9.756 (29.130)	6.835 (39.450)	-39.532 (39.754)	-33.006 (95.163)
Middle school	-58.278 (24.831)**	-71.625 (47.067)	-53.457 (30.087)*	-42.916 (37.247)	-96.887 (44.161)**	195.737 (178.201)
High school	-76.892 (32.698)**	-63.389 (53.449)	-63.965 (43.974)	-65.066 (43.756)	6.901 (97.454)	897.424 (326.438)**
Tech/College+	-42.833 (41.195)	-10.817 (58.965)	-80.166 (68.484)	-32.796 (55.740)	35.744 (88.311)	271.002 (199.648)
Log real prod assets 1	2.862 (4.248)	2.049 (7.367)	3.812 (5.293)	1.549 (5.890)	5.013 (8.139)	3.591 (11.216)
Log real prod assets 2	3.326 (4.029)	3.591 (6.448)	2.589 (5.129)	4.890 (5.711)	6.625 (8.518)	-4.025 (9.784)
Land	1.829 (1.924)	9.259 (27.384)	1.212 (1.908)	0.652 (2.983)	1.959 (3.307)	10.780 (5.431)**
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	9470	3136	6334	5226	2958	1136
R-squared	0.349	0.279	0.354	0.371	0.399	0.463
P-value for testing coefficients equal to zero						
Education	0.0315	0.4694	0.1431	0.2768	0.2389	0.0184
Assets	0.2064	0.7298	0.2753	0.3675	0.2684	0.9166
Age dummies	0.0967	0.1989	0.2611	0.0283	0.7934	0.4808
Cohort dummies	0.3269	0.1794	0.5256	0.4051	0.9021	0.0950
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.6038	Women	0.6922		
Assets		0.1392		0.9809		
Age dummies		0.0289		0.7323		
Cohort dummies		0.8070		0.4435		

Note: Also included in the regressions are age dummies and five-year cohort dummies.

Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.5: Daily Fat Intakes (gram) in CHNS 89, 91, 93 From All Food Groups

Variables	Overall	Male				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	1.117 (1.726)	-0.027 (3.485)	1.752 (2.007)	7.518 (3.611)**	2.350 (2.909)	-1.684 (3.717)
Primary	2.431 (1.800)	3.616 (3.636)	2.348 (2.090)	7.877 (3.553)**	3.570 (3.096)	0.622 (4.081)
Middle school	4.361 (1.849)**	1.721 (3.665)	5.534 (2.161)**	9.413 (3.439)**	5.910 (3.220)*	0.194 (4.974)
High school	5.121 (2.117)**	4.527 (3.866)	5.241 (2.576)**	10.576 (3.650)**	4.875 (4.224)	3.958 (8.423)
Tech/College+	6.341 (2.471)**	3.322 (3.935)	11.665 (3.891)**	13.790 (4.171)**	2.376 (4.380)	9.979 (6.410)
Log real prod assets 1	0.770 (0.312)**	-0.004 (0.472)	1.272 (0.401)**	1.263 (0.425)**	1.033 (0.711)	-1.014 (1.012)
Log real prod assets 2	-0.130 (0.297)	0.282 (0.431)	-0.490 (0.385)	-0.361 (0.392)	0.014 (0.680)	0.550 (1.022)
Land	-0.380 (0.123)**	0.038 (2.145)	-0.357 (0.125)**	-0.557 (0.171)**	-0.228 (0.206)	0.087 (0.303)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	8695	2842	5853	4681	2732	1121
R-squared	0.259	0.189	0.282	0.306	0.330	0.544
P-value for testing coefficients equal to zero						
Education	0.0356	0.5092	0.0120	0.0336	0.5248	0.5627
Assets	0.0056	0.7383	0.0010	0.0016	0.0518	0.5947
Age dummies	0.1595	0.0453	0.3547	0.6815	0.1342	0.5233
Cohort dummies	0.5570	0.0285	0.2425	0.9103	0.8234	0.3355
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.7725	0.7492	0.3157	0.4344	0.2054	0.0075
Assets	0.1574	0.6471	0.0267	0.0500	0.4419	0.5983
Age dummies	0.0062	0.1215	0.2253	0.1267	0.2148	0.5369
Cohort dummies	0.2401	0.1498	0.6458	0.9433	0.7902	0.2880
Community dummies	0.9855	0.9998	0.6613	0.0000	0.0000	0.0000
Table continues						

Table continues

Table 2.5 (cont'd)

Variables	Overall	Female				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.204 (1.148)	-0.351 (2.556)	0.085 (1.311)	-0.218 (1.890)	-0.106 (2.010)	-2.785 (3.269)
Primary	3.572 (1.343)**	1.464 (2.560)	4.017 (1.606)**	4.105 (2.158)*	-0.350 (2.265)	7.681 (5.344)
Middle school	4.288 (1.338)**	3.557 (2.554)	4.156 (1.605)**	4.989 (1.965)**	-0.131 (2.606)	16.003 (8.509)*
High school	7.216 (1.887)**	6.274 (3.014)**	8.602 (2.597)**	6.929 (2.444)**	10.329 (6.119)*	60.196 (13.668)**
Tech/College+	8.015 (2.323)**	8.027 (3.368)**	8.017 (3.820)**	7.418 (3.066)**	9.637 (5.232)*	26.666 (10.514)**
Log real prod assets 1	0.052 (0.231)	0.274 (0.438)	-0.034 (0.275)	-0.012 (0.317)	-0.086 (0.515)	0.160 (0.629)
Log real prod assets 2	0.492 (0.220)**	0.530 (0.371)	0.502 (0.274)*	0.660 (0.315)**	0.807 (0.510)	-0.022 (0.533)
Land	-0.137 (0.112)	-0.368 (0.864)	-0.140 (0.114)	-0.160 (0.177)	-0.177 (0.158)	0.679 (0.240)**
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	9470	3136	6334	5226	2958	1136
R-squared	0.262	0.192	0.282	0.291	0.340	0.468
P-value for testing coefficients equal to zero						
Education	0.0001	0.1124	0.0013	0.0030	0.2810	0.0000
Assets	0.0065	0.0763	0.0558	0.0240	0.0970	0.9513
Age dummies	0.0162	0.1710	0.4308	0.1351	0.3741	0.1184
Cohort dummies	0.5334	0.2434	0.7704	0.6907	0.8267	0.4234
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.2848	Women	0.9442		
Assets		0.1188		0.7253		
Age dummies		0.6271		0.9112		
Cohort dummies		0.0228		0.3976		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.6: Daily Protein Intakes (gram) in CHNS 89, 91, 93 From All Food Groups

Variables	Overall	Male				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.879 (1.512)	0.194 (3.021)	0.512 (1.770)	2.206 (3.875)	-3.426 (2.514)	3.314 (3.220)
Primary	1.030 (1.592)	5.723 (3.133)*	-0.981 (1.871)	0.846 (3.905)	-1.815 (2.579)	5.230 (4.118)
Middle school	0.851 (1.608)	2.115 (3.156)	0.281 (1.885)	0.681 (3.741)	-3.033 (2.749)	5.349 (4.652)
High school	0.784 (1.890)	3.562 (3.453)	-0.912 (2.291)	0.128 (3.929)	2.702 (3.734)	10.821 (8.753)
Tech/College+	2.835 (2.286)	4.907 (3.471)	2.600 (3.860)	3.993 (4.464)	-2.061 (3.903)	15.643 (6.567)**
Log real prod assets 1	1.096 (0.270)**	0.575 (0.437)	1.399 (0.341)**	1.712 (0.417)**	0.902 (0.473)*	-0.245 (0.813)
Log real prod assets 2	-0.190 (0.261)	0.055 (0.422)	-0.375 (0.330)	-0.463 (0.394)	-0.300 (0.490)	1.079 (0.713)
Land	-0.228 (0.114)**	2.852 (2)	-0.245 (0.114)**	-0.225 (0.192)	-0.290 (0.214)	-0.226 (0.318)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	8695	2842	5853	4681	2732	1121
R-squared	0.286	0.280	0.302	0.308	0.364	0.502
P-value for testing coefficients equal to zero						
Education	0.8867	0.2087	0.7938	0.6958	0.3803	0.2487
Assets	0.0000	0.2548	0.0000	0.0000	0.1055	0.1717
Age dummies	0.0017	0.0002	0.0824	0.8311	0.0509	0.3515
Cohort dummies	0.2045	0.3199	0.3883	0.6419	0.7193	0.3684
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.6882	0.5022	0.2199	0.8061	0.6673	0.6463
Assets	0.0874	0.6978	0.0959	0.0127	0.2942	0.1925
Age dummies	0.0020	0.0210	0.0429	0.9655	0.6358	0.0000
Cohort dummies	0.1800	0.1475	0.7335	0.5918	0.7799	0.0569
Community dummies	0.0067	0.9859	0.0000	0.0000	0.0000	0.0000
Table continues						

Table continues

Table 2.6 (cont'd)

Variables	Female					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.499 (1.039)	2.027 (2.074)	-0.368 (1.208)	1.110 (1.782)	-0.435 (1.651)	-1.690 (3.299)
Primary	2.941 (1.212)**	2.089 (2.309)	2.879 (1.434)**	2.717 (1.906)	1.311 (1.951)	4.204 (5.846)
Middle school	1.659 (1.199)	1.528 (2.192)	1.626 (1.466)	2.297 (1.766)	-0.825 (2.188)	10.116 (9.002)
High school	2.535 (1.650)	3.994 (2.672)	2.354 (2.189)	2.871 (2.148)	3.740 (4.852)	35.275 (20.756)*
Tech/College+	6.281 (2.166)**	8.376 (2.950)**	3.535 (3.898)	6.346 (2.920)**	6.927 (4.165)*	10.773 (11.672)
Log real prod assets 1	0.322 (0.224)	0.092 (0.380)	0.445 (0.279)	0.150 (0.332)	0.403 (0.384)	0.971 (0.694)
Log real prod assets 2	0.390 (0.220)*	0.392 (0.364)	0.343 (0.276)	0.530 (0.326)	0.696 (0.420)*	-0.616 (0.621)
Land	-0.172 (0.092)*	1.139 (1.224)	-0.210 (0.091)**	-0.145 (0.144)	-0.331 (0.162)**	0.129 (0.221)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	9470	3136	6334	5226	2958	1136
R-squared	0.265	0.244	0.288	0.283	0.356	0.449
P-value for testing coefficients equal to zero						
Education	0.0294	0.0753	0.2141	0.3725	0.4309	0.3407
Assets	0.0003	0.3718	0.0006	0.0210	0.0071	0.3720
Age dummies	0.0160	0.2034	0.0372	0.1578	0.5232	0.0000
Cohort dummies	0.5023	0.1426	0.4454	0.4634	0.9704	0.0927
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.1718	Women	0.6646		
Assets		0.3248		0.6937		
Age dummies		0.2508		0.6214		
Cohort dummies		0.5136		0.2736		

Note: Also included in the regressions are age dummies and five-year cohort dummies.
 Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.7: Percent Calorie From Fat in CHNS 89, 91, 93 From All Food Groups

Variables	Overall	Male				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.500 (0.415)	0.112 (0.865)	0.636 (0.477)	1.805 (0.863)**	0.950 (0.703)	-0.437 (0.977)
Primary	0.893 (0.437)**	1.282 (0.916)	0.778 (0.501)	2.092 (0.861)**	1.170 (0.755)	0.400 (1.099)
Middle school	1.777 (0.440)**	1.341 (0.892)	1.931 (0.510)**	2.897 (0.830)**	2.333 (0.776)**	0.263 (1.319)
High school	2.074 (0.500)**	2.062 (0.938)**	2.075 (0.606)**	3.358 (0.880)**	1.900 (1.047)*	1.416 (1.978)
Tech/College+	2.357 (0.585)**	1.707 (0.957)*	3.607 (0.913)**	4.388 (1.001)**	0.968 (1.070)	2.505 (1.650)
Log real prod assets 1	0.115 (0.070)	-0.063 (0.115)	0.233 (0.088)**	0.238 (0.096)**	0.191 (0.145)	-0.253 (0.265)
Log real prod assets 2	0.040 (0.065)	0.103 (0.106)	-0.038 (0.082)	-0.031 (0.088)	0.140 (0.138)	-0.031 (0.246)
Land	-0.121 (0.032)**	-0.331 (0.448)	-0.117 (0.033)**	-0.160 (0.044)**	-0.088 (0.045)*	0.026 (0.078)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	8695	2842	5853	4681	2732	1121
R-squared	0.372	0.262	0.361	0.406	0.449	0.553
P-value for testing coefficients equal to zero						
Education	0.0000	0.0720	0.0000	0.0001	0.0442	0.5739
Assets	0.0111	0.6197	0.0020	0.0043	0.0064	0.3696
Age dummies	0.2276	0.0025	0.2138	0.1602	0.1637	0.6588
Cohort dummies	0.2951	0.0021	0.2181	0.3432	0.8226	0.4228
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.7630	0.7472	0.6610	0.5437	0.1598	0.0072
Assets	0.3807	0.3602	0.2066	0.2043	0.4789	0.4713
Age dummies	0.0355	0.3345	0.3884	0.1071	0.3320	0.7756
Cohort dummies	0.4018	0.1761	0.9041	0.8010	0.6952	0.7649
Community dummies	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000
Table continues						

Table continues

Table 2.7 (cont'd)

Variables	Overall	Female				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.327 (0.312)	-0.077 (0.673)	0.382 (0.359)	0.173 (0.513)	0.287 (0.530)	-0.682 (1.111)
Primary	1.216 (0.359)**	1.147 (0.686)*	1.188 (0.428)**	1.135 (0.569)**	0.566 (0.611)	3.487 (1.811)*
Middle school	1.930 (0.361)**	2.040 (0.694)**	1.833 (0.428)**	2.143 (0.529)**	0.953 (0.728)	3.263 (1.837)*
High school	2.700 (0.473)**	2.755 (0.784)**	2.872 (0.638)**	2.690 (0.625)**	3.400 (1.399)**	13.855 (3.015)**
Tech/College+	3.024 (0.627)**	3.112 (0.899)**	3.321 (1.119)**	3.028 (0.822)**	3.288 (1.309)**	6.266 (2.356)**
Log real prod assets 1	0.028 (0.064)	0.036 (0.119)	0.037 (0.077)	0.036 (0.086)	-0.039 (0.132)	0.014 (0.220)
Log real prod assets 2	0.163 (0.060)**	0.199 (0.098)**	0.142 (0.076)*	0.176 (0.084)**	0.258 (0.125)**	0.026 (0.194)
Land	-0.049 (0.037)	-0.124 (0.198)	-0.049 (0.038)	-0.052 (0.060)	-0.078 (0.049)	0.154 (0.073)**
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	9470	3136	6334	5226	2958	1136
R-squared	0.385	0.284	0.363	0.410	0.452	0.538
P-value for testing coefficients equal to zero						
Education	0.0000	0.0013	0.0000	0.0000	0.0808	0.0000
Assets	0.0002	0.0154	0.0076	0.0048	0.0284	0.9695
Age dummies	0.0494	0.1789	0.7764	0.2447	0.3565	0.0090
Cohort dummies	0.8874	0.5623	0.9638	0.9321	0.4662	0.9364
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.2811	Women	0.9704		
Assets		0.1125		0.8342		
Age dummies		0.7193		0.9520		
Cohort dummies		0.0049		0.5967		

Note: Also included in the regressions are age dummies and five-year cohort dummies.

Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.8: Percent Calorie From Protein in CHNS 89, 91, 93 From All Food Groups

Variables	Male					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.139 (0.130)	0.137 (0.282)	0.089 (0.146)	0.133 (0.278)	-0.206 (0.205)	0.461 (0.337)
Primary	0.191 (0.134)	0.734 (0.301)**	-0.028 (0.149)	-0.006 (0.281)	-0.029 (0.207)	0.808 (0.400)**
Middle school	0.371 (0.136)**	0.747 (0.297)**	0.228 (0.152)	0.314 (0.270)	0.033 (0.226)	1.076 (0.478)**
High school	0.479 (0.159)**	0.862 (0.319)**	0.332 (0.185)*	0.328 (0.284)	0.863 (0.342)**	1.378 (0.805)*
Tech/College+	0.791 (0.207)**	1.255 (0.332)**	0.426 (0.343)	0.997 (0.361)**	0.138 (0.347)	1.703 (0.691)**
Log real prod assets 1	0.025 (0.023)	0.006 (0.041)	0.030 (0.029)	0.065 (0.035)*	0.021 (0.042)	-0.084 (0.081)
Log real prod assets 2	0.042 (0.022)*	0.004 (0.040)	0.056 (0.027)**	0.025 (0.032)	0.053 (0.040)	0.045 (0.079)
Land	-0.046 (0.010)**	0.135 (0.142)	-0.049 (0.010)**	-0.030 (0.016)**	-0.069 (0.020)**	-0.048 (0.033)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	8695	2842	5853	4681	2732	1121
R-squared	0.388	0.315	0.369	0.412	0.492	0.509
P-value for testing coefficients equal to zero						
Education	0.0007	0.0002	0.0906	0.0046	0.0446	0.0550
Assets	0.0003	0.9556	0.0000	0.0003	0.0740	0.5752
Age dummies	0.0510	0.0360	0.7870	0.1474	0.1057	0.7392
Cohort dummies	0.5982	0.1570	0.9335	0.3507	0.2567	0.9935
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.4432	0.8341	0.1668	0.4789	0.3750	0.6699
Assets	0.9252	0.6270	0.9908	0.4535	0.7217	0.1825
Age dummies	0.0038	0.0009	0.0244	0.7925	0.9935	0.0013
Cohort dummies	0.5190	0.6336	0.7615	0.1738	0.8771	0.9873
Community dummies	1.0000	1.0000	1.0000	0.0751	0.0000	0.0000
Table continues						

Table continues

Table 2.8 (cont'd)

Variables	Overall	Female				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.169 (0.101)*	0.280 (0.210)	0.126 (0.117)	0.169 (0.164)	0.080 (0.161)	0.285 (0.410)
Primary	0.448 (0.116)**	0.531 (0.226)**	0.407 (0.136)**	0.370 (0.175)**	0.444 (0.188)**	0.739 (0.589)
Middle school	0.581 (0.116)**	0.676 (0.225)**	0.556 (0.138)**	0.607 (0.163)**	0.501 (0.222)**	0.280 (0.959)
High school	0.790 (0.153)**	1.029 (0.268)**	0.615 (0.188)**	0.795 (0.196)**	0.666 (0.404)*	-0.396 (1.856)
Tech/College+	1.229 (0.214)**	1.391 (0.289)**	1.046 (0.445)**	1.206 (0.280)**	0.832 (0.424)**	-0.122 (1.030)
Log real prod assets 1	0.026 (0.023)	0.009 (0.040)	0.033 (0.028)	0.007 (0.034)	0.033 (0.040)	0.120 (0.080)
Log real prod assets 2	0.050 (0.022)**	0.043 (0.037)	0.051 (0.028)*	0.056 (0.032)*	0.079 (0.040)**	-0.050 (0.071)
Land	-0.040 (0.010)**	0.130 (0.087)	-0.043 (0.010)**	-0.026 (0.014)*	-0.074 (0.026)**	-0.045 (0.035)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	9470	3136	6334	5226	2958	1136
R-squared	0.391	0.310	0.371	0.416	0.491	0.511
P-value for testing coefficients equal to zero						
Education	0.0000	0.0001	0.0003	0.0000	0.0533	0.8506
Assets	0.0000	0.2191	0.0002	0.0108	0.0028	0.2812
Age dummies	0.0000	0.0033	0.0000	0.3385	0.2161	0.0000
Cohort dummies	0.1461	0.2052	0.3668	0.1131	0.7610	0.8885
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.0519	Women	0.8572		
Assets		0.1587		0.6998		
Age dummies		0.7410		0.0182		
Cohort dummies		0.4815		0.5024		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.9: Percent Calorie From Carbohydrates in CHNS 89, 91, 93 From All Food Groups

Variables	Overall	Male				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-1.182 (0.520)**	-0.966 (1.027)	-1.208 (0.615)**	-2.444 (1.092)**	-0.955 (0.858)	-0.848 (1.183)
Primary	-1.586 (0.535)**	-2.666 (1.066)**	-1.151 (0.633)*	-2.919 (1.093)**	-1.417 (0.904)	-1.972 (1.324)
Middle school	-2.881 (0.549)**	-2.564 (1.046)**	-2.967 (0.661)**	-3.824 (1.058)**	-3.113 (0.961)**	-3.561 (1.603)**
High school	-3.624 (0.632)**	-3.977 (1.125)**	-3.436 (0.791)**	-4.625 (1.124)**	-3.810 (1.337)**	-5.066 (2.415)**
Tech/College+	-4.511 (0.736)**	-3.856 (1.113)**	-6.283 (1.302)**	-6.158 (1.277)**	-3.406 (1.359)**	-4.363 (2.018)**
Log real prod assets 1	-0.152 (0.086)*	0.043 (0.136)	-0.271 (0.108)**	-0.333 (0.123)**	-0.256 (0.160)	0.492 (0.307)
Log real prod assets 2	-0.168 (0.080)**	-0.191 (0.121)	-0.106 (0.104)	-0.035 (0.113)	-0.243 (0.157)	-0.535 (0.293)*
Land	0.136 (0.041)**	0.347 (0.626)	0.130 (0.042)**	0.185 (0.056)**	0.144 (0.074)*	0.091 (0.080)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	8695	2842	5853	4681	2732	1121
R-squared	0.408	0.303	0.393	0.435	0.473	0.592
P-value for testing coefficients equal to zero						
Education	0.0000	0.0004	0.0000	0.0000	0.0058	0.0780
Assets	0.0000	0.2312	0.0000	0.0001	0.0012	0.1631
Age dummies	0.0103	0.0018	0.2347	0.4423	0.0071	0.2489
Cohort dummies	0.4791	0.0024	0.3261	0.2258	0.5081	0.1363
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.6593	0.7189	0.2273	0.5011	0.6979	0.5228
Assets	0.7978	0.6449	0.3725	0.2172	0.4559	0.1988
Age dummies	0.0000	0.3632	0.0117	0.1436	0.2170	0.0614
Cohort dummies	0.5038	0.1350	0.9598	0.5705	0.6848	0.4973
Community dummies	0.9344	0.8903	0.8620	0.0000	0.0000	0.0000

Table continues

Table 2.9 (cont'd)

Variables	Female					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-0.415 (0.353)	-0.087 (0.723)	-0.446 (0.410)	-0.228 (0.569)	-0.415 (0.600)	0.376 (1.246)
Primary	-1.689 (0.411)**	-1.522 (0.760)**	-1.686 (0.495)**	-1.541 (0.634)**	-1.206 (0.705)*	-3.871 (2.042)*
Middle school	-2.474 (0.409)**	-2.579 (0.767)**	-2.374 (0.489)**	-2.654 (0.592)**	-1.633 (0.817)**	-3.635 (2.532)
High school	-3.453 (0.532)**	-3.594 (0.865)**	-3.539 (0.721)**	-3.386 (0.699)**	-4.251 (1.543)**	-12.073 (4.407)**
Tech/College+	-4.204 (0.692)**	-4.384 (0.975)**	-4.185 (1.254)**	-4.223 (0.913)**	-4.192 (1.335)**	-6.584 (3.098)**
Log real prod assets 1	-0.078 (0.073)	-0.075 (0.137)	-0.088 (0.086)	-0.058 (0.098)	0.006 (0.143)	-0.171 (0.266)
Log real prod assets 2	-0.203 (0.069)**	-0.225 (0.117)*	-0.190 (0.084)**	-0.219 (0.096)**	-0.372 (0.134)**	0.099 (0.251)
Land	0.089 (0.042)**	0.027 (0.234)	0.091 (0.043)**	0.073 (0.067)	0.147 (0.066)**	-0.095 (0.092)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	9470	3136	6334	5226	2958	1136
R-squared	0.440	0.328	0.404	0.469	0.506	0.566
P-value for testing coefficients equal to zero						
Education	0.0000	0.0000	0.0000	0.0000	0.0160	0.0037
Assets	0.0000	0.0111	0.0001	0.0008	0.0007	0.8124
Age dummies	0.0000	0.0556	0.0033	0.4008	0.2529	0.0000
Cohort dummies	0.5839	0.3691	0.8412	0.8964	0.5112	0.7181
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.0583	Women	0.9911		
Assets		0.1468		0.9662		
Age dummies		0.9745		0.8786		
Cohort dummies		0.0036		0.6052		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.10: Food Consumption in CHNS 89,91,93 with Community Characteristics: Overall

Dependent Var	Male					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Some primary edu	0.0091 (0.0136)	0.0699 (0.0402)*	0.0187 (0.0163)	0.584 (0.419)	0.138 (0.129)	-1.357 (0.525)**
Primary	0.0094 (0.0137)	0.0770 (0.0417)*	0.0173 (0.0167)	0.854 (0.434)**	0.130 (0.133)	-1.462 (0.538)**
Middle school	-0.0023 (0.0138)	0.1590 (0.0417)**	0.0217 (0.0168)	1.805 (0.438)**	0.356 (0.135)**	-3.013 (0.551)**
High school	-0.0150 (0.0156)	0.1832 (0.0463)**	0.0191 (0.0189)	2.231 (0.500)**	0.508 (0.158)**	-3.857 (0.633)**
Tech/College+	-0.0049 (0.0186)	0.1903 (0.0501)**	0.0389 (0.0221)*	2.050 (0.580)**	0.712 (0.201)**	-4.125 (0.738)**
Log R prod assets 1	0.0099 (0.0021)**	0.0279 (0.0055)**	0.0132 (0.0025)**	0.172 (0.069)**	0.040 (0.022)*	-0.253 (0.084)**
Log R prod assets 2	-0.0054 (0.0020)**	-0.0069 (0.0054)	-0.0050 (0.0024)**	0.010 (0.065)	0.013 (0.021)	-0.095 (0.081)
Land	0.0024 (0.0012)**	-0.0036 (0.0032)	-0.0001 (0.0013)	-0.108 (0.030)**	-0.038 (0.010)**	0.126 (0.042)**
Urban Residence	0.2154 (0.0773)**	2.3329 (0.1873)**	0.5319 (0.0893)**	23.291 (2.219)**	3.938 (0.810)**	-26.542 (2.758)**
Log Price of Rice	-0.0696 (0.0171)**	-0.1576 (0.0490)**	-0.0981 (0.0214)**	-0.883 (0.373)**	-0.341 (0.138)**	2.408 (0.472)**
Log Price of Pork	0.0187 (0.0245)	0.2569 (0.0719)**	0.0786 (0.0307)**	0.618 (0.124)**	0.126 (0.043)**	-0.849 (0.159)**
Log Price of Eggs	-0.0181 (0.0038)**	-0.0318 (0.0112)**	-0.0270 (0.0047)**	-0.034 (0.062)	-0.062 (0.021)**	0.005 (0.077)
Log Price of Oil	-0.0603 (0.0248)**	-0.1561 (0.0764)**	-0.0409 (0.0306)	-0.375 (0.141)**	0.056 (0.049)	0.262 (0.167)
Water Source 2	0.0321 (0.0268)	0.3157 (0.0857)**	0.0646 (0.0328)**	4.037 (0.869)**	0.483 (0.268)*	-3.937 (1.065)**
Water Source 3	0.0911 (0.0377)**	0.2712 (0.1203)**	0.1362 (0.0468)**	1.472 (1.289)	0.586 (0.357)*	-2.829 (1.597)*

Table continues

Table 2.10 (cont'd)

Dependent Var	Female					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Some primary edu	-0.0032 (0.0091)	0.0381 (0.0297)	0.0055 (0.0118)	0.294 (0.315)	0.136 (0.102)	-0.349 (0.356)
Primary	0.0139 (0.0101)	0.1503 (0.0335)**	0.0400 (0.0131)**	1.482 (0.364)**	0.377 (0.115)**	-1.863 (0.414)**
Middle school	-0.0202 (0.0103)**	0.1616 (0.0329)**	0.0168 (0.0132)	2.044 (0.365)**	0.528 (0.117)**	-2.505 (0.412)**
High school	-0.0270 (0.0139)*	0.1985 (0.0419)**	0.0165 (0.0176)	2.818 (0.481)**	0.637 (0.152)**	-3.402 (0.539)**
Tech/College+	-0.0062 (0.0190)	0.1878 (0.0510)**	0.0610 (0.0239)**	2.518 (0.650)**	1.059 (0.214)**	-3.482 (0.713)**
Log R prod assets 1	0.0036 (0.0019)*	0.0143 (0.0052)**	0.0064 (0.0024)**	0.086 (0.064)	0.038 (0.022)*	-0.156 (0.073)**
Log R prod assets 2	-0.0002 (0.0017)	0.0047 (0.0050)	0.0014 (0.0023)	0.128 (0.061)**	0.031 (0.021)	-0.146 (0.069)**
Land	0.0019 (0.0010)*	-0.0022 (0.0031)	-0.0009 (0.0011)	-0.061 (0.036)*	-0.042 (0.009)**	0.106 (0.042)**
Urban Residence	0.2494 (0.0824)**	2.4656 (0.2005)**	0.6050 (0.1002)**	24.207 (2.334)**	4.555 (0.830)**	-28.658 (2.655)**
Log Price of Rice	-0.0332 (0.0156)**	-0.1060 (0.0457)**	-0.0756 (0.0201)**	-0.553 (0.372)	-0.469 (0.133)**	1.068 (0.407)**
Log Price of Pork	0.0161 (0.0231)	0.3100 (0.0685)**	0.0794 (0.0298)**	0.734 (0.121)**	0.131 (0.040)**	-0.824 (0.133)**
Log Price of Eggs	-0.0139 (0.0035)**	-0.0306 (0.0109)**	-0.0196 (0.0045)**	-0.095 (0.060)	-0.026 (0.020)	0.107 (0.066)
Log Price of Oil	-0.0522 (0.0212)**	-0.1853 (0.0697)**	-0.0109 (0.0276)	-0.467 (0.140)**	0.122 (0.044)**	0.392 (0.146)**
Water Source 2	0.0576 (0.0238)**	0.3074 (0.0845)**	0.0788 (0.0326)**	3.458 (0.867)**	0.350 (0.278)	-3.712 (0.980)**
Water Source 3	0.0889 (0.0391)**	0.1855 (0.1220)	0.0902 (0.0484)*	1.753 (1.275)	0.061 (0.372)	-2.370 (1.400)*

Table continues

Table 2.10 (cont'd)

Dependent Var	Male					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Water Source 4	-0.0347 (0.0244)	0.1699 (0.0724)**	0.0092 (0.0307)	3.032 (0.830)**	0.631 (0.261)**	-2.866 (0.992)**
In house no flush	0.2178 (0.0632)**	0.3453 (0.1846)*	0.1944 (0.0817)**	0.560 (2.121)	-0.981 (0.772)	3.822 (2.550)
Outside toilets	0.0690 (0.0499)	0.0294 (0.1324)	0.0442 (0.0630)	-1.596 (1.656)	-0.842 (0.664)	3.650 (1.924)*
Open pit	0.0355 (0.0481)	0.0467 (0.1297)	0.0356 (0.0622)	-0.486 (1.633)	-0.484 (0.657)	1.893 (1.907)
No toilets	0.0142 (0.0887)	-0.0652 (0.2548)	0.0421 (0.1112)	0.691 (2.847)	-0.111 (1)	2.661 (3.554)
Very little excreta	0.0785 (0.0243)**	0.1416 (0.0683)**	0.1048 (0.0297)**	0.785 (0.714)	0.363 (0.234)	-1.401 (0.904)
Some excreta	-0.0693 (0.0235)**	-0.2504 (0.0677)**	-0.0338 (0.0289)	-1.822 (0.703)**	0.603 (0.240)**	2.121 (0.885)**
Much excreta	-0.0038 (0.0992)	0.3738 (0.2838)	-0.0795 (0.1171)	-1.351 (2.552)	-1.135 (0.817)	-1.099 (3.487)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	8415	8415	8415	8415	8415	8415
R-squared	0.2352	0.2497	0.2124	0.303	0.318	0.344
P-value for testing coefficients equal to zero						
Education	0.2171	0.0176	0.8943	0.0000	0.0005	0.0000
Prod Assets	0.0000	0.0002	0.0000	0.0008	0.0054	0.0000
Age dummies	0.0001	0.0570	0.0062	0.0052	0.0201	0.0009
Cohort dummies	0.0827	0.6370	0.2266	0.0320	0.5115	0.3137
Prices	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Water source	0.0088	0.0052	0.0220	0.0000	0.0655	0.0010
Toilet type	0.0225	0.6854	0.5398	0.3035	0.6411	0.3092
Excreta	0.0000	0.0001	0.0003	0.0088	0.0270	0.0043
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

Table 2.10 (cont'd)

Dependent Var	Female					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Water Source 4	0.0240 (0.0224)	0.1668 (0.0709)**	0.0417 (0.0301)	2.383 (0.800)**	0.376 (0.263)	-2.965 (0.897)**
In house no flush	0.2497 (0.0630)**	0.5456 (0.1872)**	0.2361 (0.0828)**	3.253 (2.255)	-0.762 (0.755)	-1.617 (2.479)
Outside toilets	0.1007 (0.0438)**	0.0973 (0.1206)	0.0940 (0.0560)*	-0.225 (1.591)	-0.346 (0.596)	0.779 (1.741)
Open pit	0.0354 (0.0427)	0.0156 (0.1199)	0.0299 (0.0558)	-0.306 (1.594)	-0.307 (0.584)	0.524 (1.754)
No toilets	-0.1358 (0.0774)*	-0.2752 (0.2387)	-0.0562 (0.1002)	-2.925 (2.958)	0.628 (0.988)	1.485 (3.287)
Very little excreta	0.0247 (0.0222)	0.1039 (0.0671)	0.0494 (0.0279)*	0.709 (0.710)	0.370 (0.224)*	-1.040 (0.780)
Some excreta	-0.0320 (0.0209)	-0.2134 (0.0659)**	-0.0075 (0.0265)	-1.835 (0.690)**	0.462 (0.228)**	1.225 (0.766)
Much excreta	-0.2193 (0.0903)**	0.0839 (0.2800)	-0.0887 (0.1097)	-0.251 (2.588)	1.274 (0.821)	-1.846 (2.926)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	9162	9162	9162	9162	9162	9162
R-squared	0.2613	0.2606	0.2050	0.316	0.334	0.381
P-value for testing coefficients equal to zero						
Education	0.0039	0.0000	0.0052	0.0000	0.0000	0.0000
Prod Assets	0.0357	0.0020	0.0001	0.0001	0.0002	0.0000
Age dummies	0.0004	0.0001	0.0000	0.0001	0.0000	0.0000
Cohort dummies	0.0363	0.6476	0.4957	0.4179	0.0060	0.0468
Prices	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Water source	0.0064	0.0001	0.0111	0.0004	0.4180	0.0004
Toilet type	0.0000	0.0082	0.0049	0.2430	0.7888	0.5880
Excreta	0.0177	0.0345	0.1375	0.0118	0.0621	0.0814
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Also included in the models are individual-year age dummies, five-year-cohort dummies and community dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

2.5 Results: Determinants of Changes

During the rapid economic growth after economic reforms more health risks have resulted from deteriorating dieting habits (Guo, Popkin and Zhai 1999, Guo et al. 2000). Finding out the socioeconomic determinants of the change of people's food consumption has important policy implications. Here the changes in all outcomes between 1989 and 1991 and the changes between 1991 and 1993 are pooled together and OLS estimation is used in the basic and augmented models again to find out the effects of the changes in prices and other community characteristics on the changes in calorie, fat and protein intakes and the changes in percent of calories from fat, protein and carbohydrates.

2.5.1 Basic Models

When the outcomes are changes in calorie, fat and protein intakes and the percent of calories from fat, protein and carbohydrates the effects of the first set of regressors – education, productive assets and land together with age, cohort and community dummies – are mostly rendered insignificant, except for a few individual level variables and the community dummies. For females in the whole sample, the rural sample and the younger age group, the nonlinear effects of productive assets are quite strong. Recall in Table 4 there is no such relationship present for females. For men the levels of nutrient intakes tend to respond to the base productive asset holdings and for women the changes of nutrient intakes are more responsive to assets. In Table 13 for rural and younger females the effect of initial productive assets on changes in calorie intakes become significant and are of the same direction as the effect of productive assets on the level of calorie intakes for rural and younger men.

The land effects on changes in fat intake (Table 2.14) and changes in the percent of calories from fat (Table 2.16) become positive for all women and rural women

and on changes in percent calorie from carbohydrates (Table 2.18) become negative. This is puzzling. The effect of productive assets on changes in protein intake for younger female becomes of the same signs as the case for effects of productive assets on male protein level intake. Most relationships in Table 2.17 for changes in percent calorie from protein are insignificant except for some cohort dummies and community dummies.

Maybe the two changes measure are too close to each other, and/or taking differences on the outcomes measured with errors may exacerbate the problem and hence make it more difficult to render significant results. When a longer panel is available and outcomes are better measured the effects of some of the first set regressors should become significant again.

Table 2.11: Food Consumption in CHNS 89,91,93 with Community Characteristics in Urban Areas

Dependent Var	Urban Male					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Some primary edu	-0.0105 (0.0278)	-0.0107 (0.0748)	0.0052 (0.0329)	0.036 (0.916)	0.253 (0.285)	-0.945 (1.063)
Primary	0.0150 (0.0270)	0.0878 (0.0761)	0.0587 (0.0325)*	1.078 (0.945)	0.719 (0.301)**	-2.352 (1.096)**
Middle school	-0.0190 (0.0266)	0.0769 (0.0751)	0.0239 (0.0325)	1.106 (0.913)	0.706 (0.298)**	-2.391 (1.064)**
High school	-0.0251 (0.0279)	0.1250 (0.0781)	0.0287 (0.0340)	1.847 (0.957)*	0.898 (0.317)**	-3.756 (1.136)**
Tech/College+	-0.0213 (0.0292)	0.0967 (0.0780)	0.0490 (0.0343)	1.168 (0.971)	1.167 (0.325)**	-3.116 (1.125)**
Log real prod assets 1	0.0058 (0.0034)*	0.0092 (0.0083)	0.0065 (0.0041)	0.026 (0.117)	0 (0.041)	-0.063 (0.133)
Log real prod assets 2	-0.0005 (0.0033)	0 (0.0077)	-0.0022 (0.0040)	-0.010 (0.109)	-0.020 (0.040)	-0.063 (0.124)
Land	0.0308 (0.0181)*	0.0242 (0.0457)	0.0341 (0.0213)	-0.152 (0.493)	0.042 (0.130)	0.594 (0.635)
Log Price of Rice	-0.0473 (0.0349)	-0.1715 (0.0871)**	-0.1117 (0.0451)**	-1.754 (0.669)**	-0.763 (0.272)**	2.875 (0.778)**
Log Price of Pork	0.0624 (0.0432)	0.3767 (0.1127)**	0.0829 (0.0553)	0.743 (0.229)**	0.074 (0.085)	-0.939 (0.275)**
Log Price of Eggs	-0.0127 (0.0072)*	-0.0297 (0.0186)	-0.0148 (0.0088)*	-0.045 (0.115)	-0.021 (0.043)	-0.005 (0.134)
Log Price of Oil	-0.1477 (0.0629)**	-0.4042 (0.1683)**	-0.0722 (0.0802)	-0.937 (0.343)**	0.158 (0.122)	0.867 (0.387)**
Water Source 2	0.1044 (0.0848)	0.3974 (0.2478)	0.1613 (0.0958)*	4.868 (2.867)*	0.733 (0.806)	-6.214 (3.228)*
Water Source 3	0.2052 (0.1412)	0.2330 (0.3567)	-0.0315 (0.1722)	-0.571 (4.522)	-3.781 (1.711)**	4.148 (5.175)

Table continues

2.5.2 Augmented Models

Next, in addition to the above variables we study the effects of changes in food prices and community characteristics on the changes in outcomes. For the effects of prices the estimates that are significant in Table 2.19 remain of same signs as the effects of prices in the corresponding level estimation in Table 2.10. For changes of prices, only two relationships are still significant in the whole sample, the changes of the price of rice on the changes in male percent of calories from carbohydrates (negative) and female percent of calories from protein (positive).

Table 2.11 (cont'd)

Dependent Var	Urban Female					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Some primary edu	-0.0011 (0.0191)	0.0284 (0.0559)	0.0216 (0.0241)	-0.237 (0.676)	0.360 (0.215)*	0.002 (0.736)
Primary	0.0015 (0.0208)	0.1186 (0.0577)**	0.0333 (0.0255)	1.302 (0.717)*	0.486 (0.229)**	-1.646 (0.789)**
Middle school	-0.0239 (0.0202)	0.1173 (0.0579)**	0.0170 (0.0245)	1.813 (0.717)**	0.670 (0.227)**	-2.312 (0.791)**
High school	-0.0122 (0.0230)	0.1922 (0.0633)**	0.0401 (0.0282)	2.798 (0.798)**	0.869 (0.267)**	-3.485 (0.884)**
Tech/College+	0.0101 (0.0266)	0.1645 (0.0714)**	0.0856 (0.0323)**	2.254 (0.925)**	1.269 (0.290)**	-3.348 (1.006)**
Log real prod assets 1	0.0033 (0.0033)	0.0112 (0.0089)	0.0036 (0.0043)	0.154 (0.121)	0.001 (0.040)	-0.192 (0.138)
Log real prod assets 2	0.0004 (0.0028)	0.0060 (0.0073)	0.0025 (0.0038)	0.125 (0.101)	0.027 (0.038)	-0.124 (0.121)
Land	0.0090 (0.0129)	0.0130 (0.0249)	0.0184 (0.0137)	-0.103 (0.200)	0.164 (0.089)*	-0.066 (0.234)
Log Price of Rice	-0.0155 (0.0306)	-0.0494 (0.0776)	-0.0345 (0.0389)	-0.952 (0.685)	-0.330 (0.269)	1.356 (0.745)*
Log Price of Pork	0.0198 (0.0433)	0.3341 (0.1051)**	0.0023 (0.0538)	0.898 (0.225)**	-0.066 (0.079)	-0.821 (0.245)**
Log Price of Eggs	-0.0152 (0.0066)**	-0.0514 (0.0182)**	-0.0166 (0.0084)**	-0.144 (0.118)	-0.016 (0.042)	0.136 (0.127)
Log Price of Oil	-0.0878 (0.0520)*	-0.4409 (0.1437)**	0.0641 (0.0698)	-0.946 (0.316)**	0.439 (0.115)**	0.742 (0.348)**
Water Source 2	0.0628 (0.0727)	0.4225 (0.2309)*	0.0730 (0.1081)	5.117 (2.907)*	0.212 (0.949)	-6.146 (3.330)*
Water Source 3	0.1207 (0.1247)	0.3823 (0.3284)	-0.0636 (0.1533)	2.880 (4.222)	-2.850 (1.462)*	-1.164 (4.396)

Table continues

In urban areas the changes in the price of rice adversely affect the change in protein intakes for men. In rural areas the change in percent of calories from protein for female is negatively related with the changes in the price of rice. The changes of the price of pork do not have any statistically significant impact on any of the changes in outcomes. The effects of the changes in the prices of eggs remain strong and bear the same signs as in the level estimations. The changes in the prices of oils have negative impacts on changes in male calorie and protein intakes and changes in female calorie intakes in the whole sample and the rural areas. In the urban areas the impact of the change in the price of oil is positive for the change in percent of calories from protein consumption for female. In rural females, in addition, the impact of the

Table 2.11 (cont'd)

Dependent Var	Urban Male					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Water Source 4	-0.0185 (0.0458)	0.2667 (0.1128)**	-0.0184 (0.0562)	5.538 (1.345)**	-0.035 (0.475)	-4.791 (1.622)**
In house no flush	0.3390 (0.1495)**	0.9510 (0.3523)**	0.2899 (0.1863)	4.928 (4.317)	-0.265 (1.742)	0.709 (5.085)
Outside toilets	0.1653 (0.0950)*	0.2842 (0.2418)	0.0773 (0.1082)	0.044 (3.152)	-1.432 (1.211)	2.207 (3.600)
Open pit	0.1533 (0.1036)	0.4475 (0.2666)*	-0.0140 (0.1269)	4.362 (3.406)	-2.991 (1.279)**	-2.123 (4.029)
No toilets	0.0590 (0.1418)	0.6804 (0.3285)**	0.0279 (0.1649)	8.667 (3.973)**	-0.444 (1.548)	-9.786 (4.742)**
Very little excreta	0.0846 (0.0692)	0.2525 (0.1493)*	0.0815 (0.0843)	2.846 (1.862)	0.200 (0.670)	-5.144 (2.225)**
Some excreta	-0.3530 (0.0768)**	-0.7070 (0.1970)**	-0.2652 (0.0935)**	-4.272 (2.168)**	1.467 (0.654)**	3.178 (2.550)
Much excreta	-0.9336 (0.3917)**	-0.5271 (1.1049)	-1.6754 (0.4797)**	-10.357 (13.187)	-11.880 (3.926)**	34.373 (15.042)**
Community	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	2727	2727	2727	2727	2727	2727
R-squared	0.2052	0.1917	0.2185	0.212	0.245	0.259
P-value for testing coefficients equal to zero						
Education	0.3724	0.6362	0.2372	0.1475	0.0014	0.0037
Prod Assets	0.2176	0.7416	0.3072	0.9744	0.8263	0.5924
Age dummies	0.0000	0.0000	0.0000	0.0004	0.0182	0.0000
Cohort dummies	0.2508	0.0049	0.0558	0.0001	0.2528	0.0000
Prices	0.0613	0.0015	0.0276	0.0005	0.0225	0.0001
Water source	0.2814	0.1386	0.4408	0.0004	0.1027	0.0054
Toilet type	0.1254	0.1617	0.0784	0.0201	0.0180	0.0827
Excreta	0.0000	0.0003	0.0000	0.0430	0.0020	0.0035
Community	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

change in price of oil is strong and negative for the changes in the percent of calories from protein.

Significant effects of prices between urban and rural areas (details available upon request) on changes in macronutrients are similar. There is no implication for policy differentials between urban/rural residency although there are significant effects of the community dummies. This could be due to several reasons. One, prices as policy instruments may be more effective in changing people's consumption in specific foods, but not changes in total nutrient intakes. Two, in two years the changes in nutrient intakes are not dramatic and are measured with errors, which results in larger

Table 2.11 (cont'd)

Dependent Var	Urban Female					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Water Source 4	0.0692 (0.0394)*	0.3628 (0.1093)**	0.0912 (0.0532)*	5.691 (1.289)**	0.474 (0.488)	-6.623 (1.477)**
In house no flush	0.3333 (0.1377)**	0.9574 (0.3296)**	0.2509 (0.1659)	5.864 (4.433)	-0.890 (1.598)	-2.021 (4.636)
Outside toilets	0.1715 (0.0793)**	0.3777 (0.2267)*	0.1867 (0.1004)*	2.535 (3.070)	0.114 (1.134)	-1.979 (3.353)
Open pit	0.1285 (0.0893)	0.3945 (0.2731)	0.0293 (0.1235)	5.542 (3.602)	-1.728 (1.322)	-3.311 (4.142)
No toilets	-0.1662 (0.1196)	0.2695 (0.3327)	-0.1371 (0.1522)	4.988 (4.478)	0.285 (1.561)	-4.195 (5.057)
Very little excreta	0.0162 (0.0634)	0.1549 (0.1581)	-0.0076 (0.0811)	2.960 (1.963)	-0.290 (0.656)	-2.566 (2.064)
Some excreta	-0.1543 (0.0694)**	-0.4335 (0.1885)**	-0.0666 (0.0808)	-2.321 (2.287)	1.649 (0.642)**	0.049 (2.445)
Much excreta	-0.9837 (0.4562)**	-1.3301 (1.2628)	-1.4313 (0.5170)**	-16.275 (14.240)	-8.501 (3.823)**	29.393 (15.056)*
Community	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	3016	3016	3016	3016	3016	3016
R-squared	0.2121	0.1893	0.1846	0.219	0.253	0.268
P-value for testing coefficients equal to zero						
Education	0.5665	0.0774	0.0557	0.0024	0.0014	0.0005
Prod Assets	0.3634	0.0148	0.3173	0.0170	0.6089	0.0198
Age dummies	0.0790	0.0363	0.1718	0.0240	0.0017	0.0116
Cohort dummies	0.1250	0.1939	0.0797	0.2399	0.0446	0.0844
Prices	0.4402	0.0153	0.1704	0.0002	0.0039	0.0027
Water source	0.2248	0.0006	0.1752	0.0002	0.0892	0.0001
Toilet type	0.0059	0.2560	0.0377	0.0360	0.2486	0.4726
Excreta	0.0264	0.0466	0.0217	0.1676	0.0050	0.1496
Community	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Also included in the models are individual-year age dummies, five-year-cohort dummies and community dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

estimated standard errors and makes it harder to establish statistically significant relationships. Three, measurement errors in the changes in prices cause the estimates biased toward zero. Finally the authors in Guo, Popkin, Mroz and Zhai (1999) did not control for community level heterogeneity whereas there may be nonrandom program placement or selective migration taking place (Johnson 2003, Yaohui 1999).

Due to the stagnation of grain production and rapid increases in demand, the Chinese government implemented measures to promote grain production in 1988, including increased purchasing prices. These purchasing prices, which may have more

impact, are not available in the data set. Less than one hundred households moved in 1993 and most of them remained in the same community. The 24-hour recall estimates of food consumption may include intakes happening in a different community with different prices. All these caveats should be investigated further.

Changes in water sources at the community levels do not have significant impact on changes in outcomes, but the effects of changes in toilet types and excreta measures in a community remain strong. For men when the neighborhood has more household with little excreta there is an increase in calorie, fat, protein intake and decrease in percent calorie from carbohydrates. Similar results hold for women. There is no statistically significant difference between the parameter estimates for men and women. Men and women respond to environmental factors in similar fashions.

As we can see, not all determinants of the levels of nutrient intakes are significant determinants for the changes in nutrient intakes. When designing a policy the above results should bear some weight.

Table 2.12: Food Consumption in CHNS 89,91,93 with Community Characteristics in Rural Areas

Dependent Var	Rural Male					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Some primary edu	0.0098 (0.0159)	0.0887 (0.0482)*	0.0150 (0.0190)	0.721 (0.474)	0.061 (0.145)	-1.417 (0.617)**
Primary	0.0035 (0.0161)	0.0696 (0.0508)	-0.0024 (0.0196)	0.757 (0.496)	-0.098 (0.149)	-1.088 (0.635)*
Middle school	0.0012 (0.0161)	0.1881 (0.0508)**	0.0175 (0.0198)	2.061 (0.508)**	0.220 (0.151)	-3.236 (0.664)**
High school	-0.0179 (0.0193)	0.2009 (0.0590)**	0.0094 (0.0233)	2.362 (0.606)**	0.365 (0.185)**	-3.851 (0.792)**
Tech/College+	0.0147 (0.0299)	0.3127 (0.0807)**	0.0348 (0.0367)	3.531 (0.915)**	0.344 (0.356)	-6.479 (1.296)**
Log R prod assets 1	0.0127 (0.0027)**	0.0387 (0.0071)**	0.0171 (0.0031)**	0.260 (0.086)**	0.055 (0.027)**	-0.365 (0.108)**
Log R prod assets 2	-0.0082 (0.0026)**	-0.0139 (0.0072)*	-0.0074 (0.0031)**	-0.031 (0.082)	0.021 (0.024)	-0.048 (0.105)
Land	0.0024 (0.0011)**	-0.0034 (0.0033)	-0.0003 (0.0013)	-0.102 (0.032)**	-0.040 (0.010)**	0.113 (0.043)**
Log Price of Rice	-0.0713 (0.0204)**	-0.1294 (0.0617)**	-0.0880 (0.0253)**	-0.223 (0.467)	-0.130 (0.162)	1.988 (0.611)**
Log Price of Pork	-0.0123 (0.0322)	0.2089 (0.0989)**	0.0657 (0.0394)*	0.547 (0.156)**	0.148 (0.050)**	-0.773 (0.206)**
Log Price of Eggs	-0.0195 (0.0046)**	-0.0330 (0.0141)**	-0.0310 (0.0056)**	-0.032 (0.074)	-0.075 (0.024)**	0.014 (0.094)
Log Price of Oil	-0.0439 (0.0271)	-0.0913 (0.0866)	-0.0451 (0.0333)	-0.206 (0.158)	0.003 (0.054)	0.046 (0.191)
Water Source 2	0.0240 (0.0292)	0.2898 (0.0949)**	0.0607 (0.0359)*	3.486 (0.961)**	0.516 (0.294)*	-3.341 (1.186)**
Water Source 3	0.0816 (0.0401)**	0.2600 (0.1300)**	0.1603 (0.0501)**	1.261 (1.373)	1.108 (0.375)**	-3.047 (1.706)*

Table continues

2.6 Results: Production Function

In Chapter 1 we find strong persistency in a person's health status and it is quite reasonable to assume that lagged BMI is a good summary statistic of all past inputs and information. Here we estimate a simple health production function with health being measured by log weight or log BMI in 1993 (Table 2.20).

Lagged health (lagged log weight and height, or lagged log BMI in 1991), current physical activity levels and nutrient intakes are included and treated as endogenous. Also included in the regression are individual year age dummies. The identifying in-

Table 2.12 (cont'd)

Dependent Var	Rural Female					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Some primary edu	-0.0098 (0.0105)	0.0263 (0.0356)	-0.0058 (0.0136)	0.322 (0.362)	0.064 (0.117)	-0.332 (0.411)
Primary	0.0128 (0.0117)	0.1477 (0.0412)**	0.0369 (0.0154)**	1.417 (0.429)**	0.340 (0.133)**	-1.827 (0.494)**
Middle school	-0.0193 (0.0122)	0.1765 (0.0405)**	0.0174 (0.0159)	2.056 (0.429)**	0.499 (0.138)**	-2.520 (0.488)**
High school	-0.0305 (0.0191)	0.2101 (0.0600)**	0.0103 (0.0243)	2.854 (0.657)**	0.530 (0.184)**	-3.382 (0.738)**
Tech/College+	-0.0084 (0.0317)	0.2823 (0.0911)**	0.0452 (0.0428)	3.793 (1.185)**	0.788 (0.429)*	-4.420 (1.301)**
Log R prod assets 1	0.0044 (0.0023)*	0.0173 (0.0065)**	0.0087 (0.0029)**	0.057 (0.077)	0.056 (0.026)**	-0.141 (0.087)
Log R prod assets 2	-0.0013 (0.0022)	0.0022 (0.0065)	-0.0003 (0.0028)	0.135 (0.076)*	0.026 (0.026)	-0.154 (0.085)*
Land	0.0015 (0.0009)	-0.0024 (0.0032)	-0.0014 (0.0011)	-0.054 (0.037)	-0.046 (0.010)**	0.103 (0.043)**
Log Price of Rice	-0.0315 (0.0187)*	-0.1038 (0.0580)*	-0.0893 (0.0239)**	-0.140 (0.463)	-0.568 (0.154)**	0.710 (0.509)
Log Price of Pork	0.0048 (0.0285)	0.3062 (0.0946)**	0.1044 (0.0371)**	0.644 (0.153)**	0.206 (0.048)**	-0.813 (0.169)**
Log Price of Eggs	-0.0128 (0.0042)**	-0.0221 (0.0135)	-0.0203 (0.0053)**	-0.081 (0.070)	-0.030 (0.023)	0.090 (0.078)
Log Price of Oil	-0.0429 (0.0236)*	-0.1070 (0.0798)	-0.0273 (0.0304)	-0.246 (0.156)	0.054 (0.049)	0.231 (0.164)
Water Source 2	0.0449 (0.0264)*	0.2498 (0.0944)**	0.0566 (0.0357)	2.645 (0.961)**	0.151 (0.303)	-2.719 (1.087)**
Water Source 3	0.0746 (0.0419)*	0.1298 (0.1330)	0.0916 (0.0517)*	1.017 (1.352)	0.284 (0.390)	-1.884 (1.489)

Table continues

struments are education dummies, real productive asset splines, land cultivated and community characteristics including prices, water and sanitation conditions in 1991. For details and the goodness-of-fit of the first stage regressions see Appendix F. There is no weak instrument problem (Bound et al. 1995). The strong evidence of heteroscedasticity in male regressions suggests asymptotically correct inferences should be made using the Huber-White "sandwich" robust variance-covariance matrix. The overidentification χ^2 tests are heteroscedasticity-robust, based on Wooldridge (2002). The Wu-Hausman tests on the endogenous variables between OLS and 2SLS are carried out.

In OLS estimates, the effects of lagged log weight for men and women on current

Table 2.12 (cont'd)

Dependent Var	Rural Male					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Water Source 4	-0.0385 (0.0300)	0.0901 (0.0980)	0.0249 (0.0390)	1.048 (1.106)	0.883 (0.327)**	-1.350 (1.307)
In house no flush	0.1706 (0.0716)**	0.0440 (0.2172)	0.1659 (0.0954)*	-2.670 (2.507)	-1.069 (0.920)	6.881 (2.986)**
Outside toilets	0.0378 (0.0603)	-0.1742 (0.1558)	0.0222 (0.0821)	-3.863 (1.948)**	-0.919 (0.839)	6.920 (2.279)**
Open pit	-0.0122 (0.0567)	-0.1808 (0.1457)	0.0118 (0.0774)	-3.172 (1.901)*	-0.316 (0.816)	5.403 (2.219)**
No toilets	-0.0964 (0.1218)	-1.0618 (0.4210)**	0.0312 (0.1693)	-9.855 (4.419)**	1.345 (1.531)	18.776 (5.601)**
Very little excreta	0.0905 (0.0262)**	0.1361 (0.0778)*	0.1215 (0.0320)**	0.551 (0.793)	0.444 (0.254)*	-0.725 (1.018)
Some excreta	-0.0310 (0.0249)	-0.1620 (0.0735)**	-0.0017 (0.0308)	-1.142 (0.758)	0.512 (0.264)*	1.664 (0.969)*
Much excreta	0.0514 (0.1039)	0.3447 (0.2996)	0.0038 (0.1224)	-2.357 (2.652)	-0.642 (0.846)	-1.724 (3.666)
Community	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	5688	5688	5688	5688	5688	5688
R-squared	0.2355	0.2410	0.2300	0.282	0.292	0.319
P-value for testing coefficients equal to zero						
Education	0.4886	0.0019	0.7070	0.0000	0.0175	0.0000
Prod Assets	0.0000	0.0001	0.0000	0.0001	0.0004	0.0000
Age dummies	0.0030	0.3526	0.0660	0.1572	0.0965	0.3658
Cohort dummies	0.1665	0.4307	0.4466	0.1901	0.8196	0.6653
Prices	0.0000	0.0091	0.0000	0.0141	0.0010	0.0004
Water source	0.0344	0.0105	0.0151	0.0032	0.0077	0.0197
Toilet type	0.0439	0.6214	0.5829	0.1957	0.1476	0.0063
Excreta	0.0006	0.0285	0.0041	0.1794	0.1255	0.1303
Community	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

weight are significantly less than 1, the effects of moderate or heavy physical activities are negative, and the joint effects of all nutrient intakes are significant for men at 0.05 level and for women at 0.1 level. In the 2SLS estimates, however, the effect of lagged log weight for men is not significantly different from 1, and for women it is only marginally significant at 0.1 level. Conditional on lagged weight and height, current physical activities and nutrient inputs do not have any significant impact. The coefficients for nutrient inputs are generally larger in 2SLS regressions, indicating attenuation biases in OLS regressions.

The fact that the coefficients for lagged weight are one suggests that in a short pe-

riod of two years from 1991 to 1993, adult weight can be modelled as a random walk process. The Wu-Hausman F-statistics indicates significant differences on endogenous variables between OLS and 2SLS estimates. The overidentification tests lead to the conclusion of validity of instruments assuming that eight of all the identifying instruments are exogenous.

The results for log BMI in 1993 are very similar to the results for log weight. BMI can also be characterized as a random walk process in a short period of time.

2.7 Discussions

Using an unique longitudinal data set of adult Chinese in the early 1990s we estimate the reduced demand function for calorie, fat and protein intakes and for percent of calories from fat, protein and carbohydrates. We also study the determinants for the changes in these outcomes. The socioeconomic factors that are significantly related to nutrient consumption may not be related in the same way to the changes in nutrient consumption.

The theoretical framework of this paper is the same as in Luo (2003*b*) and similar to Foster (1995). The key findings of this paper are that education does not have significant impact on calorie intakes but does affect percent of calories from fat, from protein and from carbohydrates differently in different region and at different age. The effect of productive assets is nonlinear and in inverted U-shape for male calorie, fat and protein intakes. In rural areas the effect of productive assets is stronger than that in the urban areas. At different ages the effects of education and assets are different. The effects of prices on calorie, fat and protein intakes and the quality of diet measures can go either way. Improvements in sanitation are associated with more energy and protein intakes in urban areas. Changes in prices and community characteristics are related with the changes in nutrient consumption in different ways. The difference between urban and rural areas does not seem strong. There is a trade-off between including community dummies or not when we use current community characteristics in the model. The advantage is community dummies capture unobserved heterogeneity at the aggregate level. The disadvantage is any effect of community variables will have to be identified through the variation of them over time and in a short run the change may be too little for good identification.

There are some limitations in this paper. First, using the productive asset in the reduced model as a proxy for income and treating it as exogenous is not without criticism. In a dynamic model nutrient intakes can affect productivity and therefore the

holding of productive assets in later periods. If there was selection into more strenuous vocations with higher wages, the results will be biased upwards. Attenuation bias could also rise with imperfect measures of productive assets and land.

The model is based on the assumption of perfect insights. Ignoring uncertainty is not innocuous when there are risks other than idiosyncratic shocks. Uncertainty would be important if overweight increases risk of mortality and morbidity and people are well aware of it. The length of individual life time and the household size are all assumed to be exogenous whereas in reality both are potentially endogenous.

Food prices may be endogenous in the augmented models. It was plausible that food prices modify household income constraint and the estimated price effect may be biased.

The paper focus on macronutrients instead of specific food consumptions. Effective policies may be a little difficult to deduct from here due to the ambiguity of the effects of many of the determinants.

The corner solutions and loss to follow-up in the interview of three-day dietary recall are not considered in this paper.

Keeping these caveats in mind this paper finds that certain individual macronutrient intakes are strongly correlated with education levels and household resources. Food consumption is very sensitive to food prices and the direction can go either way. In designing food price or subsidy programs the local government should start with those less sensitive factors and proceed with care.

Table 2.12 (cont'd)

Dependent Var	Rural Female					
	Calorie Log(kcal)	Fat Log(g)	Protein Log(g)	% Calorie From		
				Fat	Protein	Carbo
Water Source 4	-0.0034 (0.0285)	0.0578 (0.0956)	0.0129 (0.0379)	0.256 (1.063)	0.300 (0.323)	-0.801 (1.175)
In house no flush	0.2081 (0.0738)**	0.2600 (0.2281)	0.2113 (0.0995)**	0.352 (2.689)	-0.914 (0.884)	0.775 (2.995)
Outside toilets	0.0654 (0.0550)	-0.1243 (0.1417)	0.0271 (0.0710)	-2.682 (1.849)	-0.981 (0.712)	3.715 (2.041)*
Open pit	-0.0037 (0.0526)	-0.1876 (0.1349)	-0.0210 (0.0682)	-2.909 (1.825)	-0.638 (0.683)	3.309 (2.010)*
No toilets	-0.1042 (0.1168)	-1.0122 (0.3646)**	-0.0059 (0.1503)	-14.044 (4.172)**	0.778 (1.351)	9.995 (4.705)**
Very little excreta	0.0393 (0.0240)	0.1206 (0.0757)	0.0723 (0.0301)**	0.459 (0.777)	0.505 (0.240)**	-0.923 (0.863)
Some excreta	-0.0089 (0.0220)	-0.1554 (0.0711)**	0.0080 (0.0285)	-1.582 (0.718)**	0.349 (0.248)	1.174 (0.810)
Much excreta	-0.1827 (0.0924)**	0.1212 (0.2903)	-0.0106 (0.1127)	-0.056 (2.645)	1.907 (0.845)**	-3.024 (3.014)
Community	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	6146	6146	6146	6146	6146	6146
R-squared	0.2635	0.2517	0.2335	0.288	0.306	0.336
P-value for testing coefficients equal to zero						
Education	0.0266	0.0001	0.0768	0.0000	0.0015	0.0000
Prod Assets	0.0757	0.0586	0.0001	0.0037	0.0002	0.0000
Age dummies	0.0004	0.0232	0.0000	0.8066	0.0000	0.0000
Cohort dummies	0.0724	0.7805	0.3657	0.7650	0.0869	0.4639
Prices	0.0002	0.0039	0.0000	0.0010	0.0000	0.0001
Water source	0.0299	0.0055	0.0486	0.0296	0.7995	0.0690
Toilet type	0.0001	0.0044	0.0486	0.0208	0.4874	0.2257
Excreta	0.0674	0.2190	0.1194	0.0788	0.0230	0.1273
Community	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Also included in the models are individual-year age dummies, five-year-cohort dummies and community dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level. The percent of households with the following water sources: (1) underground, (2) open well, (3) spring, river lake, rain or snow, and (4) water factory. Omitted is the first source. The omitted category for excreta is percent of no excreta.

Table 2.13: Changes in Daily Calorie Intakes

Variables	Male					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-22.004 (71.719)	83.932 (136.169)	-45.505 (85.809)	-26.332 (152.715)	150.730 (131.130)	-375.371 (168.039)**
Primary	4.031 (68.454)	96.633 (132.229)	-17.352 (81.802)	88.343 (144.537)	45.911 (126.826)	101.913 (186.970)
Middle school	29.540 (68.141)	146.344 (129.666)	4.105 (80.940)	117.495 (137.961)	100.360 (125.540)	-100.590 (241.401)
High school	55.567 (76.368)	101.551 (139.463)	66.382 (94.143)	137.982 (142.893)	140.372 (175.334)	148.778 (387.714)
Tech/College+	16.461 (83.286)	91.758 (133.748)	73.337 (138.261)	173.125 (158.680)	12.661 (141.672)	-379.461 (331.521)
Log real prod assets 1	12.295 (12.880)	22.400 (24.651)	10.382 (15.244)	10.171 (17.465)	13.924 (27.419)	-47.620 (52.737)
Log real prod assets 2	-9.946 (12.398)	-28.072 (26.161)	-6.517 (14.133)	-9.745 (16.572)	-8.653 (25.697)	42.328 (50.967)
Land	-1.425 (6.555)	17.909 (84.496)	-2.637 (6.667)	-4.757 (8.882)	3.730 (12.278)	-17.942 (35.433)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4326	1360	2966	2537	1309	397
R-squared	0.222	0.245	0.230	0.268	0.343	0.528
P-value for testing coefficients equal to zero						
Education	0.8294	0.8959	0.7156	0.3166	0.6533	0.1088
Assets	0.6338	0.5606	0.7813	0.8302	0.8751	0.6597
Age dummies	0.5463	0.0000	0.1507	0.7125	0.4292	0.2467
Cohort dummies	0.1656	0.2532	0.4874	0.2787	0.2711	0.6772
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.9318	0.8667	0.7461	0.3743	0.3322	0.4441
Assets	0.3019	0.8424	0.3009	0.5839	0.4602	0.8616
Age dummies	0.6728	0.0191	0.3084	0.7305	0.4014	0.4487
Cohort dummies	0.2744	0.0914	0.9517	0.4830	0.4976	0.0000
Community dummies	0.3824	0.9552	0.0454	0.0000	0.0000	0.0000
Table continues						

Table continues

Table 2.13 (cont'd)

Variables	Overall	Female			
		Urban	Rural	Age 20-39	Age 40-59
Some primary education	0.723 (37.979)	-6.294 (82.138)	-8.519 (43.678)	11.682 (60.438)	34.280 (66.951)
Primary	-32.765 (41.651)	-13.252 (85.178)	-51.783 (48.903)	19.292 (63.954)	-79.387 (73.324)
Middle school	18.762 (39.895)	-6.764 (76.409)	20.730 (47.949)	12.544 (58.00)	54.244 (83.616)
High school	8.970 (52.205)	28.728 (91.534)	-20.429 (67.511)	-8.985 (68.795)	90.523 (157.470)
Tech/College+	24.551 (67.594)	33.061 (93.752)	-49.056 (134.949)	-50.293 (90.502)	256.468 (135.595)*
Log real prod assets 1	4.938 (8.498)	6.925 (15.304)	2.936 (10.142)	15.798 (11.325)	-4.694 (17)
Log real prod assets 2	-18.124 (8.803)**	-20.166 (20.669)	-16.662 (9.907)*	-26.272 (11.621)**	-18.711 (18.656)
Land	1.202 (5.619)	2.408 (50.310)	2.022 (5.677)	6.929 (7.329)	-5.338 (11.937)
Community *year dummy	Yes	Yes	Yes	Yes	Yes
No. of Observations	4835	1515	3320	2904	1450
R-squared	0.216	0.233	0.223	0.273	0.326
P-value for testing coefficients equal to zero					
Education	0.8257	0.9857	0.7054	0.9651	0.1743
Assets	0.0546	0.6055	0.0932	0.0744	0.2826
Age dummies	0.7343	0.0444	0.4124	0.5369	0.8778
Cohort dummies	0.0039	0.0000	0.0048	0.3825	0.2266
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas					
Education	Men	0.7953	Women	0.9224	
Assets		0.7336		0.9762	
Age dummies		0.0194		0.1693	
Cohort dummies		0.3638		0.0189	

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.14: Changes in Daily Fat Intakes

Variables	Male					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-3.135 (3.998)	4.098 (7.784)	-4.399 (4.805)	0.977 (6.250)	-0.829 (9.110)	-12.643 (8.903)
Primary	-4.221 (3.730)	0.157 (7.638)	-5.085 (4.365)	3.622 (6.097)	-8.234 (8.325)	2.903 (10.690)
Middle school	-2.102 (3.695)	5.919 (7.203)	-3.939 (4.337)	4.906 (5.748)	-5.608 (7.886)	12.517 (12.913)
High school	-1.768 (4.148)	3.295 (7.590)	-2.639 (5.068)	5.052 (6.013)	1.849 (11.786)	14.029 (24.651)
Tech/College+	-4.819 (4.515)	1.218 (7.321)	-2.985 (7.703)	9.760 (7.095)	-14.339 (9.623)	-27.547 (13.405)**
Log real prod assets 1 ^a	0.192 (0.649)	-0.485 (1.138)	0.372 (0.792)	0.123 (0.815)	0.126 (1.657)	-2.112 (2.518)
Log real prod assets 2 ^a	-0.156 (0.623)	0.477 (1.334)	-0.305 (0.719)	-0.063 (0.795)	-0.307 (1.466)	2.086 (2.550)
Land	0.184 (0.294)	1.011 (5.549)	0.142 (0.291)	0.024 (0.359)	0.205 (0.584)	1.692 (1.610)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4326	1360	2966	2537	1309	397
R-squared	0.177	0.187	0.192	0.237	0.268	0.604
P-value for testing coefficients equal to zero						
Education	0.7609	0.8083	0.8967	0.6488	0.2167	0.0671
Assets	0.9572	0.9116	0.8938	0.9857	0.9645	0.6757
Age dummies	0.5201	0.0284	0.8332	0.1485	0.6929	0.3858
Cohort dummies	0.4186	0.6112	0.7617	0.1391	0.7766	0.2124
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.3500	0.4911	0.6207	0.7795	0.3326	0.1357
Assets	0.2697	0.8458	0.2960	0.2453	0.8225	0.7691
Age dummies	0.6737	0.0152	0.9315	0.4602	0.4576	0.7734
Cohort dummies	0.7481	0.5976	0.9913	0.3482	0.5393	0.1952
Community dummies	1.0000	1.0000	0.9998	0.0024	0.0000	0.0000

Table continues

Table 2.14 (cont'd)

Variables	Female					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-1.067 (2.021)	-2.800 (4.566)	-0.426 (2.288)	-0.147 (3.147)	-1.438 (3.833)	2.046 (9.267)
Primary	0.663 (2.274)	-1.606 (5.120)	1.175 (2.548)	3.440 (3.401)	-1.912 (4.158)	-20.128 (9.959)**
Middle school	-0.037 (2.163)	-2.568 (4.057)	0.529 (2.573)	-0.045 (3.104)	-0.550 (4.850)	26.097 (21.063)
High school	-1.652 (3.019)	-0.996 (4.930)	-4.011 (4.079)	-0.538 (3.960)	-7.205 (9.836)	-24.799 (17.361)
Tech/College+	5.482 (3.962)	3.524 (5.424)	5.934 (7.720)	3.956 (5.334)	4.517 (8.549)	-23.507 (31.204)
Log real prod assets 1 ^a	-0.381 (0.475)	-1.059 (0.791)	-0.279 (0.572)	-0.302 (0.646)	-0.917 (1.015)	-0.213 (1.528)
Log real prod assets 2 ^a	-0.423 (0.479)	0.500 (1.012)	-0.589 (0.556)	-0.777 (0.659)	0.102 (0.993)	-0.057 (1.710)
Land	0.581 (0.277)**	3.539 (1.915)*	0.583 (0.279)**	0.403 (0.376)	0.655 (0.593)	-1.828 (1.590)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4835	1515	3320	2904	1450	398
R-squared	0.181	0.210	0.182	0.230	0.288	0.459
P-value for testing coefficients equal to zero						
Education	0.5397	0.7661	0.7186	0.6820	0.9086	0.0894
Assets	0.0701	0.3569	0.0898	0.0741	0.4782	0.9712
Age dummies	0.4778	0.0211	0.6701	0.2877	0.7027	0.8578
Cohort dummies	0.1376	0.0000	0.0690	0.2632	0.3667	0.7798
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.8208	Women	0.8891		
Assets		0.8216		0.6209		
Age dummies		0.5012		0.2500		
Cohort dummies		0.0990		0.0088		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.15: Changes in Daily Protein Intakes

Variables	Overall	Male				
		Rural	Urban	Age 20-39	Age 40-59	Age 60+
Some primary education	1.870 (3.044)	1.254 (3.523)	5.215 (6.218)	1.978 (6.318)	6.617 (5.175)	-9.879 (7.922)
Primary	1.140 (2.960)	1.397 (3.406)	0.102 (6.389)	4.659 (6.255)	-0.323 (5.130)	5.900 (9.447)
Middle school	4.131 (2.967)	3.530 (3.416)	7.168 (6.038)	6.784 (5.852)	8.958 (5.263)*	-4.656 (11.019)
High school	5.765 (3.491)*	4.919 (4.244)	8.722 (6.618)	7.775 (6.273)	11.515 (8.436)	1.832 (18.370)
Tech/College+	-0.549 (3.896)	1.670 (6.374)	0.510 (6.315)	6.490 (7.005)	-5.988 (6.559)	-15.597 (19.268)
Log real prod assets 1 ^a	0.273 (0.589)	0.070 (0.698)	1.049 (1.127)	0.642 (0.825)	-0.289 (1.077)	-3.547 (2.391)
Log real prod assets 2 ^a	-0.483 (0.560)	-0.228 (0.636)	-1.587 (1.264)	-0.532 (0.781)	-0.032 (1.068)	1.728 (2.317)
Land	-0.020 (0.287)	-0.005 (0.291)	0.340 (3.625)	-0.029 (0.384)	-0.383 (0.561)	-0.975 (1.421)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4326	2966	1360	2537	1309	397
R-squared	0.217	0.220	0.247	0.263	0.331	0.491
P-value for testing coefficients equal to zero						
Education	0.2069	0.7619	0.2462	0.5179	0.0268	0.5536
Assets	0.6393	0.8882	0.4549	0.7365	0.9162	0.2508
Age dummies	0.2006	0.6095	0.0235	0.2663	0.3372	0.4515
Cohort dummies	0.0211	0.1489	0.1181	0.0136	0.7762	0.2401
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.5051	0.7839	0.4119	0.3497	0.6468	0.4107
Assets	0.6578	0.6373	0.8507	0.6875	0.4210	0.5087
Age dummies	0.5148	0.4623	0.3856	0.8224	0.2098	0.7942
Cohort dummies	0.4269	0.9446	0.3621	0.4104	0.0966	0.1140
Community dummies	0.9976	0.8189	0.9358	0.0093	0.0000	0.0000

Table continues

Table 2.15 (cont'd)

Variables	Female					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-0.048 (1.904)	-1.171 (4.312)	0.161 (2.151)	1.586 (2.928)	0.264 (3.448)	-3.583 (8.806)
Primary	-1.666 (2.056)	-2.299 (4.598)	-1.926 (2.370)	1.732 (3.190)	-3.017 (3.543)	-17.968 (14.132)
Middle school	-0.362 (1.939)	-3.218 (4.115)	0.106 (2.275)	-0.304 (2.729)	2.693 (4.435)	13.111 (13.145)
High school	-1.604 (2.646)	-2.417 (4.919)	-2.409 (3.391)	-1.711 (3.397)	6.953 (7.692)	-25.004 (19.383)
Tech/College+	-1.179 (3.560)	-1.571 (5.159)	-4.838 (6.816)	-0.483 (4.714)	2.196 (8.601)	-31.487 (23.569)
Log real prod assets 1 ^a	0.189 (0.449)	0.248 (0.852)	0.135 (0.529)	1.114 (0.644)*	-0.830 (0.841)	-1.472 (1.506)
Log real prod assets 2 ^a	-0.827 (0.446)*	-0.958 (1.076)	-0.739 (0.500)	-1.356 (0.646)**	-0.850 (0.845)	1.610 (1.666)
Land	0.320 (0.281)	-1.474 (2.487)	0.421 (0.281)	0.515 (0.380)	-0.203 (0.549)	0.296 (1.492)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4835	1515	3320	2904	1450	398
R-squared	0.192	0.200	0.205	0.242	0.327	0.498
P-value for testing coefficients equal to zero						
Education	0.9516	0.9797	0.8454	0.8539	0.7597	0.3055
Assets	0.0681	0.6301	0.1342	0.1073	0.0380	0.5884
Age dummies	0.0865	0.0983	0.3168	0.4999	0.5265	0.0462
Cohort dummies	0.0000	0.0000	0.0275	0.5177	0.1127	0.3327
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.8734	Women	0.9212		
Assets		0.6243		0.9826		
Age dummies		0.3324		0.5707		
Cohort dummies		0.7158		0.2107		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.16: Changes in Percent Daily Calorie From Fat

Variables	Overall	Male				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-0.932 (0.811)	0.137 (1.800)	-0.990 (0.906)	-0.025 (1.381)	-0.246 (1.552)	-2.119 (2.414)
Primary	-0.961 (0.788)	-0.316 (1.745)	-0.921 (0.882)	0.484 (1.382)	-1.221 (1.458)	-0.743 (2.828)
Middle school	-0.485 (0.772)	0.987 (1.649)	-0.717 (0.871)	0.621 (1.277)	-0.743 (1.453)	3.882 (3.255)
High school	-0.364 (0.890)	0.464 (1.711)	-0.413 (1.076)	0.746 (1.362)	1.178 (2.339)	1.809 (6.353)
Tech/College+	-1.290 (1.004)	-0.385 (1.695)	-0.819 (1.735)	1.443 (1.627)	-2.941 (1.861)	-6.549 (3.668)*
Log real prod assets 1	-0.035 (0.134)	-0.117 (0.253)	-0.014 (0.161)	-0.056 (0.189)	-0.249 (0.270)	-0.081 (0.666)
Log real prod assets 2	0.030 (0.136)	0.157 (0.297)	-0.003 (0.157)	0.029 (0.185)	0.118 (0.275)	0.307 (0.708)
Land	0.043 (0.072)	0.048 (1.122)	0.040 (0.071)	-0.011 (0.085)	0.084 (0.130)	0.331 (0.500)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4326	1360	2966	2537	1309	397
R-squared	0.190	0.202	0.206	0.234	0.327	0.504
P-value for testing coefficients equal to zero						
Education	0.5992	0.7525	0.8970	0.8985	0.3228	0.1983
Assets	0.9653	0.8633	0.9858	0.9479	0.5952	0.8783
Age dummies	0.8216	0.1053	0.1636	0.2944	0.6830	0.7597
Cohort dummies	0.6013	0.4573	0.0032	0.4316	0.8877	0.2580
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.1676	0.6282	0.4043	0.7757	0.6992	0.1914
Assets	0.3101	0.8819	0.2959	0.3237	0.9937	0.6329
Age dummies	0.9468	0.1100	0.8224	0.8232	0.6532	0.8786
Cohort dummies	0.7310	0.8936	0.9875	0.5600	0.9048	0.4012
Community dummies	0.8723	0.9974	0.5043	0.0000	0.0000	0.0000

Table continues

Table 2.16 (cont'd)

Variables	Overall	Female				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-0.116 (0.534)	-0.432 (1.205)	0.083 (0.601)	-0.136 (0.816)	-0.237 (1.012)	2.269 (2.820)
Primary	0.232 (0.586)	-0.682 (1.311)	0.510 (0.658)	0.661 (0.864)	-0.030 (1.080)	-7.510 (3.286)**
Middle school	-0.164 (0.566)	-0.729 (1.123)	-0.022 (0.663)	-0.326 (0.804)	-0.279 (1.299)	4.918 (5.114)
High school	-0.786 (0.758)	-0.883 (1.300)	-1.195 (1.016)	-0.480 (0.979)	-2.484 (2.641)	-8.038 (6.328)
Tech/College+	1.267 (1.026)	0.352 (1.477)	1.991 (1.841)	1.404 (1.389)	-1.313 (2.301)	-3.916 (5.043)
Log real prod assets 1	-0.170 (0.129)	-0.265 (0.223)	-0.176 (0.153)	-0.238 (0.172)	-0.284 (0.273)	0.592 (0.548)
Log real prod assets 2	-0.030 (0.131)	0.212 (0.274)	-0.069 (0.151)	-0.043 (0.177)	0.128 (0.271)	-0.588 (0.623)
Land	0.190 (0.074)**	1.103 (0.508)**	0.187 (0.075)**	0.107 (0.095)	0.276 (0.157)*	-0.517 (0.539)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4835	1515	3320	2904	1450	398
R-squared	0.191	0.219	0.193	0.238	0.301	0.463
P-value for testing coefficients equal to zero						
Education	0.4016	0.8998	0.4558	0.4553	0.9598	0.0679
Assets	0.0756	0.4924	0.0639	0.0561	0.5270	0.5501
Age dummies	0.2893	0.0397	0.5890	0.4061	0.3691	0.7376
Cohort dummies	0.2198	0.0009	0.3252	0.4197	0.8006	0.7726
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.8993	Women	0.8881		
Assets		0.8911		0.6046		
Age dummies		0.1542		0.3679		
Cohort dummies		0.0014		0.2814		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.17: Changes in Percent Daily Calorie From Protein

Variables	Male					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.260 (0.253)	0.402 (0.569)	0.284 (0.282)	0.134 (0.451)	0.138 (0.418)	0.725 (0.825)
Primary	0.089 (0.243)	-0.471 (0.550)	0.273 (0.269)	0.204 (0.462)	-0.259 (0.403)	0.464 (0.986)
Middle school	0.318 (0.239)	0.168 (0.519)	0.403 (0.266)	0.197 (0.422)	0.692 (0.414)*	-0.539 (1.121)
High school	0.402 (0.275)	0.499 (0.551)	0.315 (0.326)	0.209 (0.452)	0.724 (0.681)	0.438 (1.949)
Tech/College+	-0.142 (0.356)	-0.330 (0.565)	-0.092 (0.608)	0.103 (0.577)	-0.914 (0.618)	-0.255 (1.610)
Log real prod assets 1	-0.037 (0.048)	0.008 (0.095)	-0.056 (0.056)	-0.007 (0.066)	-0.077 (0.088)	-0.255 (0.285)
Log real prod assets 2	-0.025 (0.044)	-0.064 (0.107)	-0.007 (0.049)	-0.007 (0.061)	-0.016 (0.083)	0.064 (0.303)
Land	-0.014 (0.021)	-0.146 (0.256)	-0.004 (0.021)	0.003 (0.028)	-0.059 (0.045)	-0.094 (0.172)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4326	1360	2966	2537	1309	397
R-squared	0.224	0.243	0.234	0.268	0.355	0.480
P-value for testing coefficients equal to zero						
Education	0.2967	0.1263	0.6925	0.9968	0.0215	0.8488
Assets	0.1325	0.7350	0.1894	0.9481	0.3811	0.5341
Age dummies	0.0156	0.1790	0.0260	0.1102	0.0334	0.9976
Cohort dummies	0.0003	0.0424	0.0001	0.0080	0.2369	0.2193
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.4142	0.3020	0.7278	0.6600	0.8562	0.9618
Assets	0.7295	0.8493	0.6161	0.6488	0.9254	0.8424
Age dummies	0.1736	0.3889	0.0192	0.4026	0.3675	0.7795
Cohort dummies	0.1069	0.6199	0.2318	0.0544	0.2916	0.0004
Community dummies	0.6687	0.9512	0.1420	0.0000	0.0000	0.0000
Table continues						

Table continues

Table 2.17 (cont'd)

Variables	Female					
	Overall	Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-0.007 (0.180)	0.001 (0.439)	0.039 (0.202)	0.066 (0.271)	-0.131 (0.322)	0.961 (1.013)
Primary	-0.125 (0.192)	-0.181 (0.445)	-0.115 (0.216)	0.048 (0.286)	-0.060 (0.331)	-0.527 (1.455)
Middle school	-0.159 (0.191)	-0.390 (0.414)	-0.125 (0.222)	-0.218 (0.254)	0.217 (0.436)	-0.102 (1.511)
High school	-0.328 (0.255)	-0.564 (0.490)	-0.265 (0.311)	-0.297 (0.314)	0.272 (0.701)	-3.581 (5.313)
Tech/College+	-0.337 (0.382)	-0.377 (0.550)	-0.521 (0.693)	0.316 (0.476)	-1.468 (0.889)*	-0.650 (1.890)
Log real prod assets 1	-0.008 (0.044)	-0.044 (0.090)	0.006 (0.052)	0.074 (0.066)	-0.105 (0.079)	-0.164 (0.197)
Log real prod assets 2	-0.021 (0.042)	0.017 (0.096)	-0.027 (0.048)	-0.048 (0.062)	-0.024 (0.075)	0.118 (0.207)
Land	0.027 (0.025)	-0.159 (0.208)	0.036 (0.025)	0.031 (0.034)	-0.026 (0.047)	0.231 (0.191)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4835	1515	3320	2904	1450	398
R-squared	0.181	0.184	0.201	0.219	0.319	0.418
P-value for testing coefficients equal to zero						
Education	0.8016	0.8558	0.8972	0.5040	0.4310	0.8555
Assets	0.6281	0.8732	0.7405	0.5225	0.0898	0.7017
Age dummies	0.2391	0.8643	0.0000	0.6139	0.1824	0.3471
Cohort dummies	0.1878	0.6950	0.0000	0.1632	0.6716	0.4710
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	Men	0.4317	Women	0.9828		
Assets		0.8394		0.8859		
Age dummies		0.9503		0.0733		
Cohort dummies		0.9550		0.0000		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.18: Changes in Percent Daily Calorie From Carbohydrates

Variables	Overall	Male				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	0.928 (0.963)	0.039 (2.057)	0.685 (1.097)	0.388 (1.644)	0.737 (1.838)	2.286 (2.755)
Primary	1.206 (0.939)	0.865 (2.033)	0.796 (1.057)	0.012 (1.633)	2.173 (1.743)	1.522 (3.086)
Middle school	0.621 (0.916)	-0.361 (1.914)	0.453 (1.044)	-0.157 (1.503)	0.561 (1.769)	-1.198 (3.482)
High school	0.330 (1.050)	-0.583 (2.011)	0.256 (1.275)	-0.343 (1.626)	-0.924 (2.654)	-0.472 (6.960)
Tech/College+	2.898 (1.197)**	2.349 (2.005)	2.446 (2.126)	-0.096 (1.912)	6.228 (2.220)**	10.671 (4.348)**
Log real prod assets 1	0.078 (0.162)	0.183 (0.305)	0.049 (0.195)	0.122 (0.211)	0.013 (0.373)	0.698 (0.661)
Log real prod assets 2	0.034 (0.163)	-0.037 (0.337)	0.046 (0.190)	0.020 (0.212)	0.108 (0.364)	-0.950 (0.760)
Land	-0.037 (0.087)	0.151 (1.348)	-0.047 (0.088)	0.015 (0.107)	-0.149 (0.168)	0.205 (0.675)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4326	1360	2966	2537	1309	397
R-squared	0.211	0.213	0.228	0.259	0.333	0.533
P-value for testing coefficients equal to zero						
Education	0.0898	0.1680	0.8599	0.9920	0.0079	0.2015
Assets	0.5641	0.7335	0.7394	0.5738	0.8817	0.4399
Age dummies	0.5432	0.2395	0.2392	0.3086	0.6770	0.2204
Cohort dummies	0.4559	0.1680	0.0231	0.2134	0.9144	0.0981
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between men and women						
Education	0.0164	0.1791	0.3060	0.4792	0.4333	0.1912
Assets	0.6941	0.9430	0.6038	0.8370	0.7511	0.2836
Age dummies	0.9138	0.3257	0.9348	0.8293	0.8766	0.3114
Cohort dummies	0.5679	0.7125	0.9811	0.5531	0.7388	0.1036
Community dummies	0.1690	0.9884	0.0143	0.0000	0.0000	0.0000

Table continues

Table 2.18 (cont'd)

Variables	Overall	Female				
		Urban	Rural	Age 20-39	Age 40-59	Age 60+
Some primary education	-0.032 (0.595)	0.087 (1.302)	-0.264 (0.678)	-0.225 (0.905)	0.010 (1.100)	-2.707 (3.015)
Primary	-0.212 (0.629)	0.336 (1.340)	-0.447 (0.725)	-0.846 (0.946)	-0.288 (1.170)	7.063 (3.160)**
Middle school	0.258 (0.620)	0.760 (1.222)	0.079 (0.730)	0.327 (0.878)	-0.093 (1.449)	-3.903 (4.698)
High school	1.083 (0.821)	0.999 (1.402)	1.574 (1.108)	0.672 (1.060)	1.803 (2.731)	10.429 (6.528)
Tech/College+	-1.051 (1.063)	-0.380 (1.567)	-1.573 (1.899)	-1.818 (1.457)	2.376 (2.423)	6.408 (3.781)*
Log real prod assets 1	0.161 (0.138)	0.304 (0.242)	0.158 (0.164)	0.181 (0.187)	0.328 (0.291)	-0.493 (0.560)
Log real prod assets 2	0.076 (0.141)	-0.181 (0.299)	0.106 (0.163)	0.078 (0.192)	-0.014 (0.298)	0.550 (0.620)
Land	-0.195 (0.085)**	-1.057 (0.514)**	-0.197 (0.087)**	-0.135 (0.111)	-0.207 (0.175)	0.354 (0.619)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	4835	1515	3320	2904	1450	398
R-squared	0.192	0.214	0.197	0.243	0.300	0.492
P-value for testing coefficients equal to zero						
Education	0.3291	0.8787	0.4094	0.2254	0.9136	0.0305
Assets	0.0617	0.4407	0.0726	0.1549	0.2959	0.6390
Age dummies	0.0262	0.0485	0.2203	0.4784	0.4726	0.0752
Cohort dummies	0.0002	0.0023	0.0438	0.2164	0.8841	0.6150
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-values for testing between urban and rural areas						
Education	Men	0.9899	Women	0.9320		
Assets		0.9275		0.6931		
Age dummies		0.5492		0.8247		
Cohort dummies		0.2157		0.7029		

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.19: Changes in Food Consumption in CHNS 89,91,93 with Community Characteristics

Dependent Variables	Male				
	Δ Calorie (kcal)	Δ Fat (g)	Δ Protein (g)	$\Delta\%$ Calorie From Fat Protein Carbo	
Initial Some primary edu	-23.314 (73.566)	-2.823 (4.051)	0.380 (3.148)	-0.996 (0.818)	0.118 (0.256)
Initial Primary	5.235 (71.461)	-2.380 (3.727)	0.188 (3.108)	-0.503 (0.784)	0.014 (0.245)
Initial Middle school	34.647 (71.528)	-0.527 (3.704)	4.117 (3.106)	-0.192 (0.773)	0.355 (0.240)
Initial High school	33.886 (80.303)	-1.097 (4.275)	4.260 (3.661)	-0.190 (0.901)	0.368 (0.281)
Initial Tech/College+	18.601 (87.366)	-3.300 (4.607)	-1.219 (4.027)	-0.949 (1.009)	-0.171 (0.362)
Initial Log real prod assets 1	8.019 (13.301)	0.120 (0.685)	0.173 (0.616)	-0.063 (0.139)	-0.026 (0.048)
Initial Log real prod assets 2	-11.611 (12.865)	-0.230 (0.650)	-0.689 (0.581)	0.051 (0.140)	-0.050 (0.044)
Initial Land	-2.128 (6.595)	0.143 (0.288)	-0.076 (0.322)	0.044 (0.068)	-0.016 (0.024)
Initial Urban Residence	510.414 (504.605)	7.552 (23.014)	37.335 (35.455)	4.157 (5.962)	2.295 (2.739)
Δ Price of Rice	-21.699 (52.548)	-1.189 (2.691)	-3.145 (2.928)	-0.171 (0.605)	-0.399 (0.245)
Δ Price of Pork	-6.230 (19.060)	-0.442 (1.010)	0.353 (0.958)	-0.223 (0.231)	0.036 (0.081)
Δ Price of Eggs	-35.147 (8.393)**	-1.259 (0.426)**	-1.338 (0.402)**	-0.217 (0.098)**	-0.028 (0.034)
Δ Price of Oil	-38.266 (16.429)**	0.579 (0.880)	-1.618 (0.837)*	0.245 (0.214)	-0.021 (0.083)
Δ Water Source 2	131.718 (141.332)	8.658 (7.205)	2.148 (6.441)	3.133 (1.611)*	-0.160 (0.495)
Δ Water Source 3	226.994 (189.954)	-1.655 (10.178)	4.515 (7.700)	-3.100 (2.020)	-0.241 (0.587)
					(2.610)*

Table continues

Table 2.19 (cont'd)

Dependent Variables	Female					
	ΔCalorie (kcal)	ΔFat (g)	ΔProtein (g)	Fat	Δ% Calorie From Protein	Carbo
Initial Some primary edu	16.348 (38.552)	-0.017 (1.989)	1.059 (1.964)	0.136 (0.536)	0.095 (0.184)	-0.359 (0.602)
Initial Primary	-76.347 (42.833)*	-1.123 (2.282)	-2.089 (2.148)	-0.047 (0.604)	0.064 (0.194)	-0.119 (0.652)
Initial Middle school	-3.603 (41.455)	-0.338 (2.247)	-0.993 (2.033)	-0.106 (0.593)	-0.115 (0.196)	0.153 (0.649)
Initial High school	-26.942 (53.245)	-3.623 (3.019)	-1.421 (2.735)	-1.026 (0.794)	-0.072 (0.265)	1.075 (0.850)
Initial Tech/College+	23.723 (68.816)	6.964 (3.966)*	-0.752 (3.559)	1.955 (1.058)*	-0.215 (0.397)	-1.889 (1.082)*
Initial Log real prod assets 1	3.237 (9.156)	-0.074 (0.470)	0.159 (0.457)	-0.076 (0.131)	-0.006 (0.043)	0.081 (0.144)
Initial Log real prod assets 2	-18.239 (9.412)*	-0.636 (0.484)	-0.994 (0.464)**	-0.095 (0.135)	-0.039 (0.042)	0.153 (0.148)
Initial Land	-2.410 (5.744)	0.606 (0.274)**	0.248 (0.291)	0.215 (0.076)**	0.033 (0.026)	-0.224 (0.085)**
Initial Urban Residence	-238.676 (349.149)	21.420 (14.385)	-34.335 (19.624)*	10.558 (4.932)**	-4.123 (1.194)**	-6.451 (5.262)
ΔPrice of Rice	4.823 (45.751)	1.076 (2.301)	-3.096 (2.398)	0.233 (0.581)	-0.597 (0.200)**	0.318 (0.631)
ΔPrice of Pork	-11.947 (15.503)	-0.289 (0.906)	-0.164 (0.782)	-0.124 (0.249)	0.013 (0.073)	0.111 (0.268)
ΔPrice of Eggs	-18.927 (6.762)**	-1.061 (0.389)**	-0.719 (0.342)**	-0.239 (0.102)**	-0.001 (0.031)	0.266 (0.112)**
ΔPrice of Oil	-28.309 (13.184)**	-0.565 (0.854)	-0.891 (0.650)	-0.020 (0.225)	0.047 (0.066)	-0.003 (0.231)
ΔWater Source 2	196.545 (106.869)*	16.604 (6.185)**	8.367 (5.632)	3.978 (1.592)**	0.195 (0.521)	-3.899 (1.811)**
ΔWater Source 3	127.316 (146.962)	5.290 (8.063)	10.152 (6.610)	-0.467 (2.230)	0.483 (0.587)	0.209 (2.271)

Table continues

Table 2.19 (cont'd)

Dependent Variables	Male					
	Δ Calorie (kcal)	Δ Fat (g)	Δ Protein (g)	Fat	$\Delta\%$ Calorie From Protein	Carbo
Δ Water Source 4	110.746 (124.037)	7.032 (7.581)	1.374 (6.371)	1.370 (1.619)	-0.458 (0.491)	0.191 (1.834)
Δ In house no flush	682.461 (385.167)*	11.099 (20.727)	56.619 (20.551)**	3.049 (4.769)	2.957 (1.754)*	-0.930 (5.778)
Δ Outside toilets	8.745 (318.205)	-9.406 (16.471)	4.308 (17.220)	-0.226 (3.679)	-0.548 (1.531)	5.354 (4.537)
Δ Open pit	-264.783 (333.334)	-3.517 (16.832)	-6.494 (18.076)	2.589 (3.697)	-0.693 (1.562)	2.275 (4.599)
Δ No toilets	-115.416 (441.689)	-13.365 (23.169)	9.741 (21.796)	0.863 (5.328)	0.838 (2.037)	3.512 (6.590)
Δ Very little excreta	323.483 (106.577)**	12.738 (5.377)**	20.394 (5.276)**	2.423 (1.187)**	1.345 (0.422)**	-6.073 (1.543)**
Δ Some excreta	-125.639 (109.341)	-6.734 (5.286)	-0.538 (5.305)	-1.662 (1.253)	0.535 (0.429)	1.814 (1.563)
Δ Much excreta	410.592 (495.224)	0.111 (21.250)	6.141 (19.796)	-3.292 (4.641)	-0.824 (1.379)	-1.360 (6.174)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	4083	4083	4083	4083	4083	4083
R-squared	0.129	0.091	0.120	0.105	0.124	0.114
P-value for testing coefficients equal to zero						
Education	0.9006	0.8602	0.1634	0.6116	0.1644	0.1017
Prod Assets	0.6525	0.9214	0.2763	0.9023	0.0468	0.3049
Age dummies	0.4681	0.2552	0.1847	0.2125	0.0014	0.1293
Cohort dummies	0.3207	0.4720	0.0038	0.6281	0.0000	0.3528
Δ Prices	0.0000	0.0219	0.0002	0.0769	0.2907	0.0029
Δ Water source	0.5784	0.5719	0.9423	0.0380	0.8314	0.1611
Δ Toilet type	0.0119	0.7722	0.0010	0.5432	0.0518	0.3906
Δ Excreta	0.0019	0.0166	0.0006	0.0435	0.0108	0.0001
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

Table 2.19 (cont'd)

Dependent Variables	Female					
	ΔCalorie (kcal)	ΔFat (g)	ΔProtein (g)	Fat	Protein	Carbo
ΔWater Source 4	178.701 (103.465)*	12.405 (6.294)**	6.471 (5.564)	2.223 (1.509)	-0.146 (0.509)	-2.128 (1.641)
ΔIn house no flush	887.829 (335.243)**	51.047 (22.868)**	44.905 (17.122)**	7.340 (5.129)	1.241 (1.628)	-6.118 (5.558)
ΔOutside toilets	425.175 (259.548)	30.023 (18.918)	17.592 (12.888)	4.873 (4.030)	0.254 (1.300)	-4.357 (4.217)
ΔOpen pit	196.423 (272.283)	33.232 (19.774)*	5.162 (13.701)	7.526 (4.203)*	-0.281 (1.343)	-6.749 (4.432)
ΔNo toilets	-394.535 (361.104)	-17.286 (25.093)	-12.363 (17.428)	-4.247 (6.214)	0.746 (1.721)	2.327 (6.447)
ΔVery little excreta	-24.147 (89.623)	0.770 (4.626)	10.471 (4.290)**	0.449 (1.225)	1.586 (0.380)**	-2.272 (1.308)*
ΔSome excreta	29.837 (93.572)	-2.973 (4.497)	7.615 (4.416)*	-1.765 (1.161)	0.889 (0.379)**	0.613 (1.294)
ΔMuch excreta	-142.550 (423.578)	12.471 (21.254)	11.386 (17.455)	0.979 (4.974)	2.305 (1.260)*	-3.465 (5.439)
Community Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	4561	4561	4561	4561	4561	4561
R-squared	0.121	0.091	0.097	0.098	0.096	0.102
P-value for testing coefficients equal to zero						
Education	0.2814	0.1724	0.7944	0.1122	0.8739	0.0850
Prod Assets	0.0427	0.1151	0.0210	0.1966	0.2977	0.0790
Age dummies	0.7028	0.4245	0.1154	0.4271	0.0820	0.0320
Cohort dummies	0.0593	0.2139	0.0009	0.2982	0.0212	0.0000
ΔPrices	0.0030	0.0612	0.0113	0.1638	0.0436	0.0808
ΔWater source	0.1938	0.0393	0.3135	0.0568	0.7457	0.1315
ΔToilet type	0.0000	0.0355	0.0003	0.1389	0.8234	0.3697
ΔExcreta	0.9470	0.8050	0.0666	0.3816	0.0001	0.2175
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Also included in the models are individual-year age dummies, five-year-cohort dummies and community dummies. Person-level robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table 2.20: A Simple Production Function Analysis on Log(weight) in 1993 and Log(BMI) in 1993

	Male Log weight 1993		Male Log BMI 1993		Female Log weight 1993		Female Log BMI 1993	
	IV	OLS	IV	OLS	IV	OLS	IV	OLS
Log weight 1991	1.071 (0.134)**	0.755 (0.030)**			0.848 (0.091)**	0.715 (0.026)**		
Log height 1991	-0.427 (0.508)	0.248 (0.102)**			0.717 (0.422)*	0.405 (0.077)**		
Log BMI 1991			1.002 (0.130)**	0.707 (0.032)**	-0.004 (0.010)	0.001 (0.001)	0.901 (0.091)**	0.673 (0.027)**
Moderate to heavy PA	0.005 (0.025)	-0.012 (0.006)**	0.009 (0.023)	-0.011 (0.006)**	0.028 (0.028)	-0.017 (0.005)**	0.008 (0.026)	-0.016 (0.005)**
Calories (100Kcal)	0.012 (0.009)	-0.002 (0.001)*	0.002 (0.009)	-0.002 (0.001)**	-0.002 (0.011)	0.001 (0.001)	-0.010 (0.011)	0.001 (0.001)
Fat (gram)	-0.003 (0.002)*	-0.000 (0.000)	-0.001 (0.002)	0.000 (0.000)	-0.003 (0.002)	-0.0001 (0.000)	-0.001 (0.002)	0.0001 (0.000)
Protein (gram)	-0.002 (0.003)	0.0005 (0.000)**	-0.001 (0.003)	0.001 (0)**	0.001 (0.003)	-0.000 (0.000)	0.002 (0.003)	0.000 (0.000)
% Calories from Fat	0.009 (0.006)	0.001 (0.001)	0.004 (0.005)	0.001 (0.001)	0.008 (0.006)	-0.0004 (0.001)	-0.001 (0.006)	-0.0004 (0.001)
% Calories from Protein	0.015 (0.018)	-0.002 (0.002)	0.006 (0.018)	-0.003 (0.002)	-0.004 (0.020)	0.001 (0.002)	-0.004 (0.021)	0.001 (0.002)
Constant	1.620 (2.283)	-0.184 (0.493)	-0.018 (0.461)	0.981 (0.102)**	-2.967 (1.888)	-0.920 (0.323)**	0.478 (0.404)	0.999 (0.086)**
Number of obs	2556	2556	2556	2556	2897	2897	2897	2897
R-squared	0.474	0.593	0.300	0.431	0.512	0.642	0.319	0.500
P-values for Tests of Coefficients								
Log weight or BMI 1991=1	0.5967	0.0000	0.9871	0.0000	0.0934	0.0000	0.2782	0.0000
Intakes	0.2729	0.0248	0.5433	0.0019	0.3989	0.0974	0.1718	0.1212
Age	0.3537	0.0885	0.7461	0.7754	0.2146	0.0014	0.1690	0.0000
Wu-Hausman test of endogenous variables								
F-statistics	4.50		3.08		3.79		5.81	
p-value	0.0000		0.0031		0.0002		0.0000	
Overidentification Test								
Chi-sq(20)	15.15		10.61		26.77		26.61	
p-value	0.7130		0.9364		0.1102		0.1141	

Note: Also included in all regressions are individual year age dummies. Lagged weight, height and BMI, current physical activity levels, and current nutrient intakes are endogenous, and instrumented with education, real productive assets and land cultivated, and community characteristics including prices, water and sanitation conditions in 1991. For the goodness-of-fit of the first stage regressions, see Appendix F. * indicates the coefficient is statistically significant at 0.1 level; ** at 0.05 level.

Chapter 3: Age, Cohort and Year

Analysis in the Socioeconomic

Determinants of Adult BMI in the 1990s

3.1 Introduction

The age-period-cohort (APC) analysis in the medical research literature aims at determining whether the phenomenon under examination is an origin-related process (cohort effect dominant), an aging process or a function of time (instantaneous process). In the study of body mass index (BMI) there are reasons to believe the three effects may all be important. In different phases in life BMI reflects different aspects of body growth. Before reaching maturity BMI is most responsive to increases in height and for the elderly shrinking becomes the main factor for reduced BMI. Do men and women have distinctive life-cycle BMI profiles? For the people born during the severe famine in the late 1950s and early 1960s in China, did the impact carry on to their adulthood or have they made up the difference over time? Year (or time, period) effects may arise from improved living standards or/and medical technologies that reduce chronic diseases or even the changing ideas of what is fashionable. However, because of the linear dependency of age, period and cohort, the level effects of the three factors can not be identified without further normalization or exclusion assumptions based on prior information. Convenient normalizations may lead to erroneous interpretation of the data.

In the upper panel of Figure (3.1) the Lowess smoothed BMI of each five-year

cohort shows that at age less than 55 men born of a five-year later cohort have higher BMIs and the differences are larger for the 35 to 50 age group than for the younger cohorts. However, the cohort differences for women were not as pronounced. For women the inverse U shape of BMI versus age is more evident than for men. The lower panel of Figure (3.1) depicts each three-year cohort BMI against age. As we can see the cohort differences for men become less significant and so does it for women. In the upper panel for the five-year-cohort graphs two cohorts are of particular interest. The July 1957 - June 1962 ("x") and July 1967 - June 1972 cohort ("o"), the former spans the famine years and the latter represents the first half of the cultural revolution. Correspondingly in the lower panel the two three-year cohorts are the July 1957 - June 1960 ("x") and July 1969 - June 1971 cohort ("o"). It seems that the impact of famine is more severe on women although later in life there seems to be over-compensation after age 35.

3.2 Identification of Second Differences

Deaton (1997) suggests that when data are plentiful it is reasonable to use dummy variables for all three sets of effects to allow the data to choose any pattern (page 124). For individual i in cohort c at time t , denote his or her age as a . Assume the following additive model for variable y_i :

$$y_{i,c,a,t} = u + \alpha_c + \beta_a + \gamma_t + \epsilon_{i,c,a,t} \quad (3.1)$$

Suppose a ranges from age $a_1 \sim a_A$, $t \equiv t_1 \sim t_T$ and $c \equiv t - a$ takes values from $c_1 \sim c_P$. To avoid the trivial linear dependencies associated with using dummy variables, the usual normalizing constraints are either $\{\alpha_1 = 0, \beta_1 = 0, \gamma_1 = 0\}$ or $\left\{ \sum_{c=1}^P \alpha_c = 0, \sum_{a=1}^A \beta_a = 0, \sum_{t=1}^T \gamma_t = 0 \right\}$ (Heckman and Robb 1985, Holford 1983).

It has been pointed out in the specification above yet another constraint is neces-

sary due to the linear relationship between c , a and t (Holford 1983, and the citations within). However, as shown in the appendix, provided that age and cohort categories are consecutive, the number of constraints needed to identify all three effects depend on the spacing of the periods in which the data were taken. For annual vital or prevalence data, the number of constraints necessary is one; for biannual survey data, the number of constraints needed is two; for data taken every four years, the number of constraints needed is four; and for our data taken in 1989, 91, 93 and 97 the number of constraints necessary is two. If we leave it to the computer to arbitrarily impose two constraints the results are meaningless (Figure 3.2).

McKenzie (2002) provides an approach to identify the second derivatives (or second differences) of these effects without any normalizing conditions. The changes in slopes effectively provide information on changes in growth rates and the convexity or concavity of the curves. The McKenzie approach applies to both repeated cross-sections and genuine panel data. Below we illustrate how the method works in the latter case since our data are of panel structure and McKenzie (2002) laid out the former case. In the former case replace the individual information by corresponding group means and the method follows. We can use both the individual and group observations to see how sensitive the two approaches are to the data. Assume data are collected annually.

Within each individual let $\Delta_t y_i \equiv y_{i,c,a,t} - y_{i,c,a-1,t-1}$. First time-differencing y_i eliminates the cohort effect:

$$\begin{aligned} \Delta_t y_i &= (\alpha_c + \beta_a + \gamma_t + \epsilon_{i,c,a,t}) - (\alpha_c + \beta_{a-1} + \gamma_{t-1} + \epsilon_{i,c,a-1,t-1}) \quad (3.2) \\ &= \beta_a - \beta_{a-1} + \gamma_t - \gamma_{t-1} + \epsilon_{i,c,a,t} - \epsilon_{i,c,a-1,t-1} \\ &\equiv \beta_a - \beta_{a-1} + \gamma_t - \gamma_{t-1} + \Delta_t \epsilon_i \end{aligned}$$

For another individual from a one-year earlier cohort

$$y_{j,c-1,a+1,t} = u + \alpha_{c-1} + \beta_{a+1} + \gamma_t + \epsilon_{j,c-1,a+1,t} \quad (3.3)$$

the first difference of y_j is

$$\begin{aligned} \Delta_t y_j &\equiv y_{j,c-1,a+1,t} - y_{j,c-1,a,t-1} \\ &= \beta_{a+1} - \beta_a + \gamma_t - \gamma_{t-1} + \Delta_t \epsilon_j \end{aligned} \quad (3.4)$$

Subtracting (3.2) from (3.4) eliminates the time effects and gives

$$\begin{aligned} \Delta_c \Delta_t y_{i,j} &= (\beta_{a+1} - \beta_a + \gamma_t - \gamma_{t-1} + \Delta_t \epsilon_j) - (\beta_a - \beta_{a-1} + \gamma_t - \gamma_{t-1} + \Delta_t \epsilon_i) \\ &= (\beta_{a+1} - \beta_a) - (\beta_a - \beta_{a-1}) + \Delta_t \epsilon_j - \Delta_t \epsilon_i \\ &\equiv \delta_a^A + \Delta_c \Delta_t \epsilon_{i,j} \end{aligned} \quad (3.5)$$

where δ_a^A denotes the second difference of age effects, or the difference in the slope between age $a + 1$ and a from the slope between age a and $a - 1$. Structure the data as in (3.5) and we can estimate δ_a^A . Hence the changes in the slopes of age effects can be identified without any normalizing assumption.

For the individual i in (3.1) at time period $t + 1$ we have

$$y_{i,c,a+1,t+1} = u + \alpha_c + \beta_{a+1} + \gamma_{t+1} + \epsilon_{i,c,a+1,t+1} \quad (3.6)$$

Subtracting (3.1) from (3.6) gives

$$\begin{aligned} \Delta_{t+1} y_i &= (\alpha_c + \beta_{a+1} + \gamma_{t+1} + \epsilon_{i,c,a+1,t+1}) - (\alpha_c + \beta_a + \gamma_t + \epsilon_{i,c,a,t}) \\ &= \beta_{a+1} - \beta_a + \gamma_{t+1} - \gamma_t + \Delta_{t+1} \epsilon_i \end{aligned} \quad (3.7)$$

Denote the difference between (3.7) and (3.4) as $\Delta_a \Delta_t y_{i,j}$.

$$\begin{aligned}
\Delta_a \Delta_t y_{i,j} &= (\beta_{a+1} - \beta_a + \gamma_{t+1} - \gamma_t + \Delta_{t+1} \epsilon_i) - (\beta_{a+1} - \beta_a + \gamma_t - \gamma_{t-1} + \Delta_t \epsilon_j) \\
&= (\gamma_{t+1} - \gamma_t) - (\gamma_t - \gamma_{t-1}) + \Delta_{t+1} \epsilon_i - \Delta_t \epsilon_j \\
&\equiv \delta_t^T + \Delta_a \Delta_t \epsilon_{i,j}
\end{aligned} \tag{3.8}$$

where δ_t^T denotes the second difference of time effects around t . Structuring the data as in (3.8) allows us to estimate δ_t^T .

Subtracting (3.3) from (3.1) gives

$$\begin{aligned}
\Delta_c y_{i,j} &= (\alpha_c + \beta_a + \gamma_t + \epsilon_{i,c,a,t}) - (\alpha_{c-1} + \beta_{a+1} + \gamma_t + \epsilon_{j,c-1,a+1,t}) \\
&= \alpha_c - \alpha_{c-1} + \beta_a - \beta_{a+1} + \Delta_c \epsilon_{i,j}
\end{aligned} \tag{3.9}$$

Do the same for another pair of observations at time $t+1$, $y_{m,c,a+1,t+1} = u + \alpha_c + \beta_{a+1} + \gamma_{t+1} + \epsilon_{m,c,a+1,t+1}$ and $y_{n,c,a+1,t+1} = c + \alpha_c + \beta_{a+1} + \gamma_{t+1} + \epsilon_{n,c,a+1,t+1}$ and we get

$$\Delta_c y_{m,n} = \alpha_{c+1} - \alpha_c + \beta_a - \beta_{a+1} + \Delta_c \epsilon_{m,n} \tag{3.10}$$

Subtracting (3.9) from (3.10) eliminates the age effects and gives

$$\begin{aligned}
\Delta_c \Delta_c y_{i,j,m,n} &= (\alpha_{c+1} - \alpha_c + \beta_a - \beta_{a+1} + \Delta_c \epsilon_{m,n}) - (\alpha_c - \alpha_{c-1} + \beta_a - \beta_{a+1} + \Delta_c \epsilon_{i,j}) \\
&= (\alpha_{c+1} - \alpha_c) - (\alpha_c - \alpha_{c-1}) + \Delta_c \epsilon_{m,n} - \Delta_c \epsilon_{i,j} \\
&\equiv \delta_c^C + \Delta_c \Delta_c \epsilon_{i,j,m,n}
\end{aligned} \tag{3.11}$$

where δ_c^C denotes the second difference of cohort effects around c , and c takes values from 1915 to 1976. Constructing the data as in (3.11) we can estimate δ_c^C .

When only repeated cross section data are available, the pseudo-panel version of

equation (3.1) is

$$\bar{y}_{c,a,t} = u + \alpha_c + \beta_a + \gamma_t + \bar{\epsilon}_{c,a,t}$$

and all the above manipulations follow through and δ_a^A , δ_t^T and δ_c^C can be estimated. Since we have true longitudinal data, we can apply both the genuine panel approach based on individual data and the pseudo-panel approach based on group means. The Mckenzie method works the best when the data are evenly spaced in time. Hence we perform the analysis using two-year-apart data in 1989, 91 and 93 and four-year-apart data in 1989, 93 and 97, resulting in the second difference estimators across two-year and four-year intervals.

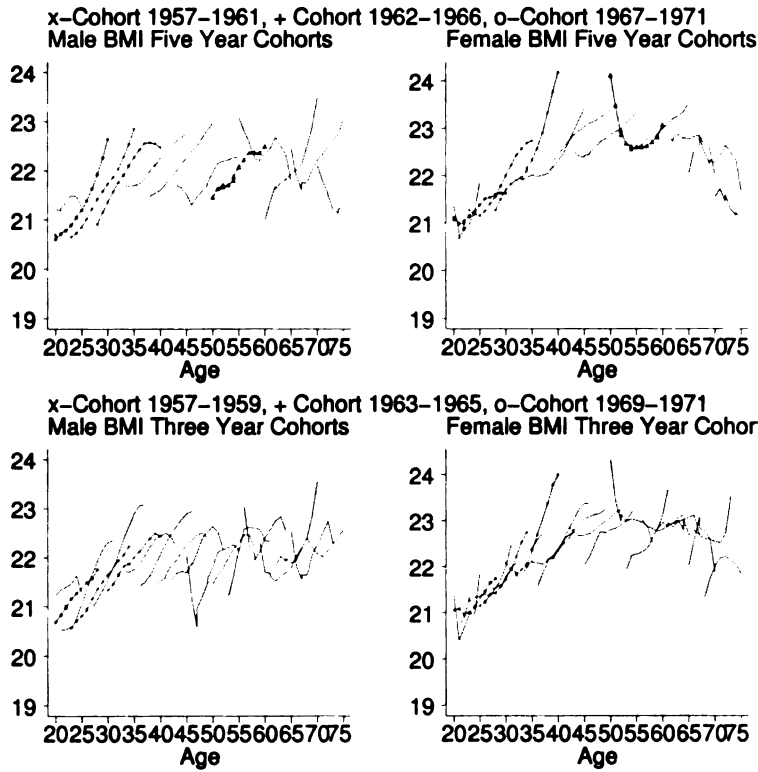


Figure 3.1: Lowess Smoothed Male and Female BMI of Each Five- and Three-Year Cohort vs. Age

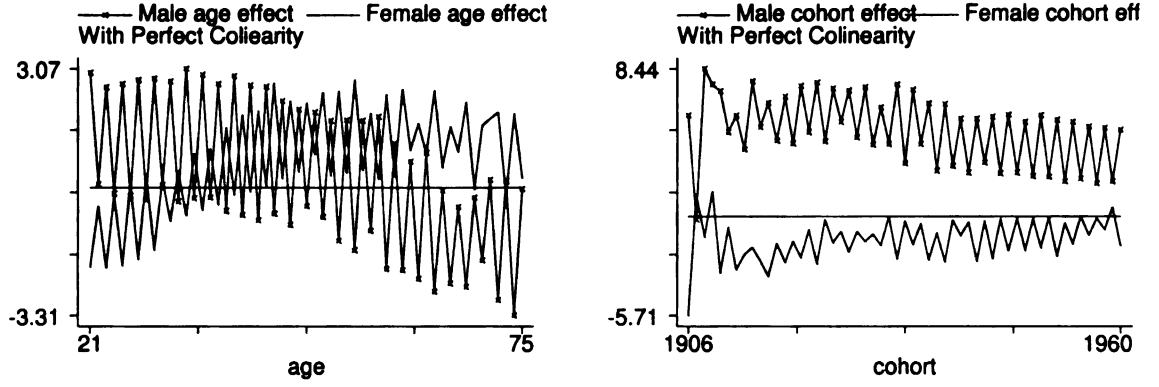


Figure 3.2: Estimated Age and Cohort Effects for Men and Women with Arbitrarily Imposed Constraints

3.3 Simulation Study

To answer the question "can the second derivatives help selecting correct estimation strategy identifying the age, year and cohort effects", we do the following simulation exercise.

Create a data set with 2500 (or 10,000 for the long panel) observations, with age ranging 20 to 59 in year 1980 and same people being surveyed for the next four (or nineteen for the long panel) years till 1984 (or 1999 for the long panel). The first example is a short five-year panel data and the second example is a relatively longer 20-year panel. The cohort is from 1921 to 1960. Generate an outcome variable with linear but no higher order age, period and cohort effects such that $x = 0.5a + 0.3c + 0.7t + \epsilon$, where $a = age - 20$, $c = cohort - 1921$, $t = year - 1980$, and $\epsilon \sim N(0, 1)$. Maintaining age and year effects but changing the cohort effect we get another outcome variable $x_1 = 0.5a - 0.3c + 0.7t + \epsilon$ in order to compare the impact of this change on estimations of age and year effects. Generate an outcome variable with quadratics of age, period and cohort effects but no interactions amongst them such that $y = 0.5a - 0.04a^2 + 0.3c - 0.02c^2 + 0.7t - 0.01t^2 + \epsilon$, where $\epsilon \sim N(0, 1)$. Generate

another variable with different cohort effects but the same age and period effects such that $y_1 = 0.5a - 0.04a^2 + 0.5c - 0.04c^2 + 0.7t - 0.01t^2 + \epsilon$, where $\epsilon \sim N(0, 1)$. Set $\partial y / \partial a = 0$, the inflection point for age profile of y is at age 26.25; so is it for y_1 . Set $\partial y / \partial c = 0$, the inflection point for cohort profile of y is at cohort 1928.5; and 1927 for y_1 .

We can estimate models with the following constraints: (1) linear cohort and year effects, (2) linear age and year effects, (3) linear age and cohort effects, (4) linear cohort and year effects with quadratic age, cohort and year effects, (5) linear age and year effects with quadratic age, cohort and year effects, (6) linear age and cohort effects with quadratic age, cohort and year effects, (7) linear cohort and year effects with quadratic cohort and year effects, (8) linear age and year effects with quadratic age and year effects, and (9) linear age and cohort effects with quadratic age and cohort effects. None of these models are correct specifications according to the data generating process. Tables 3.1 to 3.4 summarize the regression results for the above specifications for x and x_1 ; and Tables 3.5 to 3.8 for y and y_1 .

3.3.1 Models with linear and/or squared terms

Estimates for x and x_1

When the data generating process is linear in age, period and cohort, clearly the linear dependency $a \equiv t - c$ requires one of the linear effects be excluded from the model. However, when, for example in Table 3.1 for $x = 0.5a + 0.3c + 0.7t + \epsilon$, age is left out, the estimated coefficient for cohort should reflect the combined effect of age and cohort, i.e. $x = 0.5(t - c) + 0.3c + 0.7t + \epsilon = -0.2c + 1.2t + \epsilon$. In column 1 of Table 3.1 we can see that is in fact the case. Similarly, when cohort effect is excluded from the model the coefficients for age and year effects are also combined effects. It is impossible to disentangle individual effects from such estimators. Table 3.2 of the short panel for x_1 displays the same feature. In Table 3.3 and Table 3.4 when the

panel gets longer the point estimates for the combined effects are more accurate.

Using the McKenzie approach we can estimate the second derivatives of age, regressing the corresponding second differences of the outcome variables on age as a continuous variable, as seen at the bottom of Tables 3.1 to 3.4. None of the second derivatives of age, period or cohort are significant, suggesting the model should not include higher order terms. This is confirmed by columns 4 to 9 in Tables 3.1 to 3.4. When the second derivatives estimated without normalization assumption indicate there is no higher order effects the square terms should not be included in the modelling.

The difference between x and x_1 lies only in the cohort effect (0.3 for x and -0.3 for x_1). As we can see in column 3, 6 and 9 models excluding year effect estimate the combined age and year effects consistently; and in column 2, 5, and 8 models excluding cohort effect give rise to the combined age and cohort effects that are affected by the coefficients of c in the data generating processes.

Estimates for y and y_1

When the data generating process is with linear and second order effects of age, period and cohort factors and the model is specified as in columns 4 to 6 from Tables 3.5 to 3.8, we can see the second order effects are correctly estimated. However, the linear terms are still combined effects. For example in column 6 then the linear year effect is excluded the estimated model in effect is $y = 0.5a - 0.04a^2 + 0.3c - 0.02c^2 + 0.7t - 0.01t^2 + \epsilon = 0.5a - 0.04a^2 + 0.3c - 0.02c^2 + 0.7(a + c) - 0.01t^2 + \epsilon = 1.2a + c - 0.04a^2 - 0.02c^2 - 0.01t^2 + \epsilon$. As we can see all combined linear effects (age and year, cohort and year) are fairly accurately estimated.

When linear age effect is set to zero, there is no reason to believe the second order age effect should be different than zero. Hence in column 7 both linear and squared terms of age effects are set to zero. In such case, the estimated model in

effect is $y = 0.5a - 0.04a^2 + 0.3c - 0.02c^2 + 0.7t - 0.01t^2 + \epsilon = 0.5(t - c) - 0.04(t - c)^2 + 0.3c - 0.02c^2 + 0.7t - 0.01t^2 + \epsilon = -0.2c + 1.2t - 0.06c^2 - 0.05t^2 + 0.04tc$. As we can see in the long panel case (Table 3.7 column7) the coefficients for c^2 and t^2 are correctly estimated; the estimator for coefficient of c ($= 3.677$) is actually the combined effect of c and tc evaluated at $t = 97$; and the estimator for coefficient of t ($= -0.361$) is the combined effect of t and ct evaluated at $c = 39$.

Using the McKenzie approach we estimate the second derivatives of age, cohort and year effects. The t-statistic strongly suggests the second order age effect should not be zero ($p=.000$), the second order cohort effect is also nonzero ($p=.038$), and the second order year effect is not different from zero ($p=.338$). At least the McKenzie approach instructs us to include both squared terms of age and cohort.

The difference between y and y_1 is in the cohort parameters. The fact that the estimated age coefficients between Tables 3.7 and 3.8 in columns 3, 6 and 9 stay close but the estimated age coefficients between Tables 3.7 and 3.8 in column 2, 5 and 8 change dramatically indicates the two data generating processes vary mainly in the cohort factor.

3.3.2 Models with dummy variables

Clearly models with only linear terms or even linear and squared terms are not estimating parameters we are interested in. When the data generating process is defined as above for y and y_1 , it is quite impossible to separately estimate the coefficient for a , c or t . However, in a model with no interactions where $y = \beta_1 a + \beta_2 a^2 + \beta_3 c + \beta_4 c^2 + \beta_5 t + \beta_6 t^2 + \epsilon$, if we set $\partial y / \partial a \equiv 0$, the inflection point for age profile of y is at age a such that $\beta_1 + 2\beta_2 a = 0$. We can estimate regressions with dummy variables that include: (1) age and cohort dummies, (2) age and year dummies, (3) cohort and year dummies, (4) age, five-year-cohort and year dummies, or (5) age, three-year-cohort, and year dummies. All these models get around the exact

linear dependency between age, period and cohort through different identification assumptions. In model (1) year effect is assumed to be zero, the estimated age and cohort profiles are combined profiles; and similar assumptions are made for cohort and age in model (2) and (3). In model (4) it is assumed every five-year-cohort has the same cohort effect and in model (4) it is so for every three-year cohort. After plotting out the estimated age profile, we can find the inflection point and the relationship between β_1 and β_2 . For example, if we can approximately eyeball the inflection point of age at 6 years after the beginning of all estimated age periods, we can say $\beta_1 + 12\beta_2 = 0$. At the mean time we can use the McKenzie approach to estimate β_2 . Hence we will have an estimate of β_1 .

Profiles for x and x_1

Throughout Figure 3.3 to 3.6 the small circled lines are the true profiles from the known data generating processes.

Figure 3.3 displays on the left the age, cohort, and year effects for x each identified through four of the above five different identifications and on the right for x_1 both for the short panel. Figure 3.4 corresponds to the longer panel for x and x_1 . For the age profiles in Figure 3.3 and 3.4, we can see when the model (short dashed line) includes age and cohort dummies the age profile (with slope about 1.2) is the combined age and year effects and is not affected by the coefficients on c in the data generating processes (0.3 for x and -.03 for x_1). So is the case for year effects (lower panel dotted line) identified through models excluding age dummies because these year profiles are combined year and age effects that are not affected by the coefficients on c in the data generating processes.

When the age profile is identified by models including cohort dummies, be they one-year, three-year or five-year cohort dummies, the age profile is underestimated on the left for x and overestimated on the right for x_1 . As we can see the model with

age, three-year-cohort and year dummies is closest to the true profiles of age, cohort and year.

3.3.3 Profiles for y and y_1

Figure 3.5 and 3.6 show the results from the dummy-variable models for y and y_1 . Again we can see all models are estimating combined profiles instead of the individual profiles that we are interested in. The five-year and three-year cohort specifications give rise to very similar profiles and are generally closer to the true profiles from the known data generating processes.

One basic assumption in age, period, cohort analysis without interaction terms is that the age effect at any given age across different birth cohorts are the same. Notice in our model, particularly in the short panel, the age effects for younger ages are identified through later birth cohorts and for older ages through earlier birth cohorts. If the assumption does not hold the estimated age profile is biased.

As seen in Tables 3.5 to 3.9 the estimated second derivative of age using the McKenzie approach without any normalization assumption is $-.002$; and from the age profile identified through the three-year-cohort specification in Figure 3.6 the inflection point is at 12 years. Therefore we can estimate from setting $\partial y / \partial a \equiv \beta_1 + 2\beta_2 a = 0$, where $\hat{\beta}_2 = -0.02$ and $a = 12$, hence $\hat{\beta}_1 = 0.48$ which is very close to the true linear age effect in the known data generating process of y .

In conclusion, when estimating age, period and cohort effects in general we should (1) avoid using simple linear specifications, (2) when interested in the age profile use three-year or five-year cohort specifications, and (3) interpret the combined effect correctly.

Table 3.1: Short Panel: Estimated restricted age, cohort and year effects for x
 $x = 0.5a + 0.3c + 0.7t + \epsilon$ Age ranges 20-63, year 1980-1984, and cohort 1921-1960

	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
Real	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
Coefficients									
age 0.5									
	0.198	1.182							
	(0.002)**	(0.014)**							
cohort 0.3									
	-0.198	0.984							
	(0.002)**	(0.014)**							
year 0.7									
	1.182	0.984							
	(0.014)**	(0.014)**							
ageseq 0									
cohortsq 0									
yearsqr 0									
constant									
	-1940.70	-1940.70	-1940.70						
	(28.52)**	(28.52)**	(28.52)**						
Number of obs	2500	2500	2500	2500	2500	2500	2500	2500	2500
F-stat	9840.85	9840.85	9840.85	3943.63	3943.63	3943.63	4929.72	4931.40	4931.10
Root MSE	1.0102	1.0102	1.0102	1.0094	1.0094	1.0094	1.0094	1.0092	1.0093
Estimated second derivatives of age - 0.0002 (0.001), p=.983									
Estimated second derivatives of cohort 4.60e-06 (.00002), p=.845									
Estimated second derivatives of year -3.36e-06 (.000026), p=.898									
Mean(x)=17.98									

Table 3.2: Short Panel: Estimated restricted age, cohort and year effects for x_1
 $x_1 = 0.5a - 0.3c + 0.7t + \epsilon$ Age ranges 20-63, year 1980-1984, and cohort 1921-1960

	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
Real Coefficients	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
age 0.5		0.797 (0.002)**	1.200 (0.014)**		0.790 (0.051)**	1.155 (0.056)**		0.797 (0.007)**	1.195 (0.030)**
cohort -0.3	-0.797 (0.002)**		0.403 (0.014)**	-0.790 (0.051)**		0.366 (0.057)**	-0.798 (0.007)**		0.407 (0.029)**
year 0.7	1.200 (0.014)**	0.403 (0.014)**		1.155 (0.056)**	0.366 (0.057)**		1.159 (0.050)**	0.362 (0.050)**	
agesq -0.04				.0001 (0.001)	.0001 (0.001)	.0001 (0.001)	.00001 (.0002)		.0001 (0.001)
cohortsq -0.02				-.0001 (0.001)	-.0001 (0.001)	-.0001 (0.001)	6.03e-06 (.0002)		-.0001 (0.001)
yearsq -0.01				0.010 (0.012)	0.010 (0.012)	0.010 (0.012)	0.010 (0.012)	0.010 (0.012)	
constant	-825.18 (28.49)**	-825.18 (28.49)**	-825.18 (28.49)**	-751.36 (112.6)**	-751.36 (112.6)**	-751.36 (112.6)**	-743.56 (100.56)**	-744.05 (99.76)**	-833.78 (55.30)**
Number of obs	2500	2500	2500	2500	2500	2500	2500	2500	2500
F-stats	1.1e+05	1.1e+05	1.1e+05	42955.83	42955.83	42955.83	53715.79	53715.89	53700.93
Root MSE	1.0093	1.0093	1.0093	1.0098	1.0098	1.0098	1.0096	1.0096	1.0097
Estimated second derivatives of age .00002 (.001), p=.988									
Estimated second derivatives of cohort 2.38e-06 (.00002), p=.922									
Estimated second derivatives of year -6.46e-06 (.0000), p=.803									
Mean(x_1)=6.31									

Table 3.3: Long Panel: Estimated restricted age, cohort and year effects for x
 $x = 0.3a + 0.3 + 0.7t + \epsilon$ Age ranges 20-78, year 1980-1999, and cohort 1921-1960

Real Coefficients	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
age 0.5	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
							agesq=0	cohortsq=0	yearsq=0
cohort 0.3		0.200 (0.001)**	1.201 (0.002)**		0.207 (0.008)**	1.208 (0.007)**		0.208 (0.003)**	1.209 (0.005)**
year 0.7	-0.200 (0.001)**		1.001 (0.002)**	-0.207 (0.008)**		1.001 (0.009)**	-0.194 (0.003)**		1.002 (0.005)**
agesq -0.04	1.201 (0.002)**	1.001 (0.002)**		1.208 (0.007)**	1.001 (0.009)**		1.202 (0.007)**	1	
cohortsq -0.02				-0.001 (.00008)*	-0.001 (.00008)*	-0.001 (.00008)*		-0.001 (.00006)**	-0.001 (.00007)*
yearsq -0.01				-0.0002 (.0001)	-0.0002 (.0001)	-0.0002 (.0001)	-0.002 (.00008)*		-0.0002 (.0001)
constant	-1974.78 (3.88)**	-1974.78 (3.88)**	-1974.78 (3.88)**	-0.0006 (.0003)	-0.0006 (.0003)	-0.0006 (.0003)	-0.0008 (.0003)	0.0007 (.0003)	-1976.50 (9.23)**
				-1973.90 (17.08)**	-1973.90 (17.08)**	-1973.90 (17.08)**	-1989.25 (14.87)**	-1972.39 (13.58)**	
Number of obs	10000	10000	10000	10000	10000	10000	10000	10000	10000
F-stats	2.6e+05	2.6e+05	2.6e+05	1.0e+05	1.0e+05	1.0e+05	1.3e+05	1.3e+05	1.3e+05
Root MSE	1.011	1.011	1.011	1.0108	1.0108	1.0108	1.0109	1.0107	1.0107
Estimated second derivatives of age .00002 (.0004), p=.961									
Estimated second derivatives of cohort -7.22e-07 (.00001), p=.948									
Estimated second derivatives of year -2.18e-06 (.00001), p=.840									
Mean(x)=27.02									

Table 3.4: Long Panel: Estimated restricted age, cohort and year effects for x_1
 $x_1 = 0.5a - 0.3c + 0.7t + \epsilon$ Age ranges 20-78, year 1980-1999, and cohort 1921-1960

	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
Real Coefficients	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
age 0.5		0.801 (0.001)**	1.198 (0.002)**		0.807 (0.008)**	1.205 (0.007)**		0.804 (0.003)**	1.202 (0.005)**
cohort -0.3	-0.801 (0.001)**		0.397 (0.002)**	-0.807 (0.008)**		0.397 (0.009)**	-0.800 (0.003)**		0.395 (0.005)**
year 0.7	1.198 (0.002)**	0.397 (0.002)**		1.205 (0.007)**	0.397 (0.009)**		1.202 (0.007)**	0.400 (0.007)**	
agesq -0.04				-0.0007 (.0001)	-0.0007 (.0001)	-0.0007 (.0001)		-0.0005 (.0001)	-0.0008 (.0001)
cohortsq -0.02				.00004 (.0001)	.00004 (.0001)	.00004 (.0001)	-0.00003 (.0001)		.00005 (.0001)
yearsq -0.01				-0.0001 (.0003)	-0.0001 (.0003)	-0.0001 (.0003)	-0.0002 (.0003)	-0.00016 (.0003)	
constant	-812.87 (3.87)**	-812.87 (3.87)**	-812.87 (3.87)**	-814.99 (17.04)**	-814.99 (17.04)**	-814.99 (17.04)**	-823.06 (14.83)**	-818.99 (13.55)**	-809.22 (9.21)**
Number of obs	10000	10000	10000	10000	10000	10000	10000	10000	10000
F-stats	6.6e+05	6.6e+05	6.6e+05	2.6e+05	2.6e+05	2.6e+05	3.3e+05	3.3e+05	3.3e+05
Root MSE	1.0083	1.0083	1.0083	1.0084	1.0084	1.0084	1.0084	1.0084	1.0084
Estimated second derivatives of age 9.02e-07 (.0004), p=.998									
Estimated second derivatives of cohort -3.94e-07 (.000), p=.972									
Estimated second derivatives of year -2.05e-06 (.000), p=.850									
Mean(x_1)=15.33									

Table 3.5: Short Panel: Estimated restricted age, cohort and year effects for y
 $y = 0.5a - 0.04a^2 + 0.3c - 0.02c^2 + 0.7t - 0.01t^2 + \epsilon$ Age ranges 20-63, year 1980-1984, and cohort 1921-1960

	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
Real	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
Coefficients							agesq=0	cohortsq=0	yearsq=0
age 0.5		-0.743 (0.013)**	-0.584 (0.104)**	-0.231 (0.052)**	0.231 (0.052)**	1.248 (0.057)**		1.783 (0.009)**	1.158 (0.031)**
cohort 0.3	0.743 (0.013)**		0.159 (0.105)			1.016 (0.059)**	3.083 (0.011)**		0.922 (0.029)**
year 0.7	-0.584 (0.104)**	0.159 (0.105)		1.248 (0.057)**	1.016 (0.059)**		-0.330 (0.084)**	0.178 (0.060)**	
agesq -0.04				-0.040 (0.001)**	-0.040 (0.001)**	-0.040 (0.001)**		-0.059 (0)**	-0.040 (0.001)**
cohortsq -0.02				-0.020 (0.001)**	-0.020 (0.001)**	-0.020 (0.001)**	-0.060 (0)**		-0.020 (0.001)**
yearsq -0.01				-0.023 (0.012)*	-0.023 (0.012)*	-0.023 (0.012)*	-0.063 (0.020)**	-0.005 (0.014)	
constant	-300.20 (207.50)	-300.20 (207.50)	-300.20 (207.50)	-2005.99 (115.14)**	-2005.99 (115.14)**	-2005.99 (115.14)**	-5312.89 (166.99)**	-407.84 (118.95)**	-1818.91 (56.58)**
Number of obs	2500	2500	2500	2500	2500	2500	2500	2500	2500
F-stats	1717.81	1717.81	1717.81	59622.00	59622.00	59622.00	27885.67	54667.26	74452.64
Root MSE	7.3501	7.3501	7.3501	1.0326	1.0326	1.0326	1.6765	1.2038	1.0331
Estimated second derivatives of age -.002 (.001), p=.097									
Estimated second derivatives of cohort -.00002 (.00003), p=.471									
Estimated second derivatives of year -.00002 (.00003), p=.409									
Mean(y)=-16.22									

Table 3.6: Short Panel: Estimated restricted age, cohort and year effects for y_1
 $y_1 = 0.5a - 0.04a^2 + 0.5c - 0.04c^2 + 0.7t - 0.01t^2 + \epsilon$ Age ranges 20-63, year 1980-1984, and cohort 1921-1960

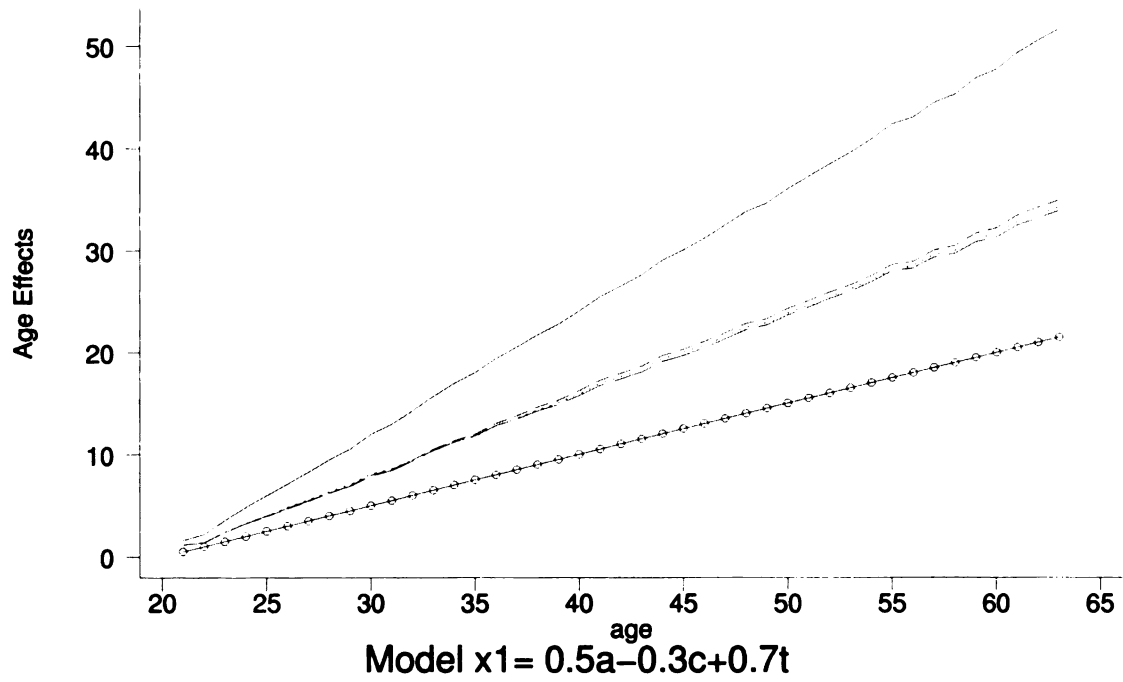
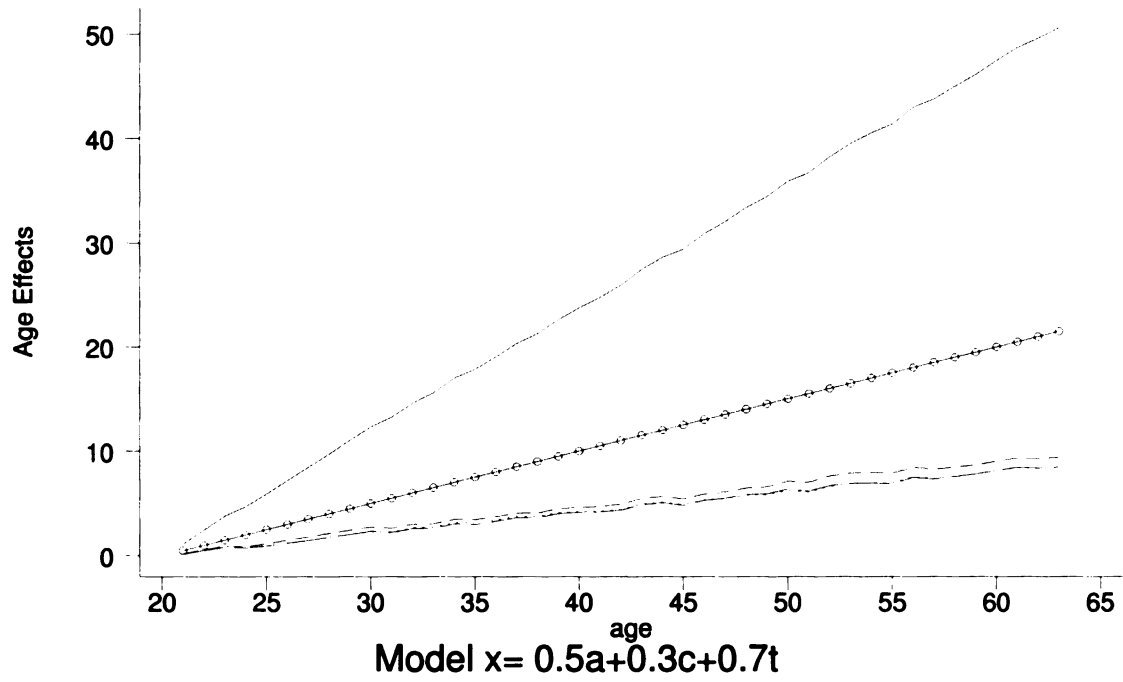
	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
Real	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
Coefficients									
cohort 0.5	0.163 (0.017)**	(0.017)**	(0.137)** -0.405 (0.138)**	-0.045 (0.051)	(0.051)	(0.056)** 1.229 (0.057)**	3.275 (0.011)**	cohortsq=0 (0.011)**	yearsq=0 (0.030)** 1.126 (0.029)**
year 0.7	-0.569 (0.137)**	-0.405 (0.138)**		1.274 (0.056)**	1.229 (0.057)**		-0.306 (0.083)**	-0.451 (0.080)**	
agesq -0.04				-0.040 (0.001)**	-0.040 (0.001)**	-0.040 (0.001)**		-0.077 (0)**	-0.041 (0.001)**
cohortsq -0.04				-0.039 (0.001)**	-0.039 (0.001)**	-0.039 (0.001)**	-0.080 (0)**		-0.039 (0.001)**
yearsq -0.01				-0.025 (0.012)**	-0.025 (0.012)**	-0.025 (0.012)**	-0.066 (0.020)**	0.011 (0.019)	
constant	787.58 (272.68)**	787.58 (272.68)**	787.58 (272.68)**	-2416.68 (112.63)**	-2416.68 (112.63)**	-2416.68 (112.63)**	-5728.83 (165.78)**	786.49 (158.10)**	-2212.26 (55.36)**
Number of obs	2500	2500	2500	2500	2500	2500	2500	2500	2500
F-stats	56.30	56.30	56.30	47225.78	47225.78	47225.78	21348.37	23151.52	58952.19
Root MSE	9.659	9.659	9.659	1.0101	1.0101	1.0101	1.6644	1.6000	1.0108
Estimated second derivatives of age -.002 (.001), p=.091									
Estimated second derivatives of cohort -.00004 (.00002), p=.064									
Estimated second derivatives of year -.00003 (.00003), p=.246									
Mean(y_1)=-22.59									

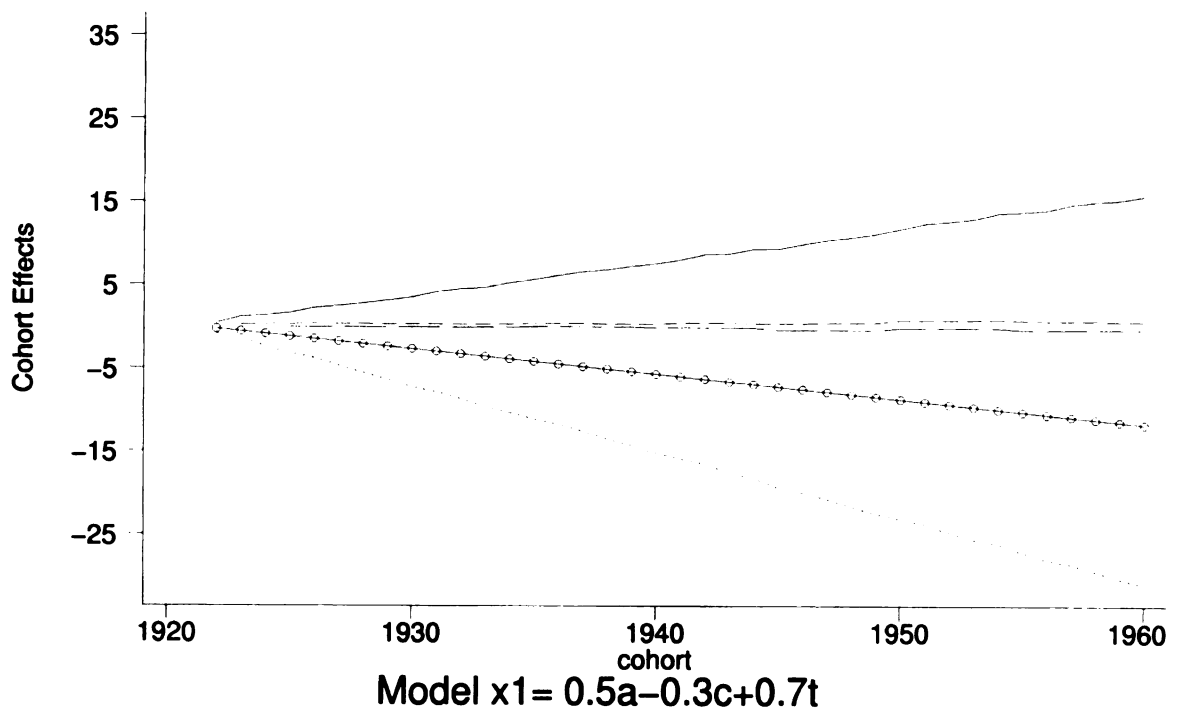
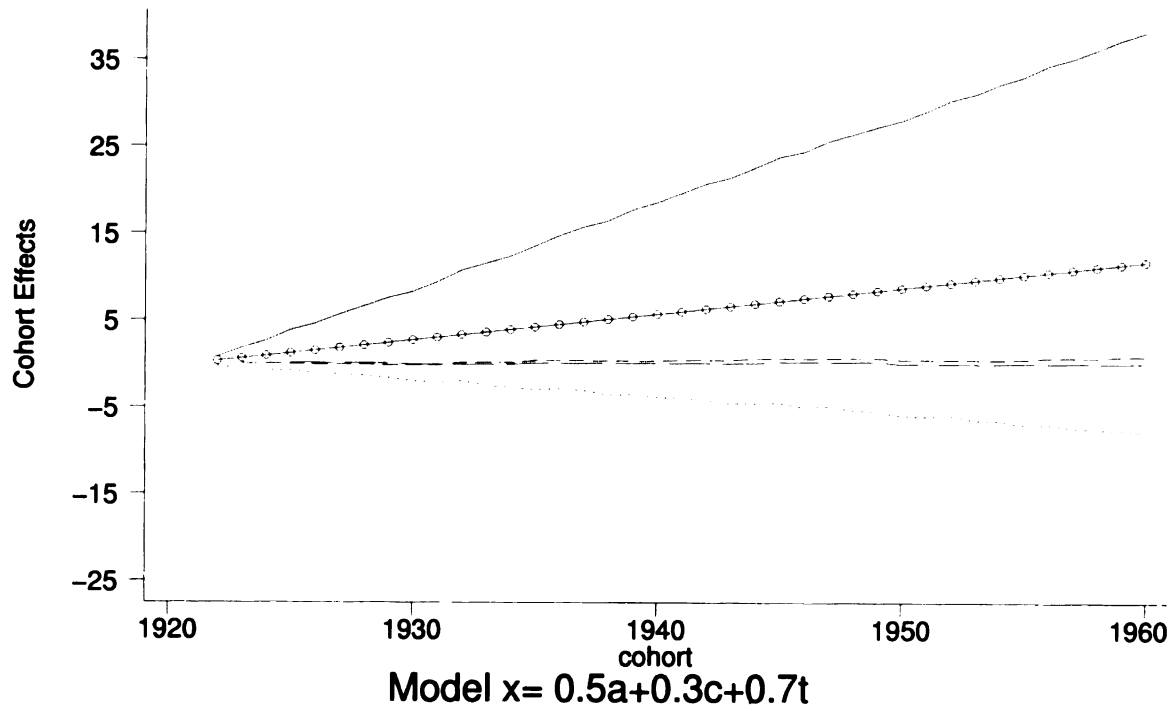
Table 3.7: Long Panel: Estimated restricted age, cohort and year effects for $y = 0.5a - 0.04a^2 + 0.3c - 0.02c^2 + 0.7t - 0.01t^2 + \epsilon$ Age ranges 20-78, year 1980-1999, and cohort 1921-1960

	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
Real Coefficients	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
age 0.5		-1.344 (0.008)**	-1.309 (0.016)**		0.201 (0.008)**	1.200 (0.007)**		1.490 (0.007)**	1.038 (0.005)**
cohort 0.3	1.344 (0.008)**		0.035 (0.018)**	-0.201 (0.008)**		0.999 (0.009)**	3.677 (0.018)**		0.792 (0.005)**
year 0.7	-1.309 (0.016)**	0.035 (0.018)**		1.200 (0.007)**	0.999 (0.009)**		-0.361 (0.036)**	0.056 (0.014)**	
agesq -0.04				-0.040 (0)**	-0.040 (0)**	-0.040 (0)**		-0.049 (0)**	-0.040 (0)**
cohortsq -0.02				-0.020 (0)**	-0.020 (0)**	-0.020 (0)**	-0.060 (0)**		-0.019 (0)**
yearsq -0.01				-0.010 (0)**	-0.010 (0)**	-0.010 (0)**	-0.050 (0.002)**	-0.001 (0.001)	
constant	-28.37 (34.84)	-28.37 (34.84)	-28.37 (34.84)	-1971.32 (17.03)**	-1971.32 (17.03)**	-1971.32 (17.03)**	-6406.17 (79.67)**	-159.18 (27.33)**	-1562.44 (9.56)**
Number of obs	10000	10000	10000	10000	10000	10000	10000	10000	10000
F-stats	18063.01	18063.01	18063.01	7.5e+05	7.5e+05	7.5e+05	29869.66	2.3e+05	8.6e+05
Root MSE	9.0776	9.0776	9.0776	1.0073	1.0073	1.0073	5.418	2.0343	1.0475
Estimated second derivatives of age -.002 (.000), p=.000									
Estimated second derivatives of cohort -.00002 (.000), p=.038									
Estimated second derivatives of year -.00001 (.000), p=.338									
Mean(y)=-24.83									

Table 3.8: Long Panel: Estimated restricted age, cohort and year effects for y_1
 $y_1 = 0.5a - 0.04a^2 + 0.5c - 0.04c^2 + 0.7t - 0.01t^2 + \epsilon$ Age ranges 20-78, year 1980-1999, and cohort 1921-1960

Real Coefficients	ModelConstraints			ModelConstraints			ModelConstraints		
	1	2	3	4	5	6	7	8	9
age 0.5	age=0	cohort=0	year=0	age=0	cohort=0	year=0	age=0	cohort=0	year=0
cohort 0.5							agesq=0	cohortsq=0	yearsq=0
year 0.7									
agesq -0.04									
cohortsq -0.04									
yearsq -0.01									
constant									
Number of obs	1094.67	1094.67	1094.67	10000	10000	10000	10000	10000	10000
F-stats	(42.42)**	(42.42)**	(42.42)**	5502.59	5502.59	5502.59	19234.64	45196.81	5.9e+05
Root MSE	11.054	11.054	11.054	1.0001	1.0001	1.0001	5.4331	3.6675	1.0444
Estimated second derivatives of age -0.002 (.000), p=.000									
Estimated second derivatives of cohort -.00005 (.000), p=.000									
Estimated second derivatives of year -9.07e-06 (.000), p=.339									
Mean(y_1)=-31.20									





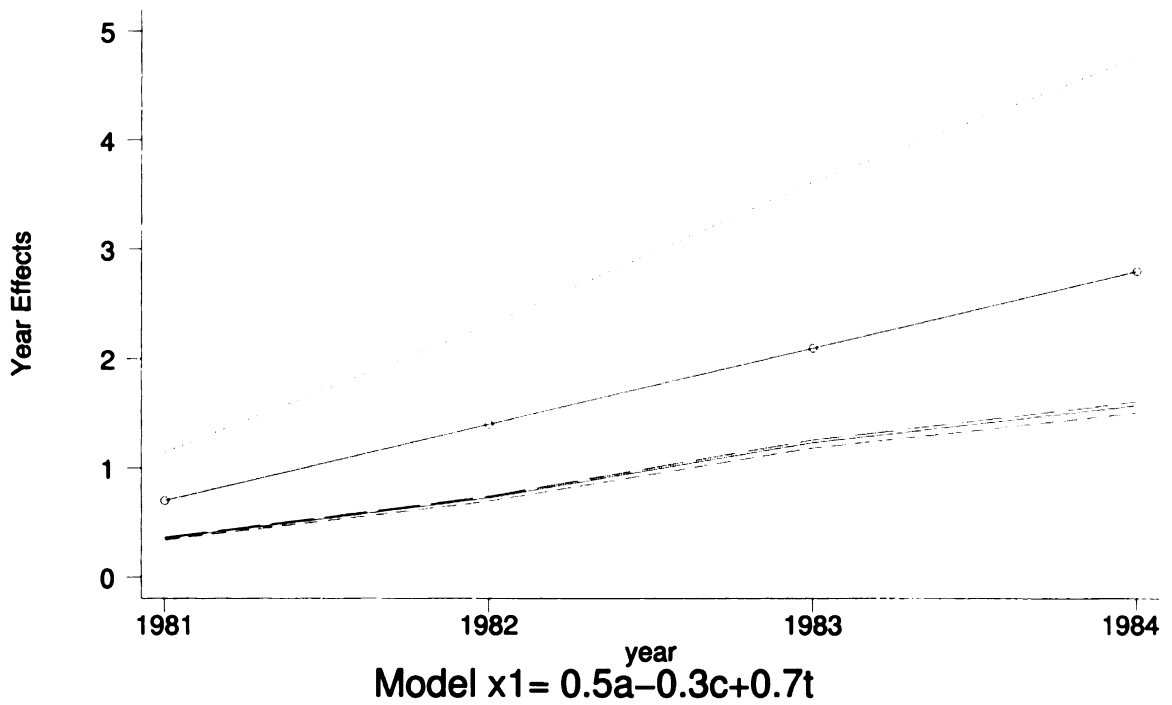
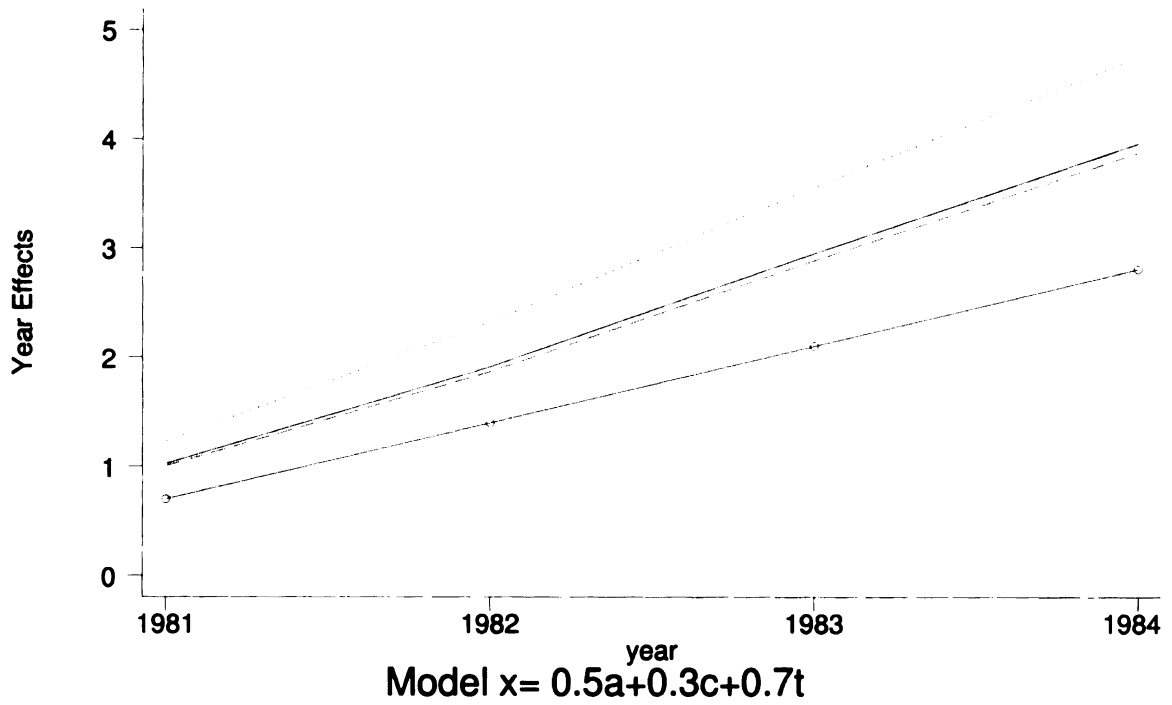
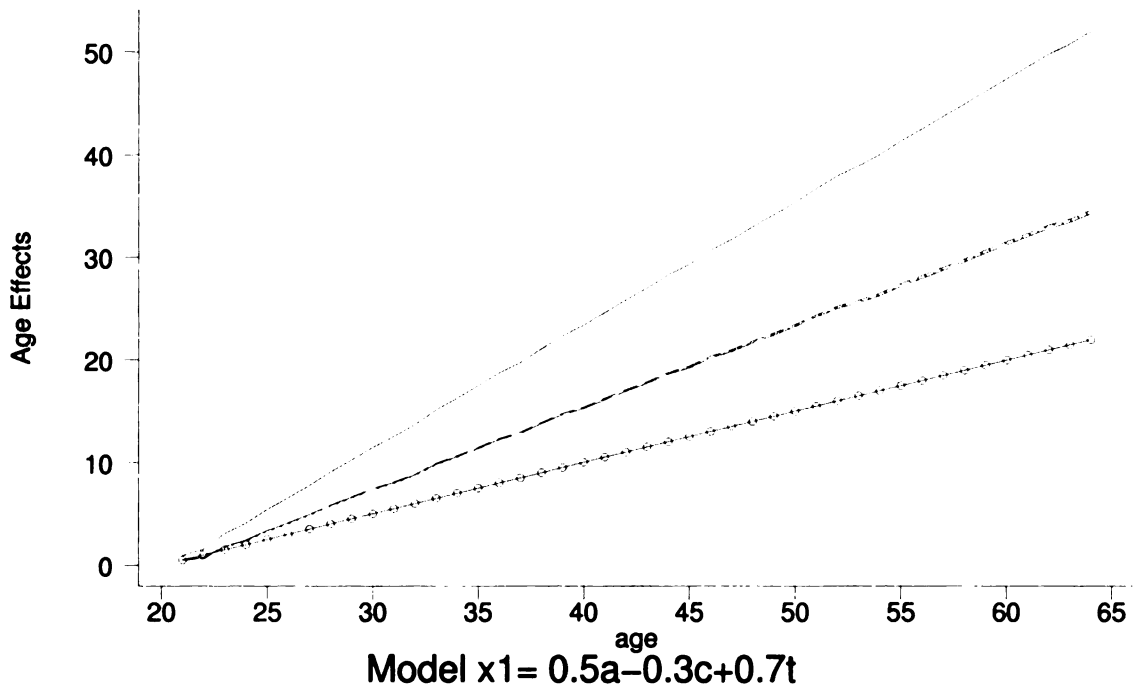
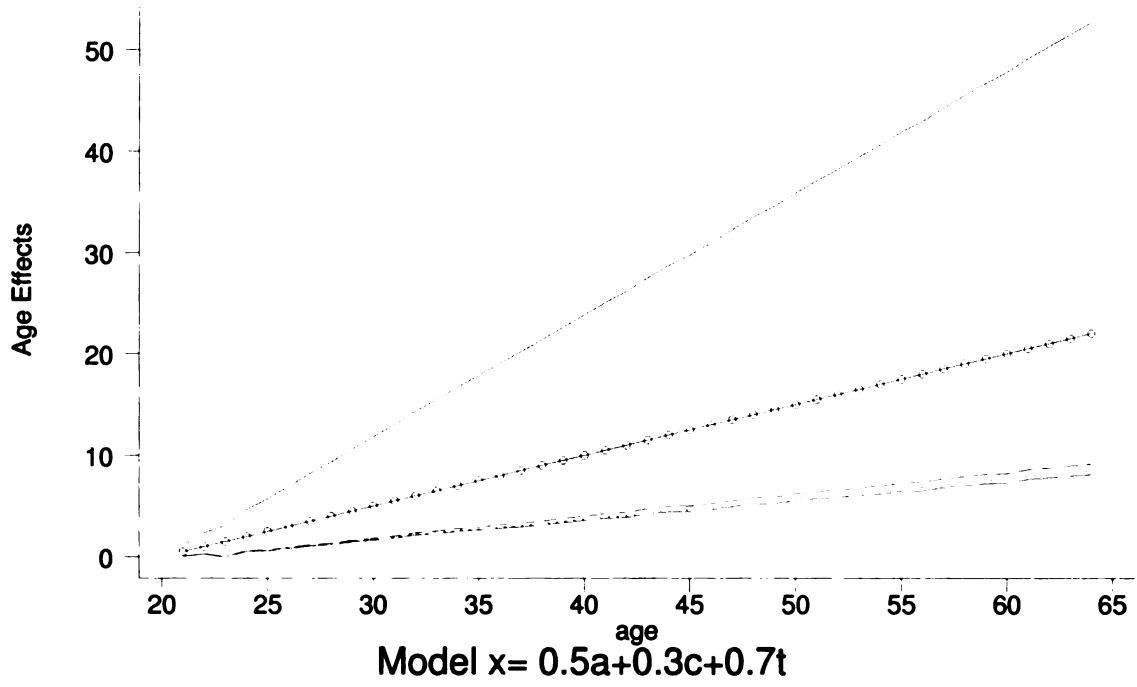
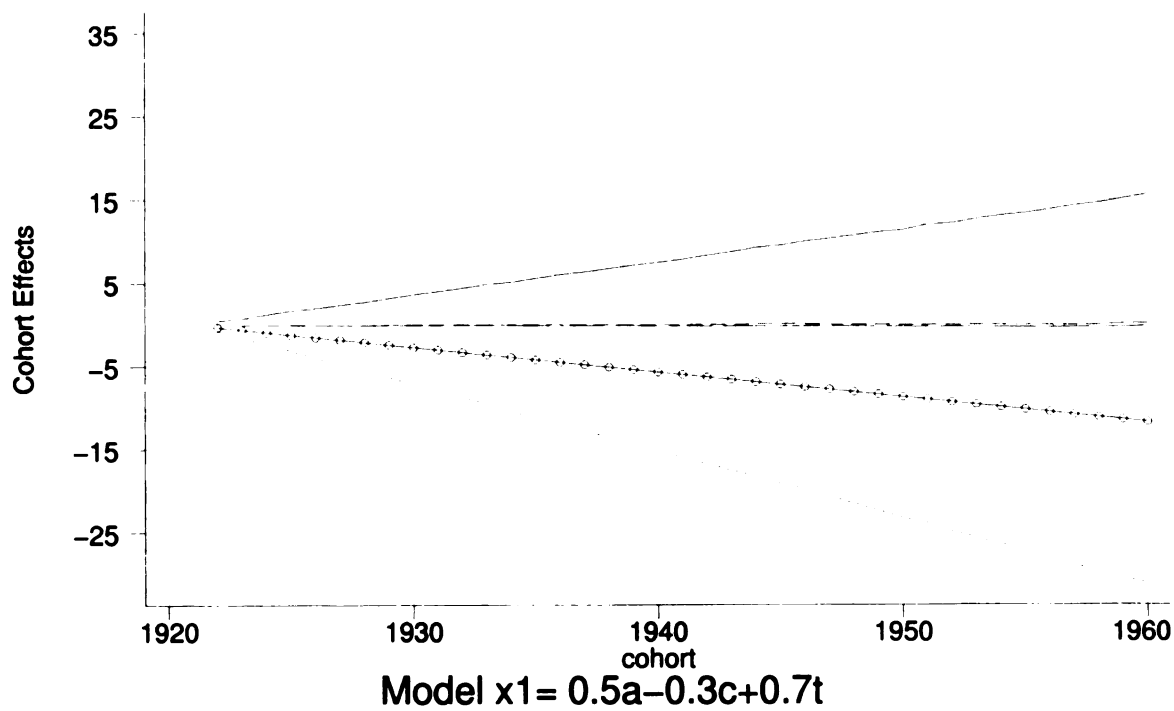
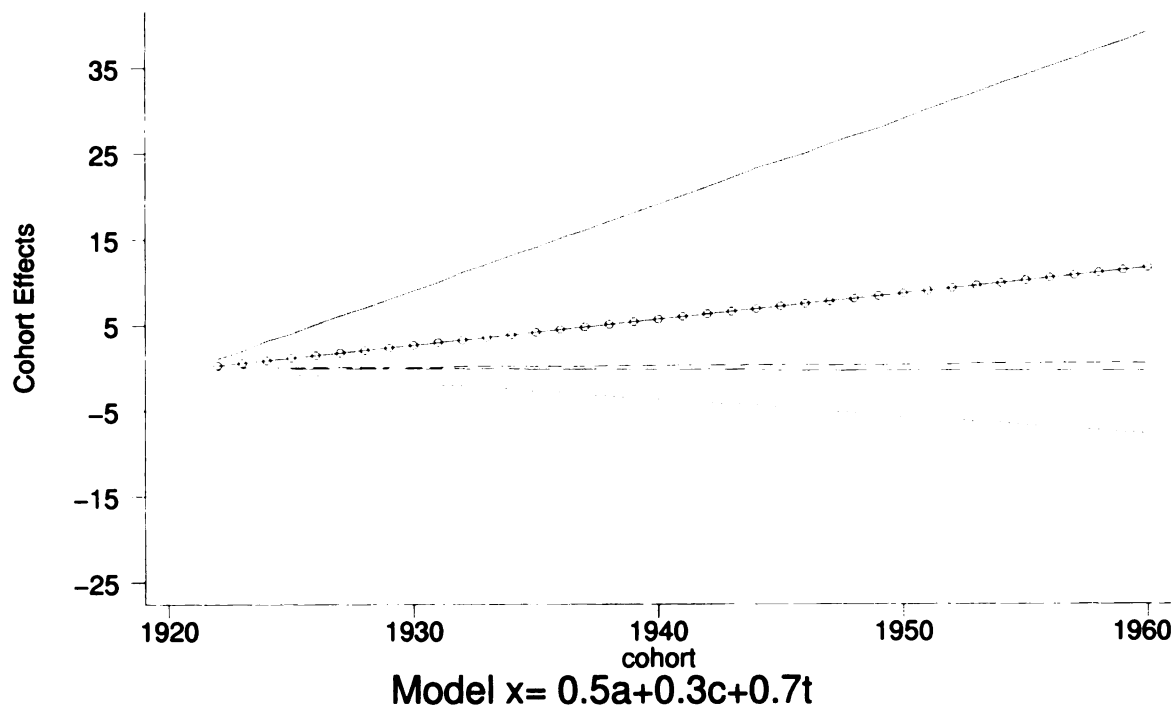


Figure 3.3: Short Panel Estimates of Age, Cohort and Year Effects for x and x_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.





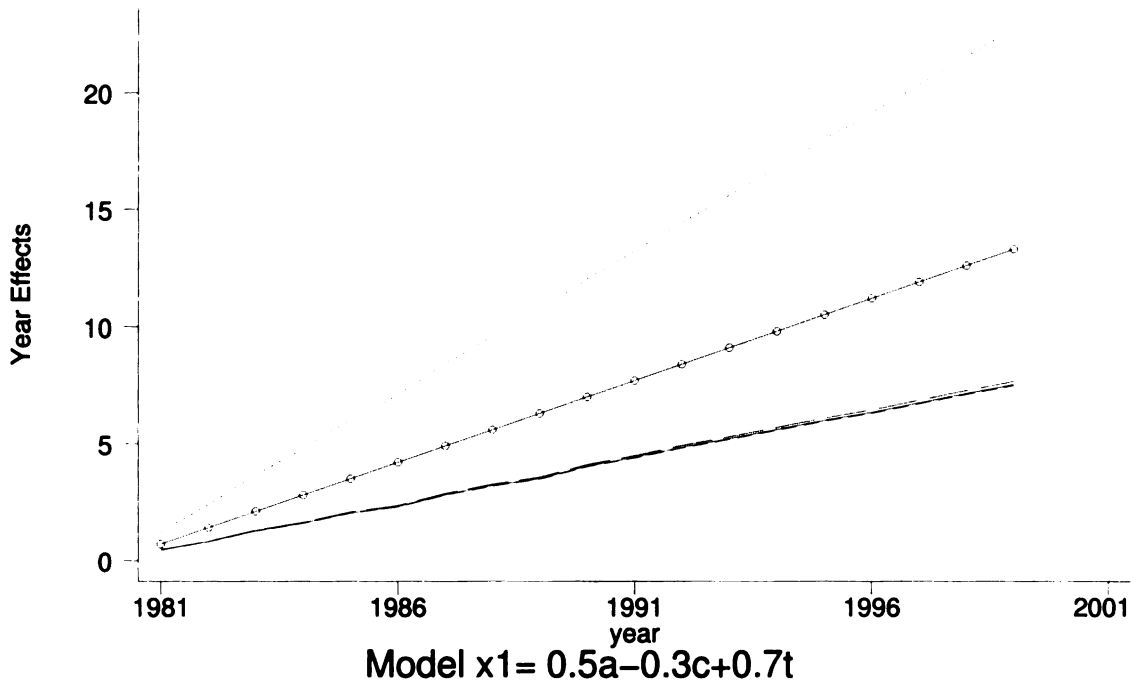
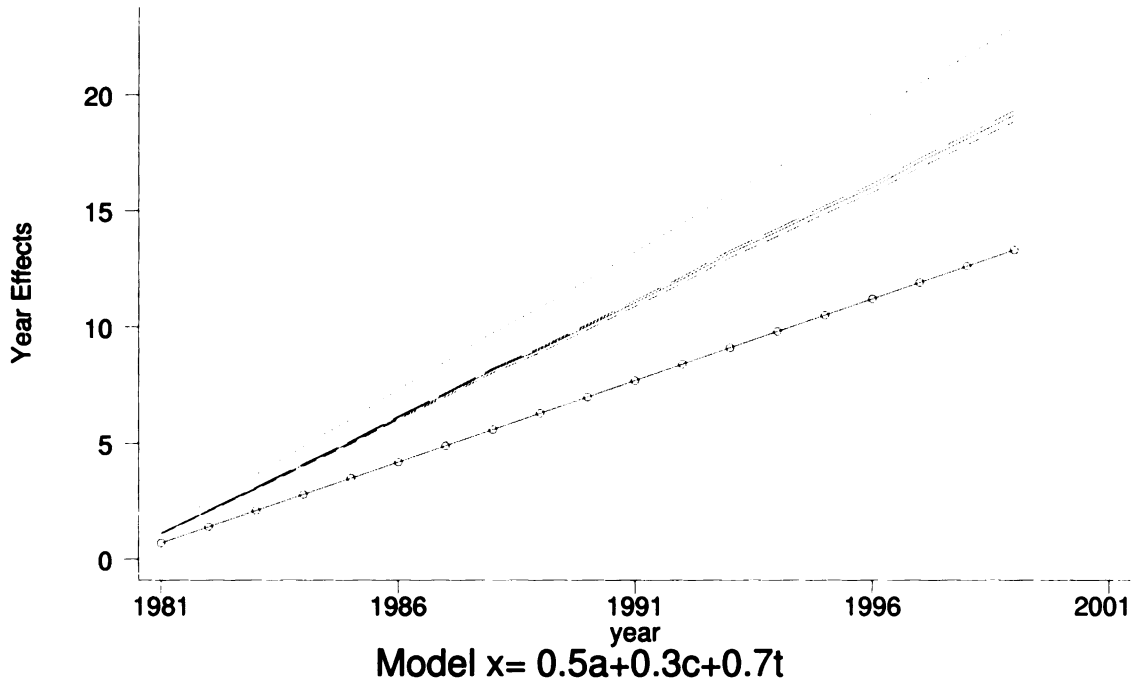
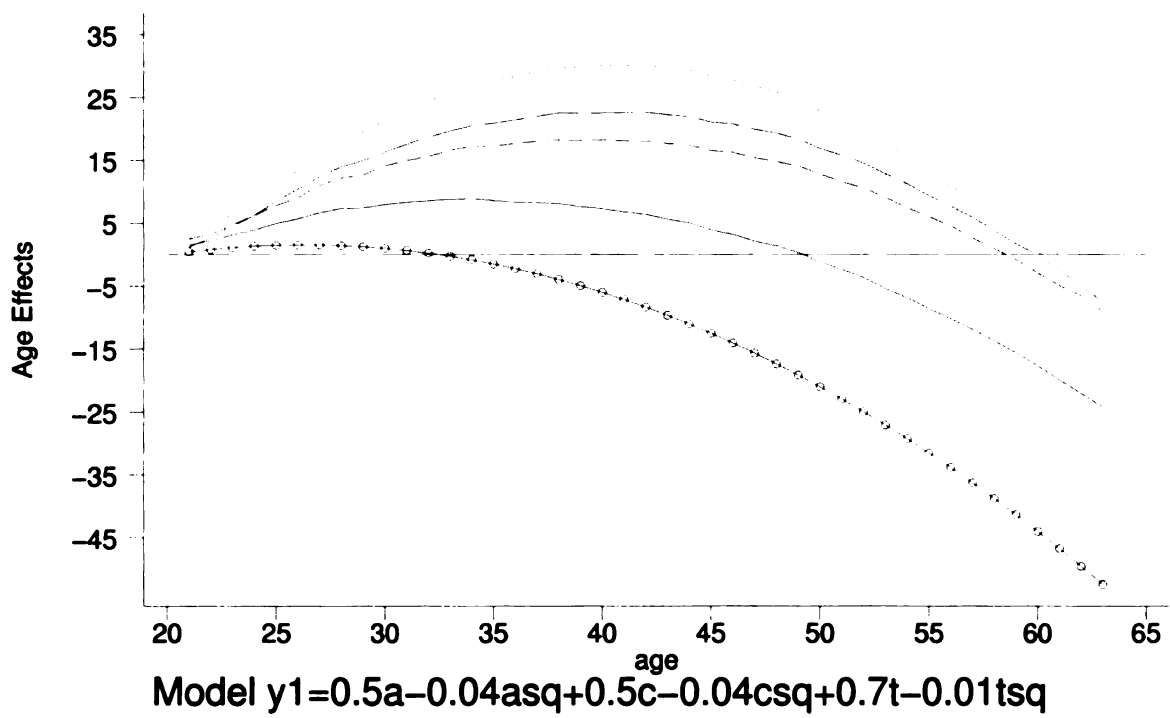
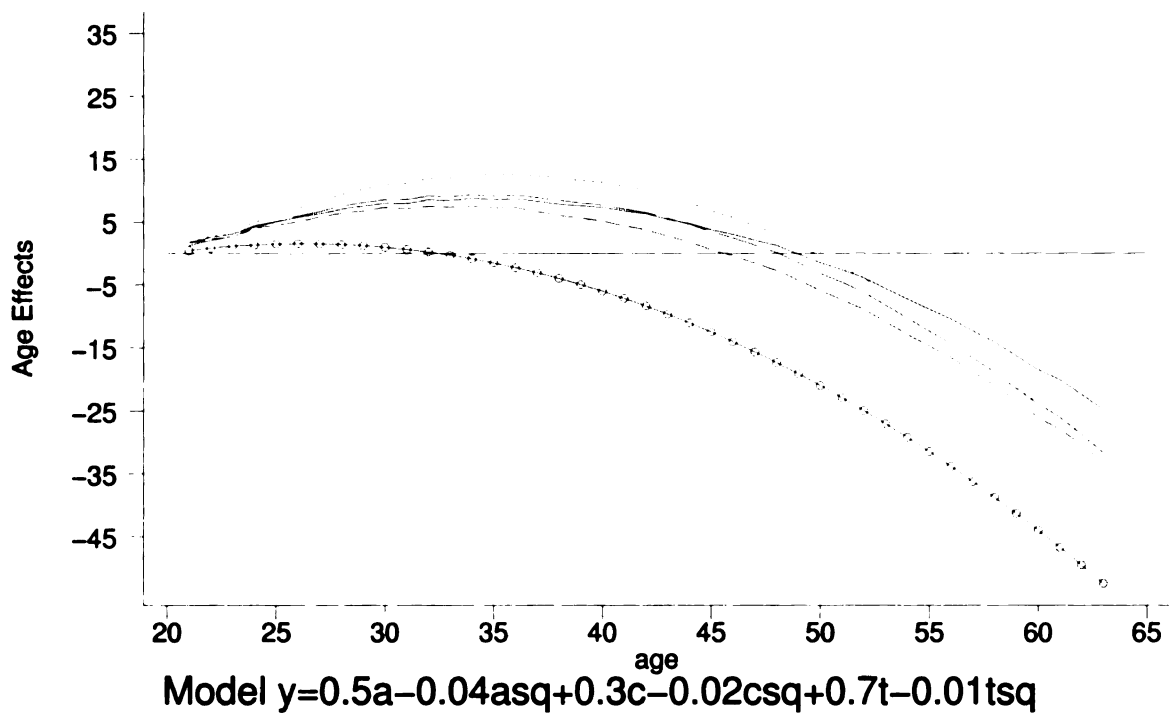
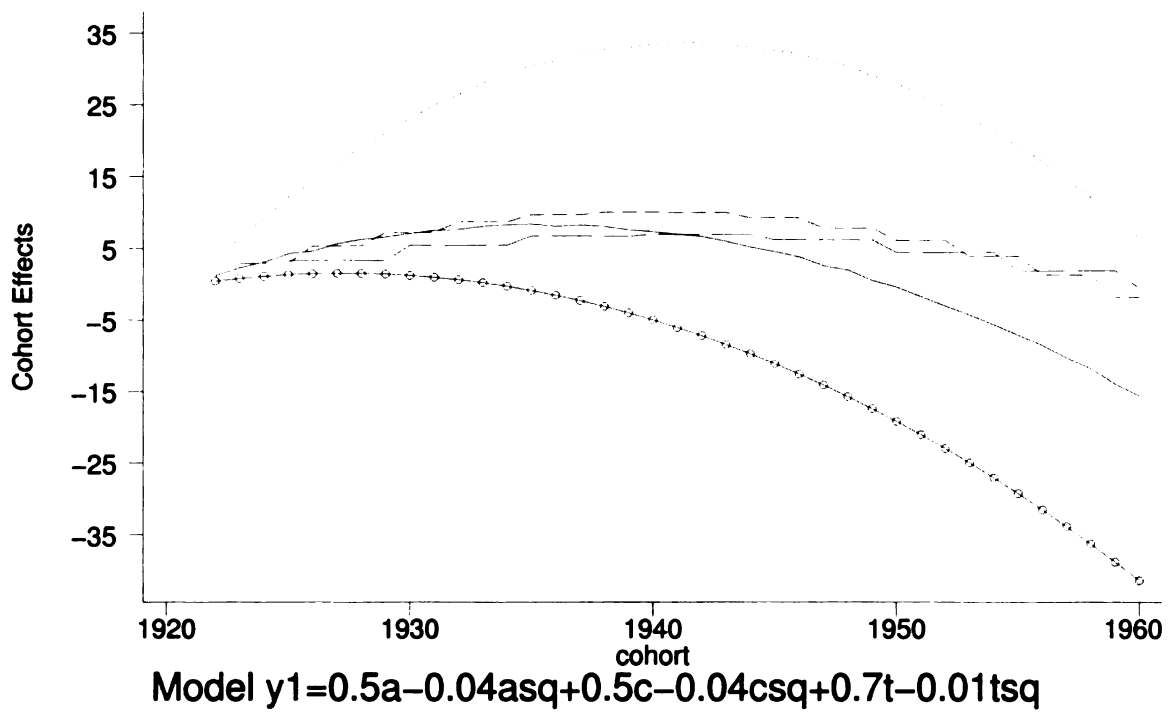
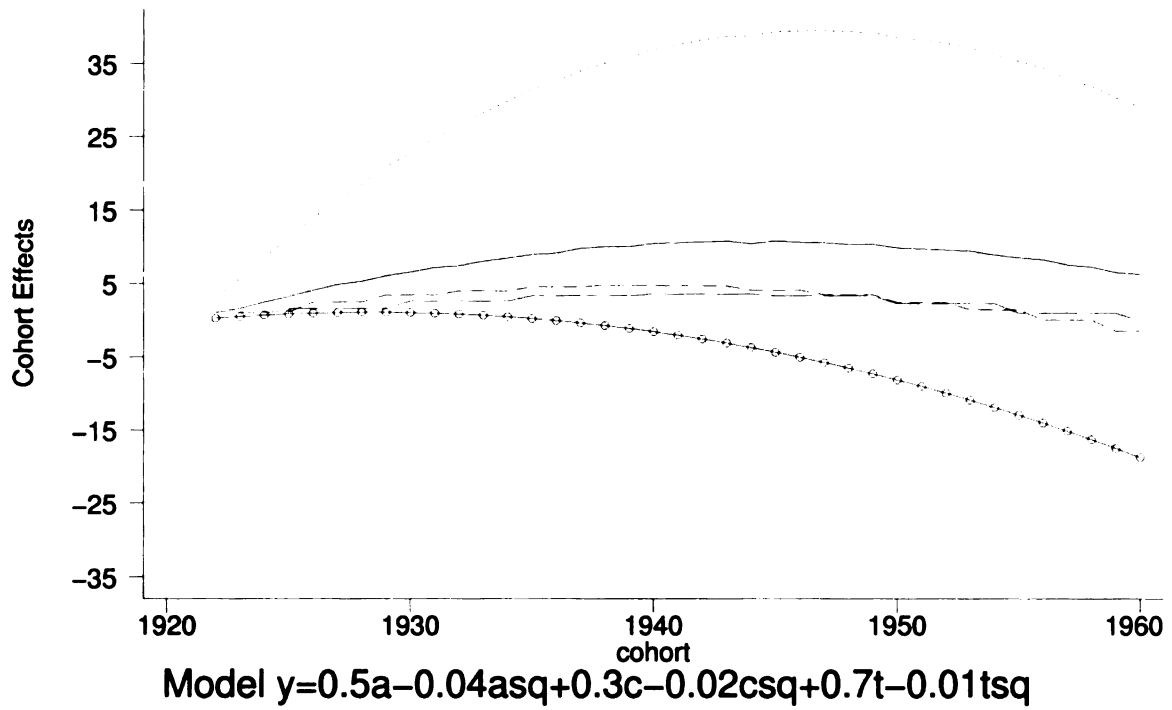


Figure 3.4: Long Panel Estimates of Age, Cohort and Year Effects for x and x_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.





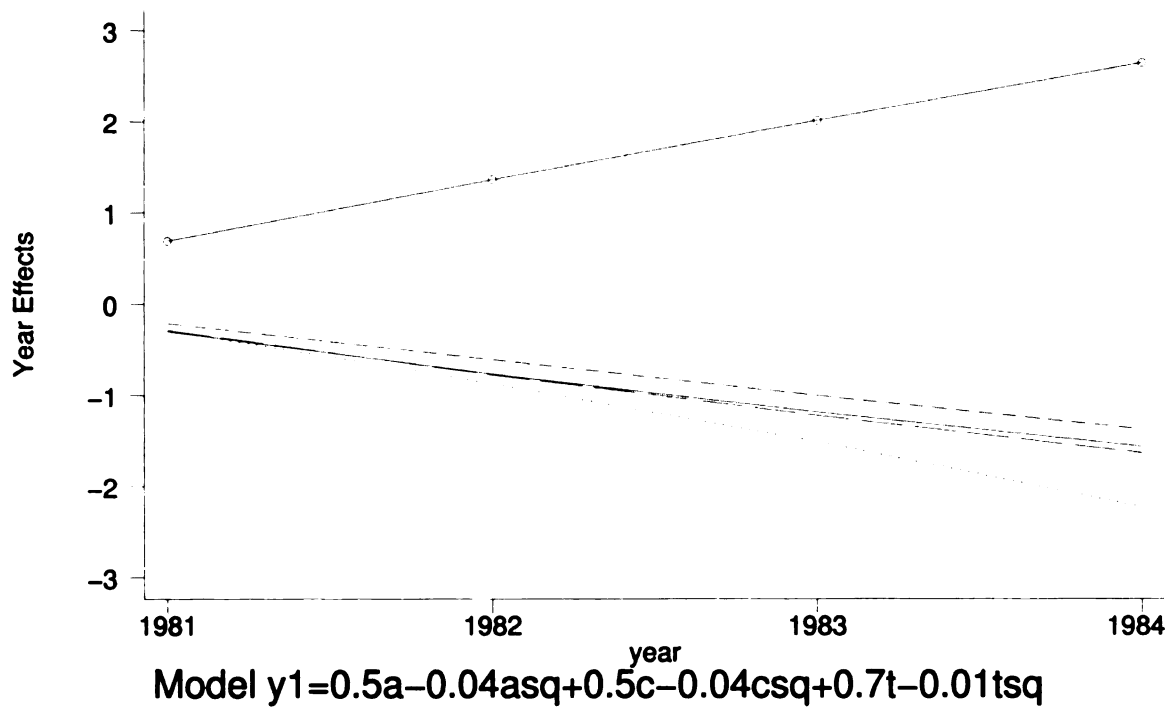
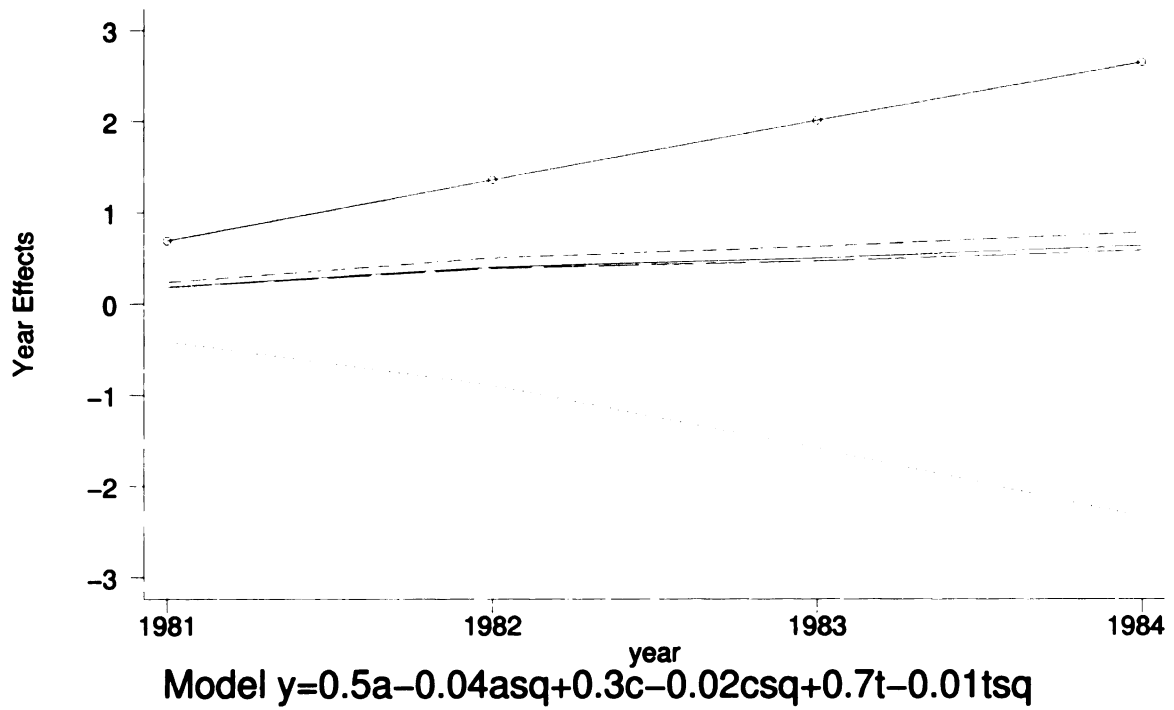
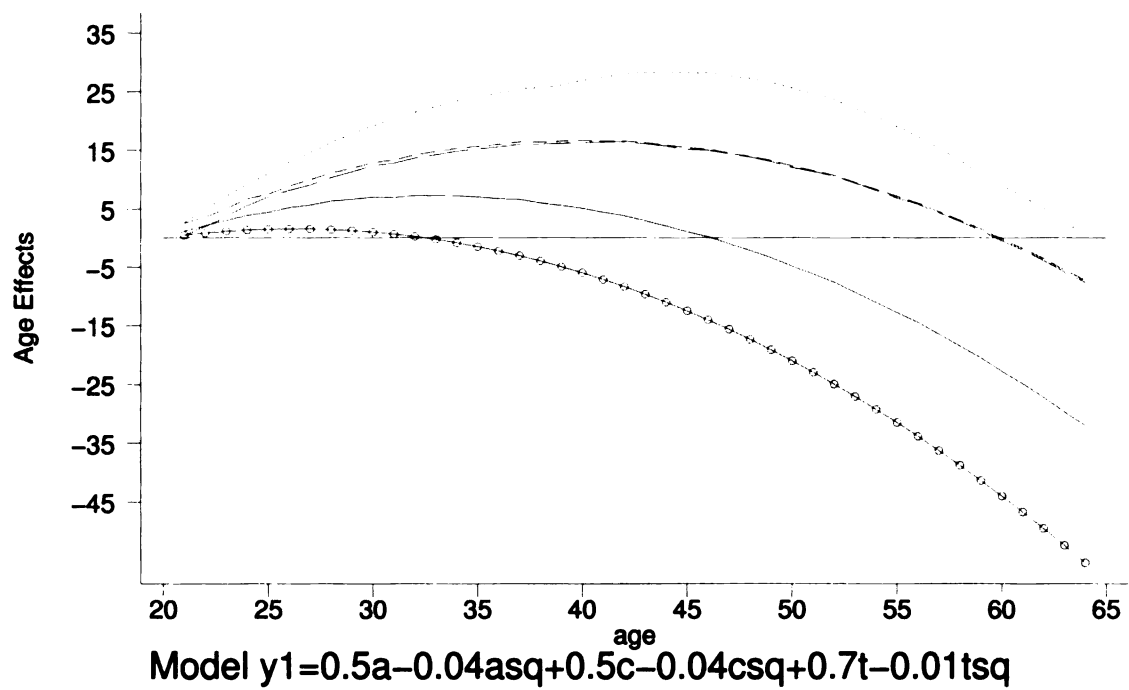
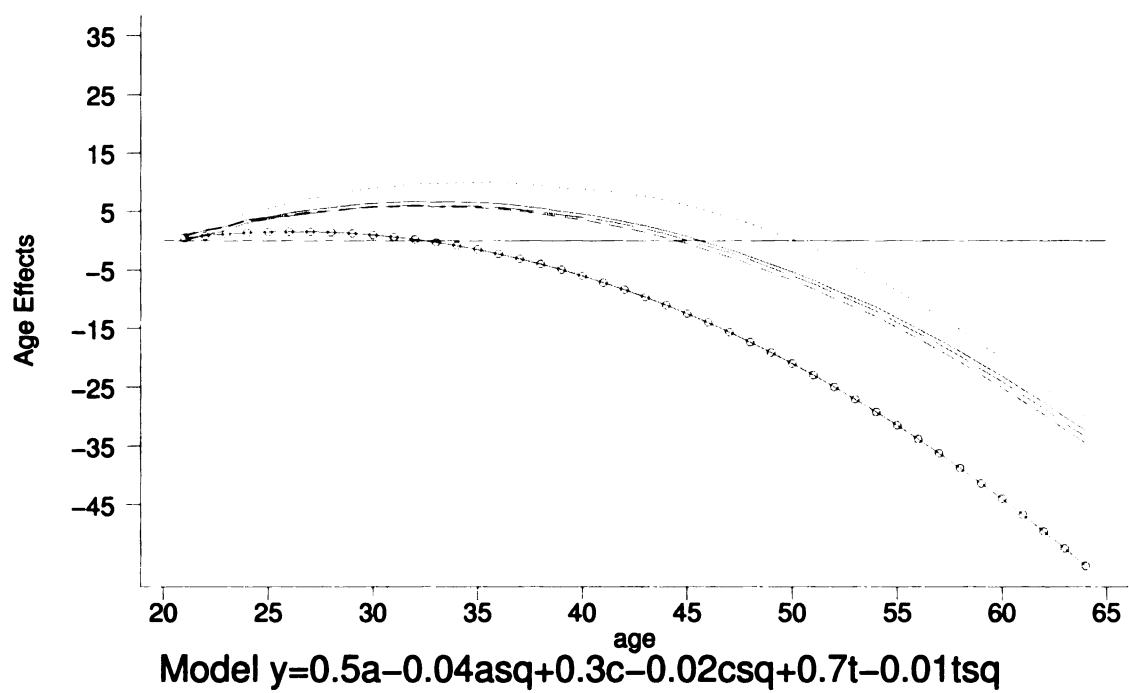
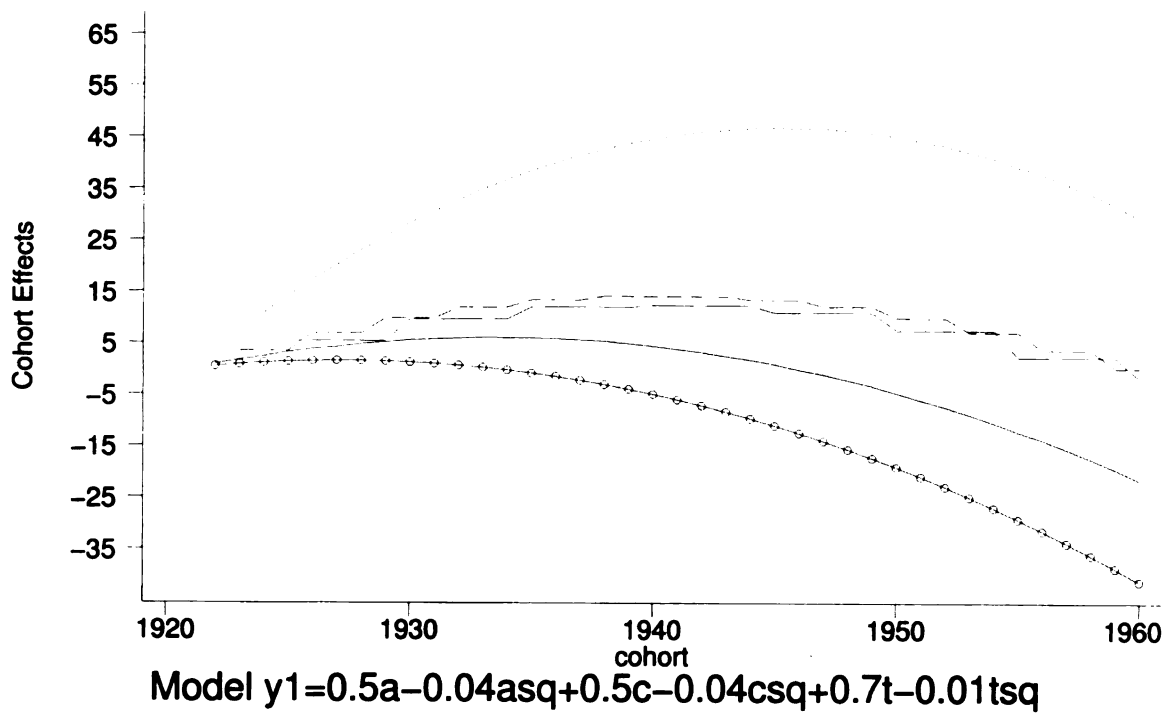
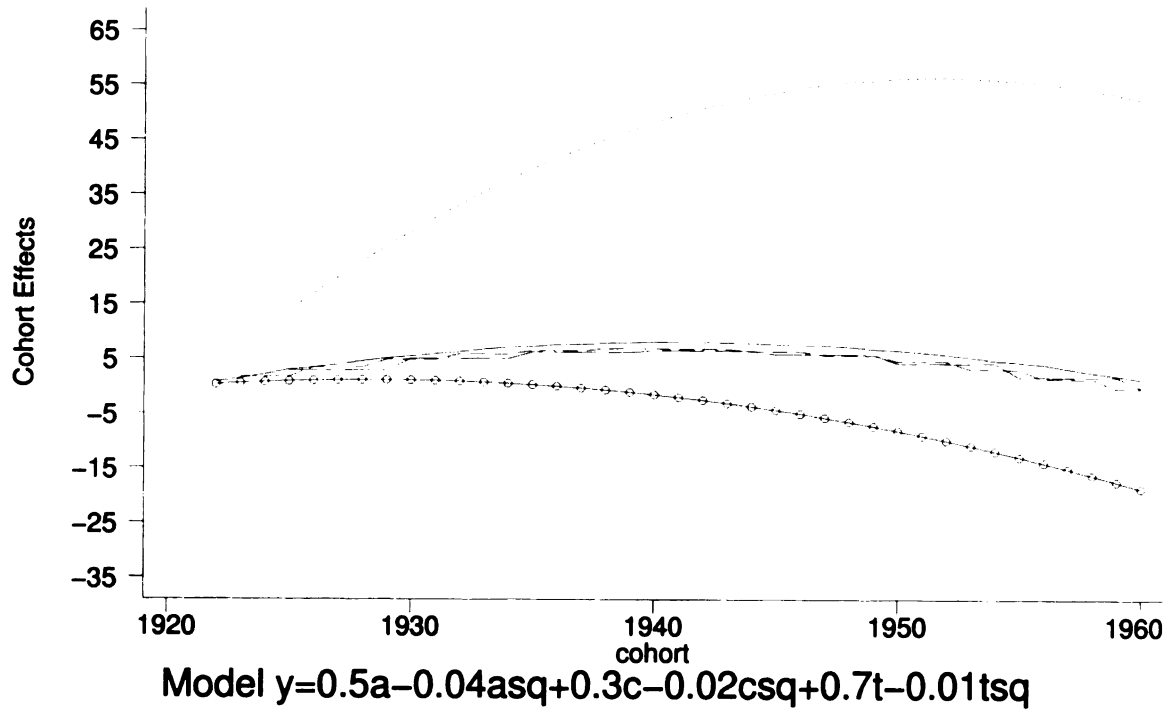


Figure 3.5: Short Panel Estimates of Age, Cohort and Year Effects for y and y_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.





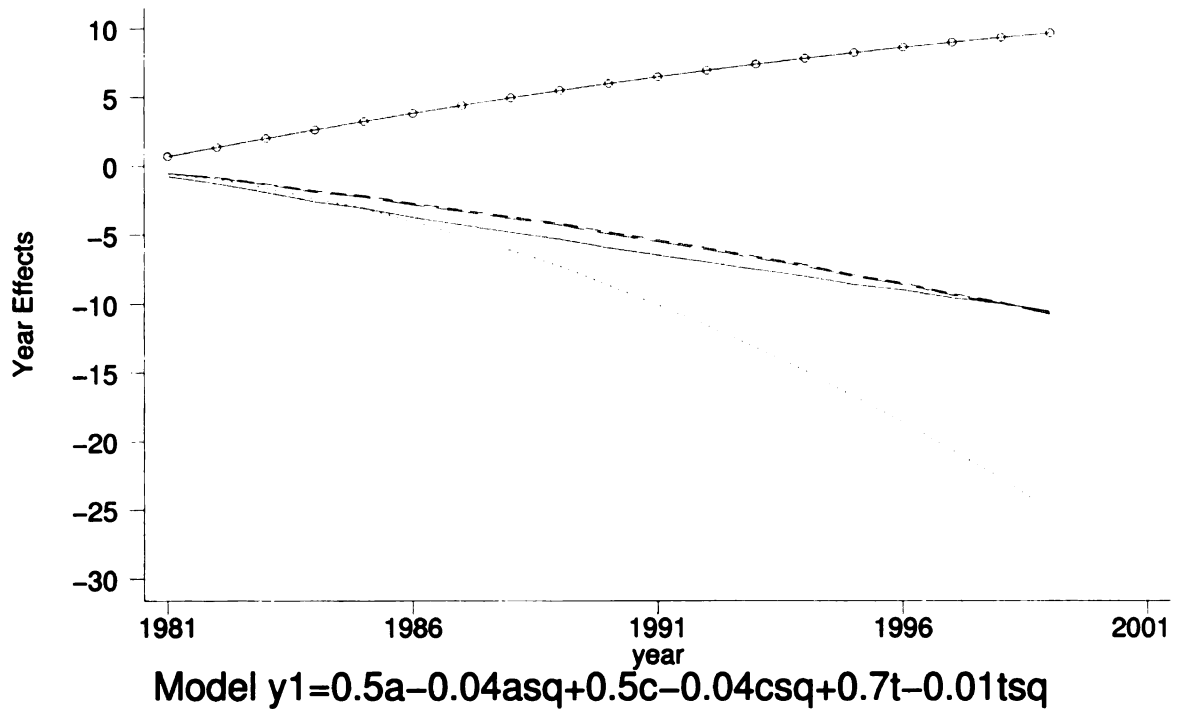
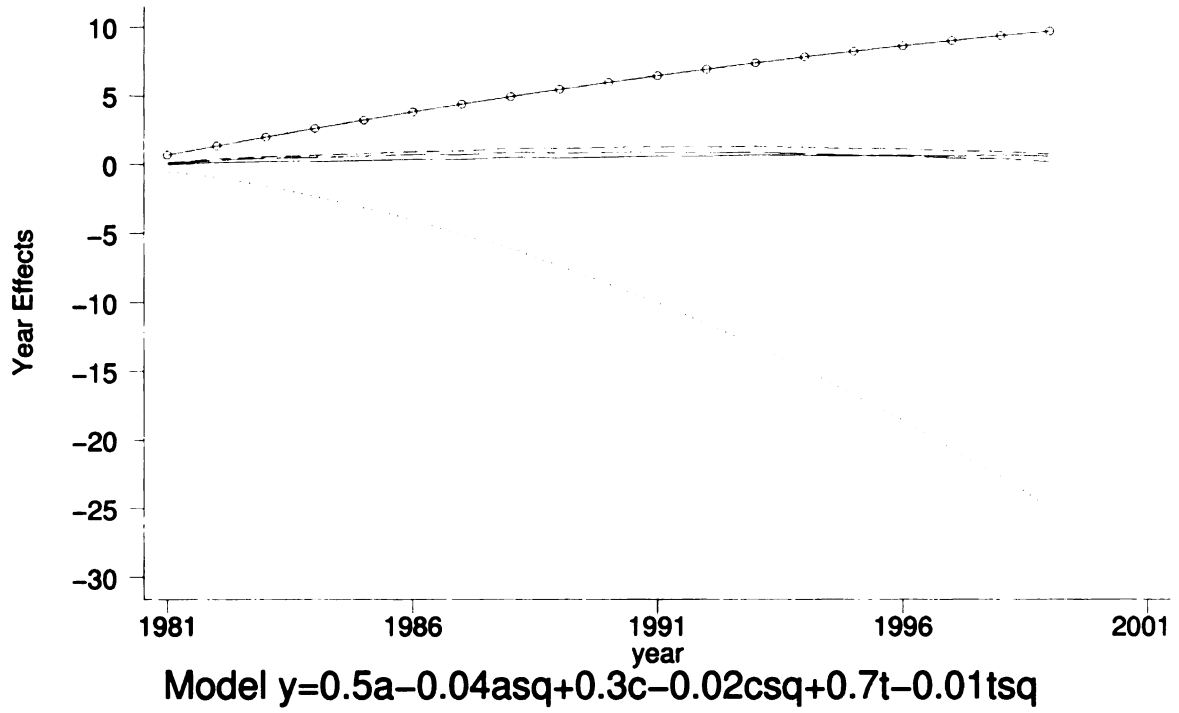


Figure 3.6: Long Panel Estimates of Age, Cohort and Year Effects for y and y_1 . Identifications achieved through (1) model age, cohort excludes year effects; (2) model age, year excludes cohort effects; (3) model age, cohort excludes year effects; (4) model age, five-year-cohort, year effects; and (5) model age, three-year-cohort, year effects.

3.4 BMI Profiles in CHNS

Figure 3.7 shows the OLS estimates of the second differences of the age and cohort effects using 1989, 1991 and 1993 data assuming the data generating process 3.1 is true for all individuals. The asterisks are from the approach based on repeated cross-sections, or the mean approach, and the connected lines are based on the genuine panel data approach, or the level approach.

The two approaches both indicate much variation in the slopes of the age and cohort effects after age 55. The standard errors based on the mean approach need to be corrected. No inference is based on the mean approach without correcting the standard errors. The standard errors in the level approach are valid provided that $\epsilon_{i,c,a,t} | (u, \alpha_c, \beta_a, \gamma_t) \sim i.i.d.(0, \sigma^2)$. The F tests of the null hypothesis that all changes in slopes equal to zero or a constant based on the individual level data are valid. The F-statistic for zero changes in male age effects is 1.71 (p=.0005), for female age effects is 1.17 (p=.1724), for changes in male cohort effects is 3.03 (p=.0000) and for changes in female cohort effects is 1.57 (p=.0026). This suggests that the age and cohort profiles are not quadratic. The growth rate in BMI is relatively constant around age 45.

Since the identification of the age and cohort effects are more reliable for those aged 20 to 45, we test the null hypothesis that the changes in slopes of the age effects in this range equal to zero or a constant and the the null hypothesis that the changes in slopes of the cohort effects in the range of 1944 to 1973 equal to zero or a constant. The F-statistic for changes in the age range for male age effects is 1.25 (p=.1881), for female is 0.55 (p=.9620), and the F-statistic for changes in the cohort range for male cohort effects is 1.62 (p=.0191), and for female is 1.59 (p=.0233). As we can see we can not reject the null of zero changes in the slopes of age effects in this range for men nor women. However, the changes in the slopes of the cohort effects from 1944 to 1973 are different from zero.

Using 1989, 1993 and 1997 data the changes in the slope of age and cohort effects have similar shape as in the two-year spaced case. Figure 3.8 indicates less variation in some ranges of age and cohort profiles for men and women. Table 9 shows the regression results of equation (3.11) with two-year spaced and four-year spaced data. Overall, there is less significant changes in five-year differences. This can be viewed as evidence that we could regroup our data into five or three-year cohorts, which essentially imposes the restriction that within each cohort group the cohort effects are the same.

3.5 Estimate the Restricted Level Effects

As demonstrated before and verified by the software in our application we need two meaningful constraints to identify all age, period and cohort effects. By grouping our data into five- or three-year cohorts we are in effect imposing more constraints than necessary. The design matrices are full rank in both cases. This is the first identification strategy.

The second strategy is to maintain some cohorts as single-year cohorts and group others, based on the changes in slopes in Table 3.9. The three-year cohorts are July 1906 - June 1909, July 1909 - June 1912, ..., July 1972 - June 1975, and the last two years July 1975 - June 1977. From Table 3.9, we can see the changes from 1927 to 1931, 1933 to 36, 36 to 39 are significantly different from zero for men and changes from 1939 to 42 and 63 to 66 are not zero for women. We can break these cohort groups down to each original birth cohort year. Doing so we still have enough restrictions to identify the system but do not impose what might be wrong constraints. Similarly, for the five-year cohorts specification the cohort groups are July 1906 - June 1912, July 1912 - June 1917, ..., July 1967 - June 1972, and July 1972 to June 1977. We break the groups 1952 - 1957 and 1957 - 1962 down to each original birth cohort

Table 3.9: Second Differences of Cohort Profiles

Differences	Two-year Spaced Data 1989, 91, 93		Differences	Four-year Spaced Data 1989, 93, 97	
	Male	Female		Male	Female
1915-1914-(1913-1912)	4.249 (2.416)*	2.247 (2.177)	1915-1914-(1911-1910)	8.710 (3.331)**	1.537 (3.601)
1916-1915-(1914-1913)	-0.772 (2.237)	-2.787 (2.309)	1916-1915-(1912-1911)	-0.033 (3.846)	-4.652 (3.601)
1917-1916-(1915-1914)	-3.077 (2.237)	-0.426 (2.309)	1917-1916-(1913-1912)	-3.387 (3.846)	2.028 (3.601)
1918-1917-(1916-1915)	0.925 (1.581)	3.490 (2.177)	1918-1917-(1914-1913)	3.942 (2.720)	2.180 (2.401)
1919-1918-(1917-1916)	0.798 (1.708)	-2.792 (2.177)	1919-1918-(1915-1914)	-5.021 (2.720)*	-0.476 (2.172)
1920-1919-(1918-1917)	-0.672 (1.708)	2.121 (1.686)	1920-1919-(1916-1915)	0.591 (2.355)	2.715 (2.547)
1921-1920-(1919-1918)	0.487 (1.708)	-1.437 (1.686)	1921-1920-(1917-1916)	1.958 (1.848)	-4.764 (2.547)*
1922-1921-(1920-1919)	-0.293 (1.708)	0.988 (1.633)	1922-1921-(1918-1917)	0.615 (1.848)	0.380 (2.401)
1923-1922-(1921-1920)	0.622 (1.528)	-1.286 (1.633)	1923-1922-(1919-1918)	-0.382 (1.923)	-2.023 (2.401)
1924-1923-(1922-1921)	-1.791 (1.528)	0.400 (1.425)	1924-1923-(1920-1919)	-3.301 (1.923)*	0.007 (1.747)
1925-1924-(1923-1922)	1.485 (1.234)	3.250 (1.425)**	1925-1924-(1921-1920)	2.590 (1.923)	-1.116 (1.572)
1926-1925-(1924-1923)	1.707 (1.118)	-2.454 (1.234)**	1926-1925-(1922-1921)	-0.034 (1.923)	-1.156 (1.801)
1927-1926-(1925-1924)	-0.641 (1.099)	-1.488 (1.234)	1927-1926-(1923-1922)	-2.661 (1.528)*	0.558 (1.801)
1928-1927-(1926-1925)	-0.767 (1.080)	2.053 (1.137)*	1928-1927-(1924-1923)	-0.905 (1.332)	-0.782 (1.572)
1929-1928-(1927-1926)	1.443 (0.854)*	-0.926 (1.104)	1929-1928-(1925-1924)	-1.284 (1.332)	1.161 (1.572)
1930-1929-(1928-1927)	-3.656 (1.015)**	-0.895 (1.104)	1930-1929-(1926-1925)	-0.637 (1.259)	-0.745 (1.273)
1931-1930-(1929-1928)	0.799 (1.015)	0.206 (1.074)	1931-1930-(1927-1926)	0.384 (1.081)	1.892 (1.273)
1932-1931-(1930-1929)	2.351 (0.986)**	1.657 (1.074)	1932-1931-(1928-1927)	-1.514 (1.081)	-0.263 (1.254)
1933-1932-(1931-1930)	1.162 (0.986)	-1.010 (1.020)	1933-1932-(1929-1928)	0.613 (0.982)	-1.262 (1.074)
1934-1933-(1932-1931)	-4.993 (1.015)**	0.814 (1.046)	1934-1933-(1930-1929)	1.874 (1.143)	1.222 (1.139)
1935-1934-(1933-1932)	2.276 (1.015)**	0.927 (1.046)	1935-1934-(1931-1930)	0.317 (1.143)	-0.077 (1.184)
1936-1935-(1934-1933)	3.519 (0.960)**	-1.396 (0.933)	1936-1935-(1932-1931)	-3.243 (1.067)**	-0.839 (1.184)
1937-1936-(1935-1934)	-3.037 (0.924)**	-1.916 (0.933)**	1937-1936-(1933-1932)	1.255 (1.004)	0.903 (1.098)
1938-1937-(1936-1935)	-1.688 (0.924)*	2.088 (0.953)**	1938-1937-(1934-1933)	1.503 (1.143)	-1.361 (1.019)
1939-1938-(1937-1936)	2.546 (0.854)**	1.047 (0.953)	1939-1938-(1935-1934)	-1.437 (1.143)	-0.548 (0.980)
1940-1939-(1938-1937)	-0.617 (1)	0.358 (0.973)	1940-1939-(1936-1935)	-1.576 (1.081)	0.990 (1.153)
1941-1940-(1939-1938)	-1.012 (1)	0.933 (0.973)	1941-1940-(1937-1936)	1.114 (1.004)	2.234 (1.153)*
1942-1941-(1940-1939)	0.701 (1.015)	-1.873 (0.984)*	1942-1941-(1938-1937)	0.426 (0.962)	-0.263 (1.029)
1943-1942-(1941-1940)	1.363 (1.015)	-2.024 (0.984)**	1943-1942-(1939-1938)	-1.131 (0.962)	-1.436 (1.029)
1944-1943-(1942-1941)	-0.250 (0.863)	2.343 (0.943)**	1944-1943-(1940-1939)	2.272 (1.126)**	0.548 (1.074)

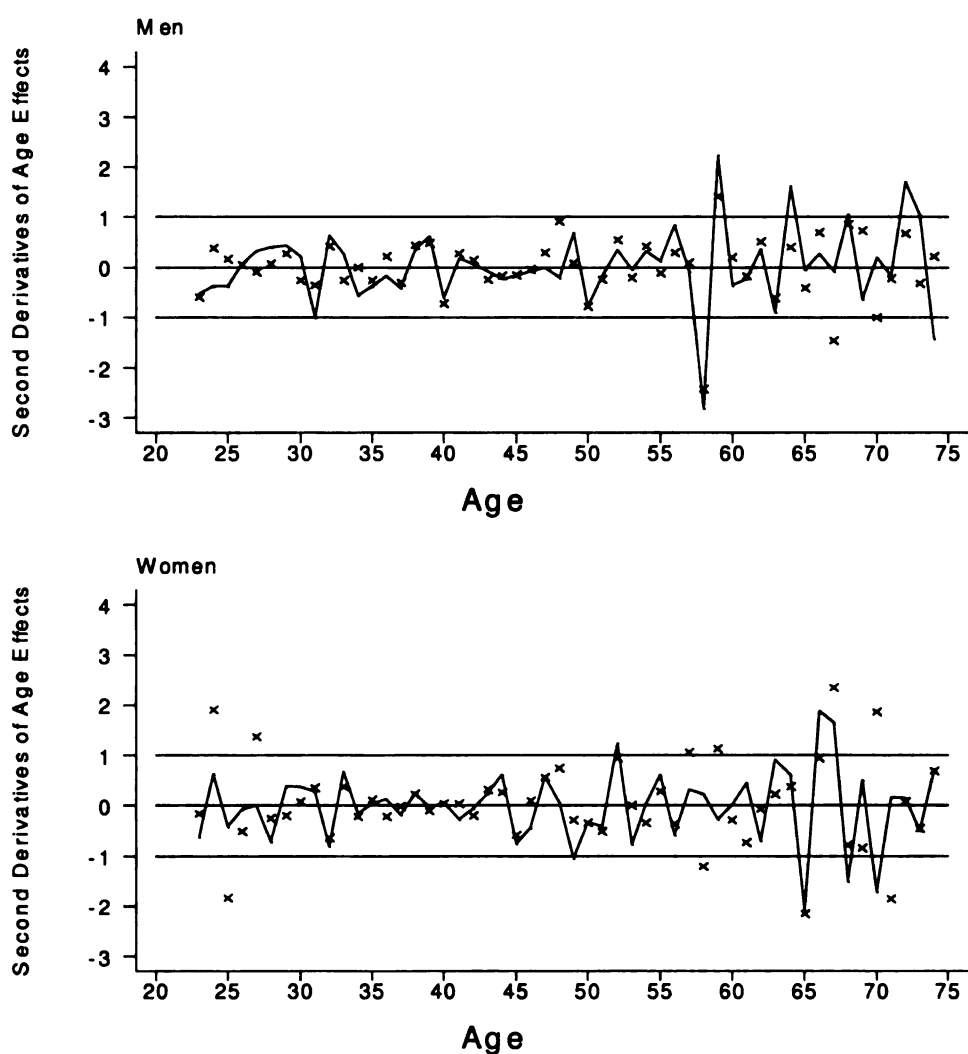
Table continues

Table 3.9 (cont'd)

Differences	Two-year Spaced Data 1989, 91, 93		Differences	Four-year Spaced Data 1989, 93, 97	
	Male	Female		Male	Female
1945-1944-(1943-1942)	-0.168 (0.845)	0.083 (0.943)	1945-1944-(1941-1940)	-0.739 (1.126)	-1.081 (1.074)
1946-1945-(1944-1943)	-0.934 (0.723)	-0.256 (0.798)	1946-1945-(1942-1941)	0.007 (0.907)	1.062 (0.980)
1947-1946-(1945-1944)	-0.072 (0.631)	-0.206 (0.692)	1947-1946-(1943-1942)	-0.104 (0.882)	1.275 (0.980)
1948-1947-(1946-1945)	-0.032 (0.523)	-0.191 (0.562)	1948-1947-(1944-1943)	-0.545 (0.814)	-2.054 (0.873)**
1949-1948-(1947-1946)	-0.017 (0.523)	-0.322 (0.530)	1949-1948-(1945-1944)	0.194 (0.710)	0.524 (0.759)
1950-1949-(1948-1947)	-0.085 (0.529)	0.223 (0.521)	1950-1949-(1946-1945)	0.372 (0.561)	-0.609 (0.618)
1951-1950-(1949-1948)	0.832 (0.529)	0.182 (0.502)	1951-1950-(1947-1946)	0.644 (0.561)	0.184 (0.596)
1952-1951-(1950-1949)	-0.349 (0.459)	-0.168 (0.498)	1952-1951-(1948-1947)	-0.717 (0.587)	-0.533 (0.590)
1953-1952-(1951-1950)	-0.500 (0.459)	-1.021 (0.474)**	1953-1952-(1949-1948)	-0.009 (0.587)	-0.606 (0.556)
1954-1953-(1952-1951)	0.355 (0.462)	0.541 (0.466)	1954-1953-(1950-1949)	-0.391 (0.533)	0.616 (0.556)
1955-1954-(1953-1952)	0.867 (0.462)*	0.866 (0.466)*	1955-1954-(1951-1950)	-0.225 (0.533)	-0.052 (0.530)
1956-1955-(1954-1953)	-0.937 (0.482)*	-0.342 (0.450)	1956-1955-(1952-1951)	0.369 (0.559)	0.237 (0.523)
1957-1956-(1955-1954)	0.119 (0.482)	-0.123 (0.459)	1957-1956-(1953-1952)	-0.159 (0.559)	1.674 (0.524)**
1958-1957-(1956-1955)	-0.175 (0.478)	-0.546 (0.533)	1958-1957-(1954-1953)	-0.130 (0.542)	-1.438 (0.629)**
1959-1958-(1957-1956)	0.010 (0.478)	0.286 (0.533)	1959-1958-(1955-1954)	-0.137 (0.542)	0.231 (0.629)
1960-1959-(1958-1957)	0.431 (0.620)	0.883 (0.617)	1960-1959-(1956-1955)	0.791 (0.691)	-0.600 (0.660)
1961-1960-(1959-1958)	-0.429 (0.631)	-0.208 (0.617)	1961-1960-(1957-1956)	-0.580 (0.702)	-0.550 (0.660)
1962-1961-(1960-1959)	1.433 (0.634)**	0.930 (0.586)	1962-1961-(1958-1957)	1.583 (0.702)**	1.696 (0.586)**
1963-1962-(1961-1960)	-0.883 (0.614)	-0.696 (0.586)	1963-1962-(1959-1958)	-0.947 (0.544)*	-1.487 (0.590)**
1964-1963-(1962-1961)	-1.289 (0.554)**	-0.604 (0.494)	1964-1963-(1960-1959)	0.256 (0.702)	1.037 (0.647)
1965-1964-(1963-1962)	0.815 (0.453)*	1.092 (0.484)**	1965-1964-(1961-1960)	0.581 (0.714)	0.292 (0.647)
1966-1965-(1964-1963)	-0.010 (0.515)	-1.588 (0.589)**	1966-1965-(1962-1961)	-1.288 (0.653)**	-1.613 (0.613)**
1967-1966-(1965-1964)	-0.757 (0.545)	0.751 (0.589)	1967-1966-(1963-1962)	0.726 (0.627)	0.690 (0.604)
1968-1967-(1966-1965)	0.885 (0.545)	-0.406 (0.550)	1968-1967-(1964-1963)	0.882 (0.582)	-0.691 (0.602)
1969-1968-(1967-1966)	0.801 (0.502)	0.226 (0.550)	1969-1968-(1965-1964)	-0.473 (0.525)	-0.237 (0.611)
1970-1969-(1968-1967)	-1.354 (0.502)**	0.615 (0.510)	1970-1969-(1966-1965)	-0.211 (0.555)	0.685 (0.644)
1971-1970-(1969-1968)	0.117 (0.638)	-1.051 (0.667)	1971-1970-(1967-1966)	0.383 (0.624)	0.303 (0.658)
1972-1971-(1970-1969)	0.969 (0.758)	-0.902 (0.829)	1972-1971-(1968-1967)	-0.371 (0.624)	-0.056 (0.647)
1973-1972-(1971-1970)	-3.588 (1.291)**	0.642 (1.425)	1973-1972-(1969-1968)	-0.734 (0.656)	-1.124 (0.791)
Number of obs	4284	4930	Number of obs	4287	4690
R-squared	0.0446	0.0205	R-squared	0.0203	0.019

because they are significantly different from zero as a group (Table 3.9).

Yet another way of breaking the perfect collinearity between age, period and cohort is, instead of defining age as period minus cohort (July to the next June cohort), to use the exact age at interview time (interview date - birth date). In this case for those who were born before July the exact linear dependency between age, period and cohort (July to the next June cohort) breaks down and the system can be identified.



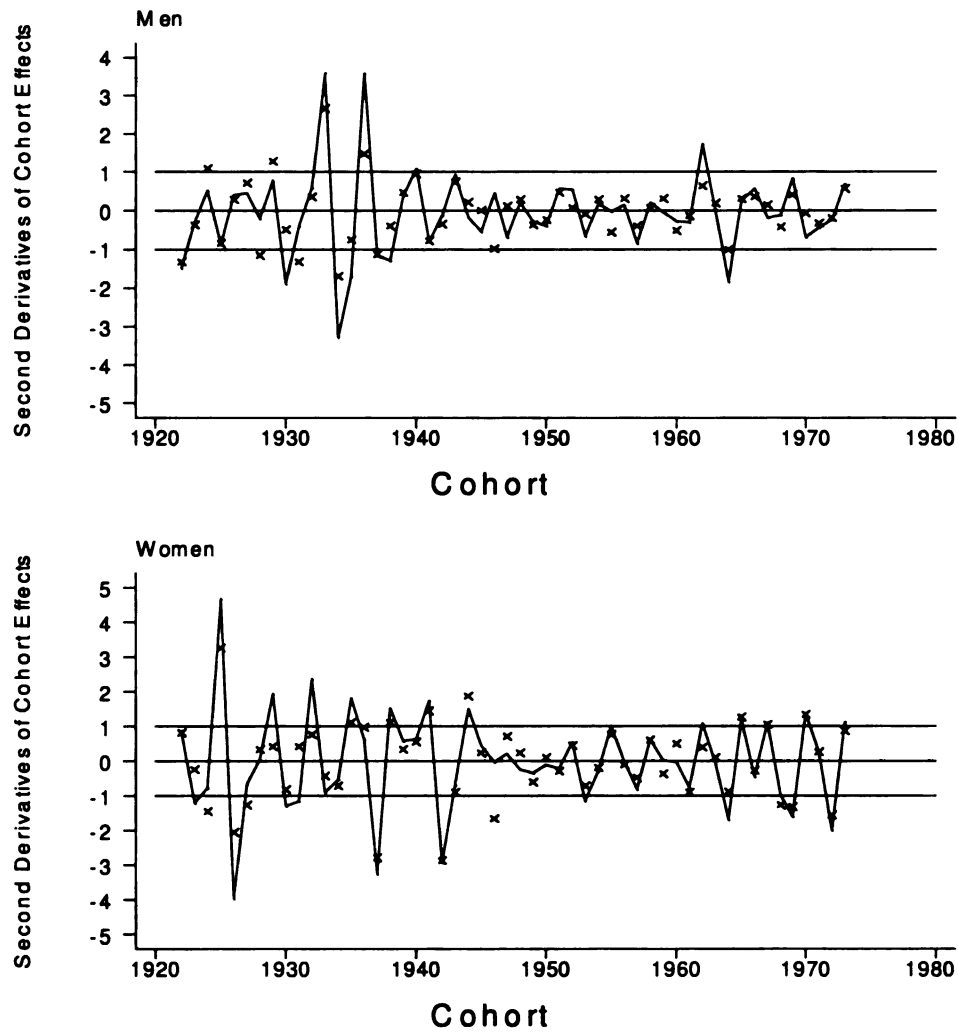
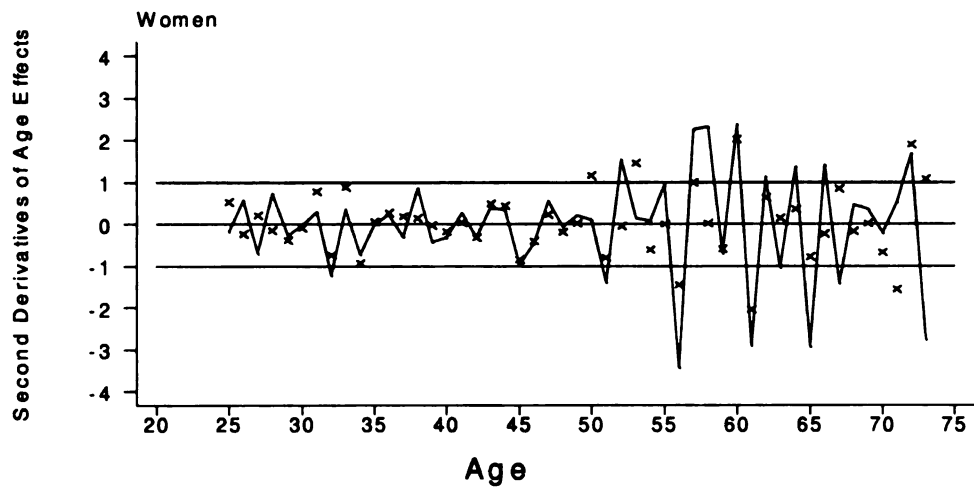
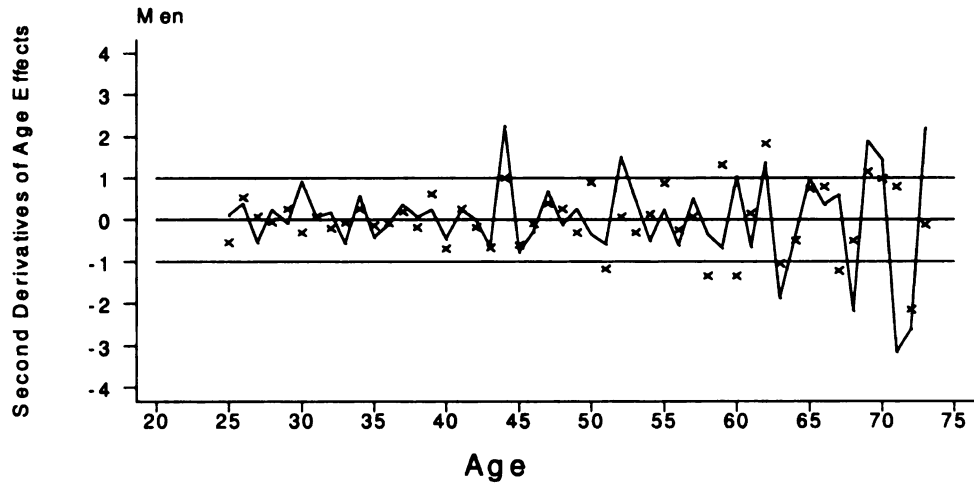


Figure 3.7: The Second Derivatives of Age and Cohort Effects for Men and Women for BMI using two-year spaced CHNS 89, 91 and 93. * The approach for repeated cross-sections based on means. – The approach for genuine panel data based on individual level data.



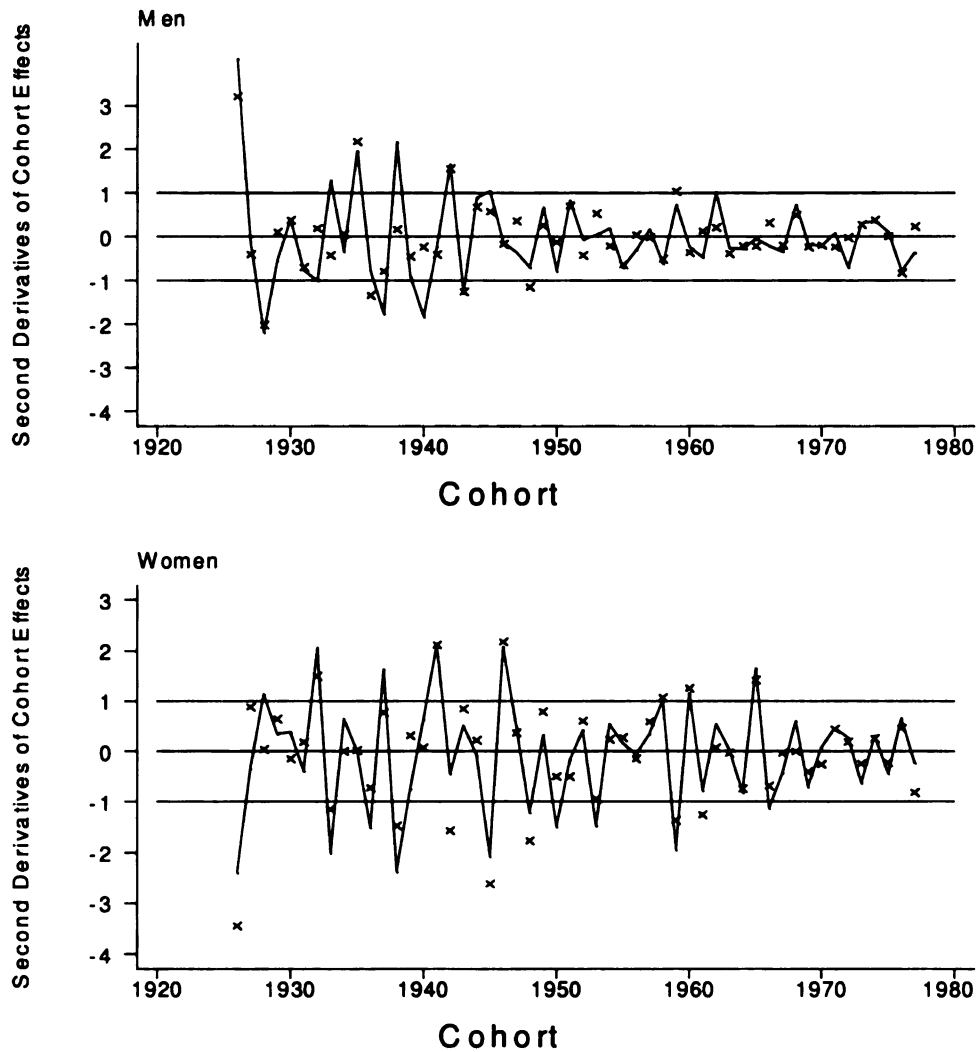


Figure 3.8: The Second Derivatives of Age and Cohort Effects for Men and Women for BMI using four-year spaced CHNS 89, 93 and 97. * The approach for repeated cross-sections based on means. – The approach for genuine panel data based on individual level data.

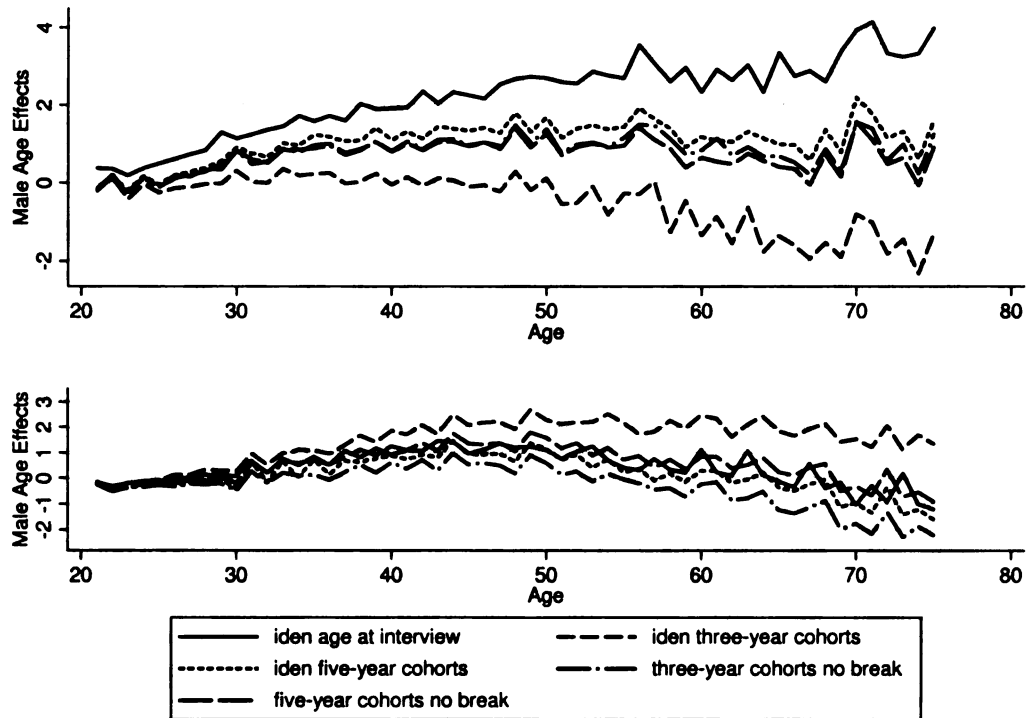


Figure 3.9: Age Effects for Men and Women, identified through (a) exact age at interview (b) three-year cohorts, (c) five-year cohorts specifications, (d) three-year cohorts without breaking down at certain groups into single years, (e) five-year cohorts without breaking down at certain groups into single years.

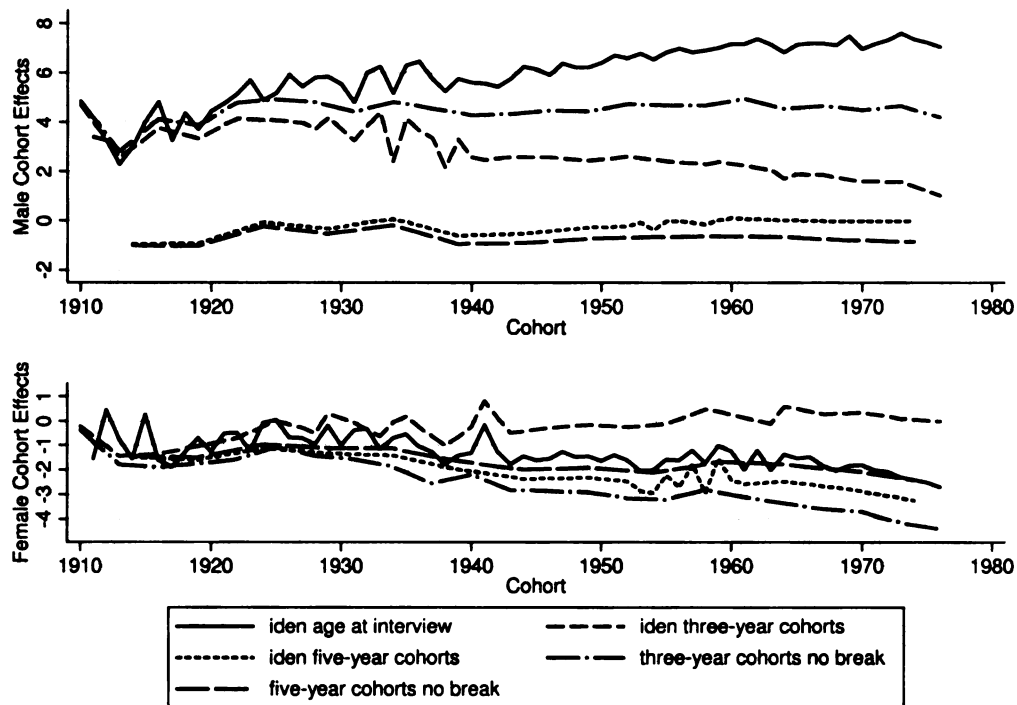


Figure 3.10: Cohort Effects for Men and Women, identified through (a) exact age at interview (b) three-year cohorts, (c) five-year cohorts specifications, (d) three-year cohorts without breaking down at certain groups into single years, (e) five-year cohorts without breaking down at certain groups into single years.

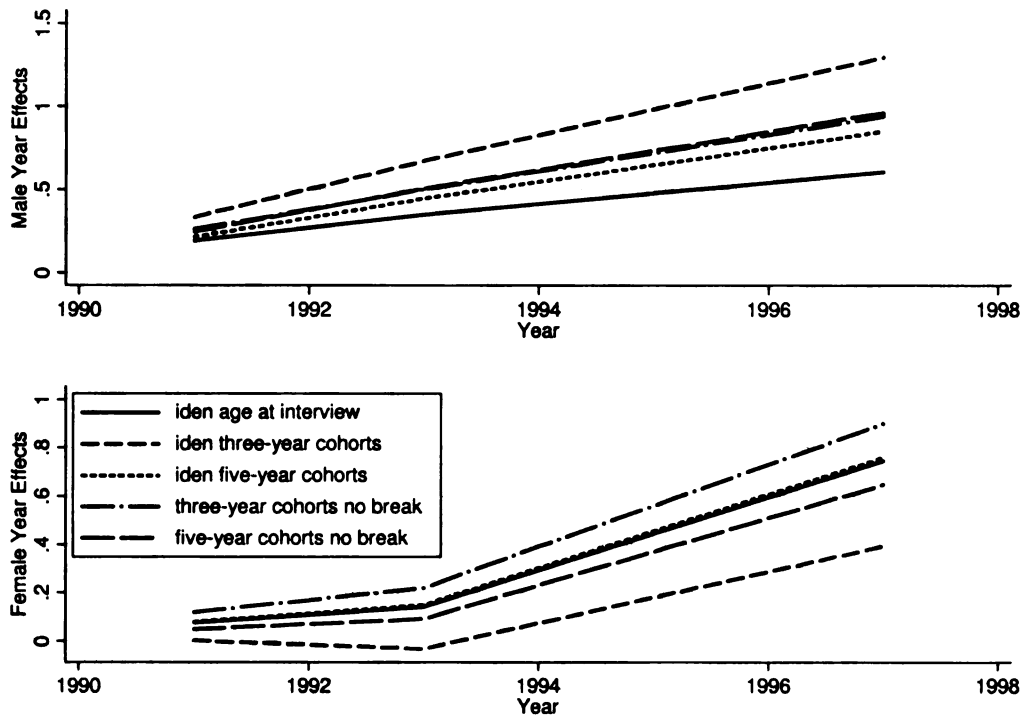


Figure 3.11: Year Effects for Men and Women, identified through (a) exact age at interview (b) three-year cohorts, (c) five-year cohorts specifications, (d) three-year cohorts without breaking down at certain groups into single years, (e) five-year cohorts without breaking down at certain groups into single years.

Figures 3.9, 3.10 and 3.11 are estimated age, period and cohort profiles for men and women with the above five different identifying strategies. The most meaningful profiles are based on the five-year cohorts strategies.

For men, the increasing trend by age (Figure 3.9) is not as clear as that for women. The female BMI-age profile peaks around 50 and then starts to decline. Shrinking is faster for female than for male. The hunch back shaped age profile for women is consistent with what we find in figure 1.

The period effect for men (Figure 3.10) is linear and increasing over time at a larger magnitude than for women. The strong linear year effect for men suggests as we pool the data from all four years of survey in the main analysis for BMI it is

important to control for the year effect.

The cohort profile (Figure 3.11) for men is quite flat, which seems to be at odd with figure 1. However, this may not be due to the failure of the model but because we have a relatively short panel that does not allow us to fully estimate the cohort effect. As we can see from figure 1, especially after we group our data to five-year cohorts, at each age there are only two to three cohort groups available to identify any cohort effect. If we have a longer panel the fitting will be improved.

There is a decreasing trend in female cohort profile and two dips around the revolution and the pre-famine years.

3.6 Concluding Remarks

This APC exercise helps us to better understand the BMI distribution against age, cohort and year for men and women. The age profile for women is inversely U-shaped. The year effect for men is strong. There are not enough data to identify cohort effects due to the short length of the survey, for any given cohort there are only data for a short span of age. In our main analysis for socioeconomic determinants of BMI we will be using the five-year cohort identification strategy.

APPENDIX

Appendix A: Survey instruments in CHNS

The China Health and Nutrition Survey (CHNS) was designed to examine the effects of health, nutrition, and family planning policies and programs implemented by national and local governments and to see how the social and economic transformation of Chinese society is affecting the health and nutritional status of its population. The impact on nutrition and health behaviors and outcomes is gauged by measuring changes in community organizations and programs as well as by measuring changes in sets of household and individual economic, demographic, and social factors (CPC online).

The survey instruments were designed by an interdisciplinary group of social scientists and biomedical researchers with extensive experience in survey research on these topics.

The health services section contains health insurance coverage of surveyed individuals, medical providers, and health facilities that the household might use. Information on illness and uses of the health system during the previous month is asked for some household members. Physical examinations that include weight, height and blood pressures were given to all adults and children starting from 1991 and to adults aged 20 to 45 in 1989. Physical functioning data were collected since 1993 for the elderly.

Table A.1: Survey instruments for health related topics

Instruments	1989	1991	1993	1997
M1a. How would you describe your health compared to that of other people your age? a		✓	✓	✓
M25. How severe was the illness or injury within the last four weeks? b	✓	✓	✓	✓
M26a. How many days unable to carry out normal activities due to this illness?		✓	✓	✓
M29. How many days hospitalized in the first facility?	✓	✓	✓	✓
M35. How many days hospitalized in the second facility?	✓	✓	✓	✓
U56-91, U48-93, U48-97 Past three months, had difficulties in daily activities due to illness?	✓	✓	✓	✓
U57-91, U49-93, U49-97 Weeks with difficulties in daily activities due to illness?	✓	✓	✓	✓
A10, U12-91, U56-93, U56-97 Pregnant? c	✓	✓	✓	✓
Weight (kg) d	✓	✓	✓	✓
Height (cm)	✓	✓	✓	✓
Upper arm circumference (cm)	✓	✓	✓	✓
Triceps skin fold (mm) e	✓	✓	✓	✓
Hip circumference (cm)		✓	✓	✓
Waist circumference (cm)		✓	✓	✓
Systolic blood pressure (mm Hg) f	✓	✓	✓	✓
Diastolic blood pressure (mm Hg)	✓	✓	✓	✓

^aThis field is not available in 1989 survey. In 1997 the same question is asked in physical exam module under U48a. Answers are coded as excellent, good, fair, or poor.

^bAnswers are coded as not severe, somewhat severe, or quite severe.

^cFemale members who are 16 and older were asked about pregnancy in 1991, so were female aged 16 to 49 in 1993 and 1997; whereas in 1989 all female members were asked this question.

^dOnly adults age 20 to 45 in 1989 survey have anthropometric measures.

^eTriceps skin fold was measured three times in 1993-97 surveys.

^fSystolic and diastolic blood pressures were measured three times in 1991-97 surveys, and once in 1989 survey.

Table A.2: ADL type questions in CHNS 93-97

Do you have difficulties in ^a	1993	1997
1. Running a kilometer?	U62	U157
2. Walking a kilometer?	U63	U158
3. Walking 200 meters?	U64	U159
4. Walking across a room?	U65	U160
5. Sitting for 2 hours?	U66	U161
6. Standing up?	U67	U162
7. Climbing one staircase? h	U70	U163
8. Climbing a few stairs?	U69	U164
9. Lifting 5 kilogram?	U71	U165
10. Squatting down, kneel, bend?	U72	U166
11. Taking a shower?	U73	U167
12. Eating?	U76	U169
13. Putting on clothes?	U79	U171
14. Combing your hair?	U80	U172
15. Using the toilet?	U81	U173

^aADL type questions were asked only members over 50 in 1993 and over 55 in 1997, according to the questionnaire headers. In 1993, activity disruption in the past 3 months was excluded. Coding in 1993 and 97 are different. In 1993, answers were coded as No Difficulty, A Little Difficult, Some Difficulty, Very Difficult, or Can't Do. In 1997, answers were coded as No Difficulty, Some Difficulty, Need Help, or Can't Do.

^bU69 and U164 questions are "do you have difficulty in climbing a few stairs without stopping"?". U70 and U163 are "do you have difficulty in climbing one staircase?" In 1993, the skip pattern is when answer to U69 is yes U70 is asked. In 1997, however, the pattern is when answer to U163 is yes U164 is asked. Different skip patterns in two years may make these outcomes incomparable.

Table A.3: Adult BMI by Education Levels in Rural and Urban Areas

	Male			Female		
	All	Urban	Rural	All	Urban	Rural
No Formal Education						
Number of observations	1126	261	865	3779	936	2843
Median BMI	21.25	22.76	20.94	21.64	22.88	21.33
Mean BMI	21.79	23.11	21.39	22.17	23.27	21.81
	(3.40)	(3.49)	(3.27)	(3.92)	(4.41)	(3.67)
% Undernourished (<18.5)	10.04%	3.83%	11.91%	12.60%	11.00%	13.12%
	(0.30)	(0.19)	(0.32)	(0.33)	(0.31)	(0.34)
% Overweight (≥25)	14.21%	28.35%	9.94%	18.26%	30.02%	14.39%
	(0.35)	(0.45)	(0.30)	(0.39)	(0.46)	(0.35)
% Obese (≥30)	1.24%	1.92%	1.04%	2.73%	5.13%	1.93%
	(0.11)	(0.14)	(0.10)	(0.16)	(0.22)	(0.14)
% Overweight high risk	0.36%	0.38%	0.35%	5.48%	11.32%	3.55%
	(0.06)	(0.06)	(0.06)	(0.23)	(0.32)	(0.19)
Some Primary Schooling						
Number of observations	2256	525	1731	2426	563	1863
Median BMI	21.12	22.10	20.94	21.75	22.60	21.50
Mean BMI	21.58	22.63	21.26	22.24	23.04	21.99
	(3.43)	(4.31)	(3.05)	(3.60)	(3.60)	(3.57)
% Undernourished (<18.5)	11.44%	10.67%	11.67%	9.03%	7.99%	9.34%
	(0.32)	(0.31)	(0.32)	(0.29)	(0.27)	(0.29)
% Overweight (≥25)	10.64%	21.90%	7.22%	15.70%	25.58%	12.72%
	(0.31)	(0.41)	(0.26)	(0.36)	(0.44)	(0.33)
% Obese (≥30)	1.55%	2.86%	1.16%	2.31%	4.44%	1.66%
	(0.12)	(0.17)	(0.11)	(0.15)	(0.21)	(0.13)
% Overweight high risk	0.75%	1.90%	0.40%	4.66%	9.41%	3.22%
	(0.09)	(0.14)	(0.06)	(0.21)	(0.29)	(0.18)
Primary School						
Number of observations	2640	567	2073	2355	573	1782
Median BMI	21.26	21.83	21.12	22.00	22.86	21.78
Mean BMI	21.63	22.30	21.45	22.53	23.17	22.33
	(3.03)	(3.59)	(2.84)	(4.65)	(3.76)	(4.89)
% Undernourished (<18.5)	9.09%	11.11%	8.54%	6.88%	6.28%	7.07%
	(0.29)	(0.31)	(0.28)	(0.25)	(0.24)	(0.26)
% Overweight (≥25)	9.77%	19.75%	7.04%	18.17%	25.31%	15.88%
	(0.30)	(0.40)	(0.26)	(0.39)	(0.44)	(0.37)
% Obese (≥30)	1.21%	2.29%	0.92%	2.12%	3.32%	1.74%
	(0.11)	(0.15)	(0.10)	(0.14)	(0.18)	(0.13)
% Overweight high risk	0.34%	1.23%	0.10%	4.16%	7.16%	3.20%
	(0.06)	(0.11)	(0.03)	(0.20)	(0.26)	(0.18)

Table continues

Table A.3 (cont'd)

	Male			Female		
	All	Urban	Rural	All	Urban	Rural
Middle School						
Number of observations	4173	1242	2931	3290	1227	2063
Median BMI	21.23	21.62	21.13	21.48	21.91	21.35
Mean BMI	21.76	22.16	21.59	22.06	22.53	21.78
	(3.20)	(3.58)	(3.00)	(3.63)	(4.10)	(3.29)
% Undernourished (<18.5)	7.36%	9.10%	6.62%	8.69%	7.66%	9.31%
	(0.26)	(0.29)	(0.25)	(0.28)	(0.27)	(0.29)
% Overweight (≥25)	11.10%	16.75%	8.70%	14.71%	20.29%	11.39%
	(0.31)	(0.37)	(0.28)	(0.35)	(0.40)	(0.32)
% Obese (≥30)	1.29%	1.53%	1.19%	1.82%	2.69%	1.31%
	(0.11)	(0.12)	(0.11)	(0.13)	(0.16)	(0.11)
% Overweight high risk	0.38%	0.32%	0.41%	3.10%	4.48%	2.28%
	(0.06)	(0.06)	(0.06)	(0.17)	(0.21)	(0.15)
High School						
Number of observations	1655	761	894	1260	713	547
Median BMI	21.62	21.97	21.28	21.44	21.55	21.24
Mean BMI	22.10	22.54	21.73	21.96	22.01	21.89
	(3.42)	(4.03)	(2.75)	(3.50)	(3.34)	(3.69)
% Undernourished (<18.5)	6.65%	7.88%	5.59%	9.29%	9.40%	9.14%
	(0.25)	(0.27)	(0.23)	(0.29)	(0.29)	(0.29)
% Overweight (≥25)	14.98%	19.58%	11.07%	14.92%	15.85%	13.71%
	(0.36)	(0.40)	(0.31)	(0.36)	(0.37)	(0.34)
% Obese (≥30)	1.27%	1.97%	0.67%	1.19%	1.12%	1.28%
	(0.11)	(0.14)	(0.08)	(0.11)	(0.11)	(0.11)
% Overweight high risk	0.73%	1.18%	0.34%	1.90%	1.82%	2.01%
	(0.08)	(0.11)	(0.06)	(0.14)	(0.13)	(0.14)
Technical School/College						
Number of observations	1015	754	261	680	526	154
Median BMI	22.31	22.29	22.40	21.43	21.44	21.40
Mean BMI	22.67	22.71	22.55	22.01	22.07	21.83
	(3.21)	(3.25)	(3.07)	(3.22)	(3.35)	(2.73)
% Undernourished (<18.5)	6.50%	6.63%	6.13%	8.68%	8.56%	9.09%
	(0.25)	(0.25)	(0.24)	(0.28)	(0.28)	(0.29)
% Overweight (≥25)	20.59%	22.15%	16.09%	15.88%	16.54%	13.64%
	(0.40)	(0.42)	(0.37)	(0.37)	(0.37)	(0.34)
% Obese (≥30)	2.07%	2.39%	1.15%	1.62%	2.09%	0.00%
	(0.14)	(0.15)	(0.11)	(0.13)	(0.14)	(0.00)
% Overweight high risk	1.08%	1.33%	0.38%	3.82%	3.61%	4.55%
	(0.10)	(0.11)	(0.06)	(0.19)	(0.19)	(0.21)

Note: Standard deviations are in parentheses. Having high risk or high abdominal fat is defined as having waist circumference > 102 cm for male or >88 cm for female. Waist circumferences were not measured before 1993.

Table A.4: Community Characteristics in CHNS 1989-97

	1989			1991			1993			1997		
	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural
Population	2566.1	3760.8	1973.9	2248.0	2463.7	2139.3	2746.7	3398.2	2429.0	2239.0	2795.5	1958.5
% in agriculture	40.80	8.53	55.71	47.27	9.22	64.79	42.70	10.07	57.68	42.48	9.58	58.40
% work out of town	18.62	14.79	20.40	17.28	19.07	16.46	19.39	15.77	21.05	21.11	17.32	22.90
Skilled worker wage	9.88	9.92	9.87	4.69	4.38	4.74	6.44	7.17	6.39	16.13	18.62	15.84
Construction worker wage	12.85	20.00	9.61	7.86	7.31	8.12	10.05	10.35	9.89	24.34	24.86	24.09
Price of rice	1.56	1.65	1.52	1.07	1.11	1.05	2.00	3.59	1.23	2.63	2.72	2.58
Price of wheat	1.39	1.39	1.39	1.12	1.06	1.15	2.24	4.19	1.31	2.56	2.55	2.57
Price of eggs	5.99	6.09	5.94	3.93	4.13	3.83	4.51	4.80	4.38	5.82	5.83	5.81
Price of pork	5.75	5.70	5.77	5.88	5.76	5.94	7.08	7.34	6.96	14.46	13.79	14.79
Price of beef	7.58	7.59	7.57	7.50	7.35	7.57	9.24	9.68	9.02	15.05	13.89	15.62
Price of fish	7.71	7.72	7.71	6.98	7.13	6.90	8.36	8.53	8.27	11.77	9.91	12.68
Clinic physicians /100 res	0.05	0.04	0.06	0.11	0.17	0.08	0.09	0.19	0.04			
Clinic nurses /100 res	0.01	0.02	0.00	0.05	0.14	0.00	0.00	0.01	0.00			
Clinic beds /100 res	0.03	0.06	0.01	0.20	0.52	0.04	0.10	0.18	0.06			
Clinic health workers /100 res	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01			
% Water underground ^a	0.40	0.25	0.48	0.38	0.19	0.48	0.39	0.13	0.53	0.41	0.14	0.54
% Water open well	0.18	0.03	0.26	0.14	0.01	0.21	0.12	0.03	0.16	0.07	0.02	0.10
% Water spring, river, lake, rain	0.08	0.04	0.10	0.07	0.02	0.09	0.07	0.04	0.09	0.03	0.01	0.05
% Water factory	0.34	0.69	0.16	0.41	0.78	0.22	0.42	0.81	0.23	0.49	0.84	0.32
% Toilet inhouse-flush	0.12	0.28	0.04	0.14	0.31	0.05	0.18	0.38	0.09	0.28	0.54	0.15
% Toilet inhouse-no-flush	0.06	0.06	0.06	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.03	0.04
% Toilet outside	0.17	0.35	0.08	0.24	0.41	0.16	0.20	0.33	0.13	0.14	0.24	0.08
% Toilet open pit	0.56	0.21	0.75	0.53	0.22	0.69	0.55	0.21	0.72	0.52	0.16	0.70
% No toilet	0.04	0.07	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.02	0.02	0.02
% Other toilet	0.04	0.03	0.05	0.04	0.01	0.06	0.01	0.00	0.01	0.01	0.00	0.02
% No excreta	0.55	0.84	0.40	0.53	0.82	0.38	0.56	0.87	0.41	0.64	0.91	0.51
% Very little excreta	0.23	0.10	0.30	0.25	0.12	0.32	0.26	0.09	0.34	0.22	0.06	0.29
% Some excreta	0.20	0.05	0.28	0.21	0.05	0.29	0.17	0.04	0.23	0.13	0.03	0.18
% Much excreta	0.02	0.00	0.03	0.01	0.00	0.02	0.02	0.00	0.02	0.01	0.00	0.01

Note: In 1997 the hospital and clinic questions were not included. The %s are aggregated from household level. For example, in 1989 within a community the average percent of households having underground water source is 40.2 percent.

Appendix B: First Stage

Regressions for Section 1.6

These are the first stage regressions in Table 1.13 where BMI in 1991 is instrumented with information from 1989.

Table B.1: First stage regressions for Table 1.13

Variables	Male		Female	
Some primary 91	0.030 (0.237)	-0.003 (0.238)	0.295 (0.229)	0.295 (0.230)
Primary 91	-0.029 (0.239)	-0.056 (0.240)	0.610 (0.247)**	0.610 (0.247)**
Middle school 91	0.109 (0.245)	0.072 (0.246)	0.091 (0.263)	0.095 (0.264)
High school 91	0.345 (0.289)	0.310 (0.290)	0.046 (0.345)	0.050 (0.345)
Tech/College+ 91	0.022 (0.337)	-0.008 (0.338)	-0.969 (0.470)**	-0.955 (0.473)**
Log real prod asset 91 1 ^a	-0.053 (0.052)	-0.056 (0.053)	0.059 (0.060)	0.059 (0.060)
Log real prod asset 91 2 ^a	0.079 (0.052)	0.079 (0.052)	-0.101 (0.060)*	-0.101 (0.060)*
Land farmed in 90	0.046 (0.033)	0.044 (0.033)	-0.007 (0.040)	-0.007 (0.040)
Married 91		0.329 (0.267)		0.081 (0.497)
Divorced Separated 91		-0.304 (0.874)		0.309 (0.988)
Widowed 91		0.022 (0.477)		0.130 (0.585)
Log R Productive Asset spline 1 in 89	-0.122 (1.240)	-0.131 (1.240)	-2.033 (1.508)	-2.035 (1.510)
Log R Productive Asset spline 2 in 89	0.461 (1.189)	0.474 (1.190)	0.714 (1.421)	0.717 (1.423)
Land farmed in 88	-1.106 (0.835)	-1.050 (0.838)	-0.678 (1.296)	-0.680 (1.296)

Table continues

Table B.1 (cont'd)

	Male		Female	
Price Rice*Log R Prod s1 89	-0.017 (0.237)	-0.011 (0.237)	0.064 (0.262)	0.066 (0.262)
Price Rice*Log R Prod s2 89	0.005 (0.242)	0.004 (0.242)	0.201 (0.259)	0.199 (0.259)
Price Rice*Land 88	0.095 (0.098)	0.092 (0.098)	-0.058 (0.115)	-0.057 (0.115)
Price Egg*Log R Prod s1 89	-0.003 (0.043)	-0.004 (0.043)	-0.049 (0.052)	-0.049 (0.053)
Price Egg*Log R Prod s2 89	-0.042 (0.046)	-0.041 (0.046)	0.041 (0.056)	0.041 (0.056)
Price Egg*Land 88	-0.007 (0.023)	-0.006 (0.023)	0.007 (0.030)	0.007 (0.030)
Price Pork*Log R Prod s1 89	-0.050 (0.067)	-0.050 (0.067)	0.111 (0.079)	0.111 (0.080)
Price Pork*Log R Prod s2 89	0.074 (0.075)	0.073 (0.075)	-0.102 (0.086)	-0.103 (0.086)
Price Pork*Land 88	0.013 (0.046)	0.011 (0.046)	0.028 (0.054)	0.028 (0.054)
Price Fish*Log R Prod s1 89	0.032 (0.023)	0.032 (0.023)	-0.001 (0.026)	-0.001 (0.026)
Price Fish*Log R Prod s2 89	-0.027 (0.024)	-0.027 (0.024)	0.019 (0.028)	0.019 (0.028)
Price Fish*Land 88	0.016 (0.014)	0.016 (0.014)	0.005 (0.013)	0.005 (0.013)
Clinic Physician*Log R Prod s1 89	-54.763 (40.758)	-55.600 (40.768)	-73.328 (49.170)	-73.211 (49.215)
Clinic Physician*Log R Prod s2 89	58.932 (40.179)	59.749 (40.191)	56.078 (49.316)	56.089 (49.374)
Clinic Physician*Land 88	6.544 (43.261)	10.576 (43.361)	-12.828 (51.190)	-13.092 (51.243)
Clinic Nurse*Log R Prod s1 89	91.537 (214.023)	87.161 (214.068)	342.405 (262.008)	343.472 (262.194)
Clinic Nurse*Log R Prod s2 89	-255.278 (236.251)	-247.959 (236.321)	-296.318 (292.011)	-296.527 (292.189)
Clinic Nurse*Land 88	1277.351 (3070.487)	1470.020 (3074.360)	-488.022 (644.643)	-489.179 (645.153)
Clinic Bed*Log R Prod s1 89	-109.046 (107.524)	-106.985 (107.548)	-4.284 (130.627)	-4.665 (130.712)
Clinic Bed*Log R Prod s2 89	176.883 (118.294)	172.719 (118.347)	12.951 (146.623)	12.986 (146.718)
Clinic Bed*Land 88	8.459 (86.225)	1.093 (86.399)	24.826 (102.190)	25.516 (102.311)

Table continues

Table B.1 (cont'd)

	Male		Female	
Water Source 2*Log R Prod s1 89	-0.354 (0.302)	-0.354 (0.303)	-0.430 (0.323)	-0.427 (0.323)
Water Source 2*Log R Prod s2 89	0.168 (0.286)	0.163 (0.286)	0.213 (0.308)	0.211 (0.309)
Water Source 2*Land 88	0.095 (0.093)	0.094 (0.093)	-0.103 (0.097)	-0.103 (0.097)
Water Source 3*Log R Prod s1 89	0.113 (0.377)	0.109 (0.377)	-0.200 (0.450)	-0.202 (0.451)
Water Source 3*Log R Prod s2 89	0.021 (0.367)	0.021 (0.367)	0.291 (0.454)	0.293 (0.455)
Water Source 3*Land 88	0.120 (0.211)	0.115 (0.211)	0.088 (0.197)	0.085 (0.198)
Water Source 4*Log R Prod s1 89	0.064 (0.204)	0.070 (0.204)	0.219 (0.242)	0.219 (0.242)
Water Source 4*Log R Prod s2 89	-0.014 (0.220)	-0.022 (0.220)	-0.269 (0.249)	-0.270 (0.250)
Water Source 4*Land 88	-0.348 (0.162)**	-0.360 (0.162)**	-0.117 (0.212)	-0.116 (0.213)
In house no flush*Log R Prod s1 89	0.845 (1.040)	0.857 (1.041)	1.436 (1.269)	1.434 (1.270)
In house no flush*Log R Prod s2 89	-0.805 (1.008)	-0.824 (1.009)	-0.042 (1.227)	-0.037 (1.228)
In house no flush*Land 88	0.875 (0.802)	0.822 (0.804)	0.294 (1.252)	0.295 (1.254)
Outside toilets*Log R Prod s1 89	0.271 (0.955)	0.260 (0.956)	1.438 (1.192)	1.437 (1.193)
Outside toilets*Log R Prod s2 89	-0.449 (0.922)	-0.441 (0.922)	-0.359 (1.141)	-0.357 (1.142)
Outside toilets*Land 88	0.695 (0.808)	0.642 (0.810)	0.582 (1.254)	0.583 (1.255)
Open pit cement or earth*Log R Prod s1 89	0.389 (0.899)	0.398 (0.899)	1.665 (1.114)	1.661 (1.115)
Open pit cement or earth*Log R Prod s2 89	-0.614 (0.869)	-0.627 (0.870)	-0.717 (1.065)	-0.713 (1.066)
Open pit cement or earth*Land 88	0.706 (0.783)	0.663 (0.785)	0.459 (1.250)	0.460 (1.251)
No toilets*Log R Prod s1 89	1 (1.477)	1.043 (1.478)	3.477 (1.831)*	3.482 (1.833)*
No toilets*Log R Prod s2 89	-1.440 (1.423)	-1.459 (1.424)	-2.415 (1.792)	-2.416 (1.793)

Table continues

Table B.1 (cont'd)

	Male		Female	
No toilets*Land 88	1.283 (1)	1.276 (1.001)	0.106 (1.322)	0.107 (1.322)
Other toilets*Log R Prod s1 89	-0.152 (1.288)	-0.158 (1.289)	-1.712 (1.522)	-1.719 (1.523)
Other toilets*Log R Prod s2 89	0.050 (1.229)	0.045 (1.229)	2.427 (1.465)*	2.430 (1.466)*
Other toilets*Land 88	0.511 (0.848)	0.465 (0.849)	0.696 (1.311)	0.699 (1.312)
Little excreta*Log R Prod s1 89	-0.091 (0.283)	-0.107 (0.283)	0.102 (0.341)	0.103 (0.341)
Little excreta*Log R Prod s2 89	0.051 (0.305)	0.062 (0.305)	-0.213 (0.361)	-0.215 (0.361)
Little excreta*Land 88	0.046 (0.158)	0.045 (0.158)	0.209 (0.187)	0.208 (0.187)
Some excreta*Log R Prod s1 89	-0.214 (0.315)	-0.207 (0.316)	0.093 (0.419)	0.090 (0.419)
Some excreta*Log R Prod s2 89	0.244 (0.320)	0.236 (0.320)	0.098 (0.430)	0.100 (0.431)
Some excreta*Land 88	0.102 (0.129)	0.098 (0.129)	0.238 (0.154)	0.238 (0.154)
No excreta*Log R Prod s1 89	0.381 (1.287)	0.366 (1.288)	2.746 (1.664)*	2.758 (1.665)*
No excreta*Log R Prod s2 89	-0.716 (1.313)	-0.699 (1.313)	-2.958 (1.592)*	-2.969 (1.593)*
No excreta*Land 88	0.175 (0.380)	0.180 (0.381)	-0.689 (0.482)	-0.690 (0.482)
Number of obs	2477	2477	2701	2701
R-squared	0.2578	0.2587	0.2379	0.2379

The percent of households with the following water sources: (1) underground, (2) open well, (3) spring, river lake, rain or snow, and (4) water factory. Omitted is the first source. The inhouse with flush toilet type is omitted. The no excreta category is omitted.

These are the first stage regressions in Table 1.14 where 1993 BMI was instrumented with information in 1991.

Table B.2: First stage regressions for Table 1.14

Var	Male		Female	
Some primary 93	0.252 (0.322)	0.246 (0.322)	0.523 (0.253)**	0.533 (0.253)**
Primary 93	-0.303 (0.359)	-0.318 (0.360)	0.449 (0.326)	0.460 (0.327)
Middle school 93	-0.194 (0.353)	-0.215 (0.354)	0.007 (0.316)	0.026 (0.317)
High school 93	0.118 (0.419)	0.113 (0.420)	-0.474 (0.431)	-0.465 (0.431)
Tech/College+ 93	0.220 (0.530)	0.179 (0.530)	-0.254 (0.665)	-0.227 (0.666)
Log real prod asset 93 1 ^a	-0.038 (0.059)	-0.041 (0.059)	0.090 (0.058)	0.096 (0.058)
Log real prod asset 93 2 ^a	0.072 (0.056)	0.073 (0.056)	0.007 (0.054)	0.001 (0.055)
Land farmed in 92	0.027 (0.021)	0.027 (0.021)	-0.015 (0.023)	-0.014 (0.024)
Married 93		0.077 (0.428)		0.332 (0.701)
Divorced Separated 93		-1.973 (1.328)		-1.110 (1.530)
Widowed 93		-0.636 (0.667)		0.647 (0.798)
Log real prod asset 91 1 ^a	0.959 (1.045)	0.925 (1.045)	-1.620 (0.990)	-1.589 (0.991)
Log real prod asset 91 2 ^a	-1.107 (0.946)	-1.062 (0.946)	0.586 (0.921)	0.568 (0.922)
Land farmed in 90	-1.383 (2.304)	-1.394 (2.303)	-2.134 (2.560)	-2.266 (2.562)

Table continues

Table B.2 (cont'd)

	Male		Female	
Price Rice*Log R Prod s1 91	0.109 (0.249)	0.109 (0.249)	0.086 (0.224)	0.079 (0.224)
Price Rice*Log R Prod s2 91	-0.305 (0.301)	-0.310 (0.300)	-0.017 (0.286)	-0.017 (0.286)
Price Rice*Land 90	-0.426 (0.313)	-0.449 (0.313)	0.353 (0.335)	0.346 (0.335)
Price Egg*Log R Prod s1 91	-0.038 (0.037)	-0.037 (0.037)	0.041 (0.035)	0.040 (0.035)
Price Egg*Log R Prod s2 91	0.064 (0.037)*	0.063 (0.037)*	-0.020 (0.039)	-0.019 (0.039)
Price Egg*Land 90	0.020 (0.030)	0.022 (0.030)	-0.029 (0.032)	-0.027 (0.032)
Price Pork*Log R Prod s1 91	0.007 (0.013)	0.008 (0.013)	0.014 (0.012)	0.014 (0.012)
Price Pork*Log R Prod s2 91	-0.010 (0.012)	-0.011 (0.012)	-0.010 (0.012)	-0.010 (0.012)
Price Pork*Land 90	-0.005 (0.011)	-0.005 (0.011)	0.007 (0.010)	0.007 (0.010)
Price Fish*Log R Prod s1 91	-0.007 (0.060)	-0.005 (0.060)	0.104 (0.056)*	0.100 (0.056)*
Price Fish*Log R Prod s2 91	0.036 (0.056)	0.032 (0.057)	-0.031 (0.054)	-0.028 (0.054)
Price Fish*Land 90	-0.054 (0.038)	-0.055 (0.038)	-0.006 (0.039)	-0.005 (0.039)
Clinic Physician*Log R Prod s1 91	106.470 (64.158)*	104.397 (64.177)	-31.799 (56.539)	-32.189 (56.568)
Clinic Physician*Log R Prod s2 91	-64.291 (59.262)	-63.582 (59.269)	1.969 (53.180)	2.593 (53.200)
Clinic Physician*Land 90	30.506 (45.891)	25.657 (45.968)	-4.420 (51.507)	-4.605 (51.565)
Clinic Nurse*Log R Prod s1 91	-5205.468 (1269.822)**	-5241.722 (1269.582)**	-971.784 (606.475)	-1003.204 (608.499)*
Clinic Nurse*Log R Prod s2 91	4192.287 (1064.981)**	4219.063 (1064.810)**	668.101 (533.770)	691.569 (535.244)
Clinic Nurse*Land 90	0.344 (1234.162)	0.875 (1233.638)	-1071.045 (913.902)	-1066.944 (914.324)
Clinic Bed*Log R Prod s1 91	-192.314 (258.593)	-178.749 (258.717)	40.569 (164.804)	34.565 (164.896)
Clinic Bed*Log R Prod s2 91	174.245 (225.482)	164.440 (225.569)	-11.620 (145.094)	-6.370 (145.175)
Clinic Bed*Land 90	71.949 (64.611)	71.989 (64.636)	62.266 (62.866)	60.736 (62.899)

Table continues

Table B.2 (cont'd)

	Male		Female	
Water Source 2*Log R Prod s1 91	0.091 (0.355)	0.090 (0.355)	-0.429 (0.325)	-0.443 (0.325)
Water Source 2*Log R Prod s2 91	0.028 (0.354)	0.025 (0.354)	0.062 (0.307)	0.077 (0.307)
Water Source 2*Land 90	-0.084 (0.162)	-0.080 (0.162)	0.093 (0.177)	0.094 (0.177)
Water Source 3*Log R Prod s1 91	0.239 (0.779)	0.258 (0.779)	0.296 (0.592)	0.310 (0.592)
Water Source 3*Log R Prod s2 91	-0.192 (0.665)	-0.209 (0.665)	-0.455 (0.535)	-0.474 (0.535)
Water Source 3*Land 90	-0.594 (0.279)**	-0.582 (0.279)**	0.262 (0.275)	0.276 (0.277)
Water Source 4*Log R Prod s1 91	-0.451 (0.318)	-0.446 (0.318)	-0.148 (0.300)	-0.156 (0.300)
Water Source 4*Log R Prod s2 91	0.585 (0.302)*	0.588 (0.302)*	-0.057 (0.281)	-0.052 (0.281)
Water Source 4*Land 90	0.031 (0.159)	0.033 (0.159)	-0.072 (0.170)	-0.068 (0.170)
In house no flush*Log R Prod s1 91	-1.462 (1.017)	-1.435 (1.018)	1.475 (0.969)	1.481 (0.970)
In house no flush*Log R Prod s2 91	1.210 (1.281)	1.185 (1.282)	-0.705 (0.931)	-0.707 (0.932)
In house no flush*Land 90	1.875 (2.421)	1.942 (2.420)	1.557 (2.670)	1.705 (2.673)
Outside toilets*Log R Prod s1 91	-1.069 (0.725)	-1.058 (0.724)	0.812 (0.721)	0.827 (0.721)
Outside toiletsI*Log R Prod s2 91	1.020 (0.676)	1.012 (0.676)	-0.236 (0.682)	-0.253 (0.683)
Outside toilets*Land 90	2.109 (2.265)	2.150 (2.264)	2.176 (2.523)	2.309 (2.526)
Open pit cement or earth*Log R Prod s1 91	-1.201 (0.806)	-1.192 (0.806)	0.721 (0.771)	0.727 (0.772)
Open pit cement or earth*Log R Prod s2 91	1.263 (0.742)*	1.250 (0.742)*	-0.225 (0.714)	-0.233 (0.714)
Open pit cement or earth*Land 90	2.122 (2.295)	2.172 (2.295)	1.802 (2.556)	1.932 (2.558)
No toilets*Log R Prod s1 91	-0.474 (1.605)	-0.455 (1.608)	0.230 (1.281)	0.324 (1.288)
No toilets*Log R Prod s2 91	-0.356 (1.643)	-0.379 (1.648)	0.182 (1.361)	0.124 (1.371)
No toilets*Land 90	4.220 (2.657)	4.262 (2.656)	0.041 (2.930)	0.117 (2.934)

Table continues

Table B.2 (cont'd)

	Male		Female	
Other toilets*Log R Prod s1 91	-2.283 (1.425)	-2.344 (1.425)*	0.521 (1.274)	0.512 (1.275)
Other toilets*Log R Prod s2 91	2.056 (1.372)	2.105 (1.374)	0.454 (1.246)	0.458 (1.247)
Other toilets*Land 90	2.439 (2.421)	2.472 (2.420)	2.494 (2.675)	2.636 (2.677)
Little excreta*Log R Prod s1 91	0.829 (0.455)*	0.826 (0.455)*	0.179 (0.411)	0.178 (0.411)
Little excreta*Log R Prod s2 91	-0.825 (0.405)**	-0.808 (0.406)**	-0.061 (0.366)	-0.063 (0.366)
Little excreta*Land 90	-0.063 (0.210)	-0.072 (0.210)	-0.193 (0.206)	-0.203 (0.206)
Some excreta*Log R Prod s1 91	0.191 (0.362)	0.193 (0.362)	-0.449 (0.364)	-0.445 (0.365)
Some excreta*Log R Prod s2 91	-0.233 (0.368)	-0.230 (0.368)	0.422 (0.358)	0.416 (0.358)
Some excreta*Land 90	0.158 (0.243)	0.117 (0.244)	0.045 (0.235)	0.046 (0.236)
No excreta*Log R Prod s1 91	-5.475 (3.943)	-5.386 (3.944)	2.813 (3.167)	2.791 (3.168)
No excreta*Log R Prod s2 91	3.331 (3.297)	3.272 (3.298)	-2.423 (2.811)	-2.431 (2.812)
No excreta*Land 90	0.324 (1.681)	0.352 (1.682)	0.747 (1.695)	0.902 (1.701)
Number of obs	1815	1815	1948	1948
R-squared	0.2585	0.2606	0.2462	0.2472

The percent of households with the following water sources: (1) underground, (2) open well, (3) spring, river lake, rain or snow, and (4) water factory. Omitted is the first source. The inhouse with flush toilet type is omitted. The no excreta category is omitted. For first stage regressions for Table 1.15 and Table 1.16 please contact me

Appendix C: Linear Probability

Models for Undernourishment and Overweight

Table C.1: Probability Model for Overweight: Overall, Urban and Rural Areas

Variables	Overall		Male		Rural	
			Urban			
Some primary education	-0.003 (0.014)	-0.006 (0.014)	-0.027 (0.040)	-0.032 (0.039)	-0.0004 (0.014)	-0.001 (0.014)
Primary	-0.008 (0.015)	-0.012 (0.015)	-0.035 (0.040)	-0.040 (0.040)	-0.009 (0.015)	-0.010 (0.015)
Middle school	0.001 (0.015)	-0.003 (0.015)	-0.050 (0.040)	-0.055 (0.040)	0.006 (0.015)	0.004 (0.015)
High school	0.020 (0.018)	0.016 (0.018)	-0.012 (0.042)	-0.017 (0.042)	0.022 (0.020)	0.020 (0.020)
Tech/College+	-0.006 (0.023)	-0.009 (0.023)	-0.042 (0.044)	-0.047 (0.044)	-0.005 (0.030)	-0.007 (0.030)
Log real prod assets 1 ^a	0 (0.002)	-0.001 (0.002)	0.004 (0.005)	0.003 (0.005)	-0.002 (0.002)	-0.002 (0.002)
Log real prod assets 2 ^a	0.005 (0.002)**	0.005 (0.002)**	0.003 (0.005)	0.003 (0.005)	0.005 (0.002)**	0.005 (0.002)**
Land	-0.001 (0.001)**	-0.001 (0.001)**	-0.006 (0.007)	-0.006 (0.007)	-0.001 (0.001)*	-0.001 (0.001)*
Married		0.030 (0.011)**		0.056 (0.024)**		0.006 (0.012)
Divorced Separated		-0.018 (0.034)		0.015 (0.073)		-0.061 (0.029)**
Widowed		-0.017 (0.022)		-0.046 (0.054)		-0.021 (0.022)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	12760	12760	4087	4087	8673	8673
R-squared	0.179	0.180	0.192	0.194	0.151	0.151
P-value for testing coefficients equal to zero						
Education	0.3828	0.3673	0.4839	0.4212	0.4059	0.4199
Assets	0.0196	0.0213	0.1717	0.1849	0.0416	0.0404
Age dummies	0.8750	0.8563	0.9648	0.9601	0.5659	0.5596
Cohort dummies	0.5106	0.5191	0.2658	0.2694	0.6820	0.6824
Marital Status		0.0019		0.0127		0.0567
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

Table C.1 (cont'd)

Variables	Overall		Female		Rural	
			Urban			
Some primary education	0.006 (0.013)	0.005 (0.013)	-0.016 (0.030)	-0.021 (0.030)	0.002 (0.014)	0.002 (0.014)
Primary	0.022 (0.015)	0.021 (0.015)	-0.030 (0.032)	-0.037 (0.032)	0.025 (0.017)	0.024 (0.017)
Middle school	-0.001 (0.016)	-0.001 (0.016)	-0.007 (0.033)	-0.014 (0.034)	-0.008 (0.017)	-0.008 (0.017)
High school	-0.025 (0.020)	-0.024 (0.020)	-0.030 (0.038)	-0.037 (0.038)	-0.010 (0.024)	-0.009 (0.024)
Tech/College+	-0.051 (0.025)**	-0.051 (0.026)**	-0.036 (0.039)	-0.043 (0.040)	-0.092 (0.044)**	-0.089 (0.044)**
Log real prod assets 1 ^a	0.002 (0.002)	0.002 (0.002)	0.005 (0.005)	0.005 (0.005)	0.001 (0.003)	0.001 (0.003)
Log real prod assets 2 ^a	-0.001 (0.002)	-0.001 (0.002)	-0.003 (0.005)	-0.003 (0.005)	0.001 (0.003)	0.001 (0.003)
Land	0.001 (0.001)	0.001 (0.001)	-0.008 (0.008)	-0.008 (0.008)	0 (0.001)	0 (0.001)
Married		0.018 (0.014)		-0.012 (0.027)		0.018 (0.016)
Divorced Separated		0.026 (0.051)		0.001 (0.069)		0.004 (0.070)
Widowed		-0.012 (0.023)		-0.079 (0.041)*		0.012 (0.027)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	13652	13652	4493	4493	9159	9159
R-squared	0.167	0.167	0.191	0.192	0.152	0.152
P-value for testing coefficients equal to zero						
Education	0.0190	0.0284	0.7922	0.7249	0.0423	0.0530
Assets	0.4221	0.4321	0.6578	0.6616	0.5516	0.5652
Age dummies	0.0817	0.0860	0.1730	0.1943	0.5504	0.5268
Cohort dummies	0.0148	0.0154	0.0744	0.0714	0.1238	0.1306
Marital Status		0.2056		0.1902		0.7229
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table C.2: Probability Model of Overweight: By Age Groups

Variables	Male					
	Age 20-39		Age 40-59		Age 60+	
Some primary education	0.001 (0.026)	-0.004 (0.026)	-0.024 (0.023)	-0.027 (0.023)	-0.022 (0.034)	-0.022 (0.034)
Primary	-0.012 (0.026)	-0.018 (0.026)	-0.024 (0.024)	-0.027 (0.024)	-0.015 (0.040)	-0.015 (0.040)
Middle school	-0.007 (0.025)	-0.013 (0.025)	-0.015 (0.026)	-0.019 (0.026)	-0.084 (0.047)*	-0.084 (0.047)*
High school	0.014 (0.028)	0.008 (0.028)	0.006 (0.035)	0.003 (0.035)	-0.014 (0.071)	-0.011 (0.071)
Tech/College+	-0.011 (0.032)	-0.016 (0.032)	0.008 (0.042)	0.002 (0.042)	-0.133 (0.066)**	-0.134 (0.066)**
Log real prod assets 1 ^a	-0.001 (0.002)	-0.001 (0.002)	-0.003 (0.004)	-0.003 (0.004)	0.008 (0.009)	0.008 (0.009)
Log real prod assets 2 ^a	0.006 (0.003)**	0.006 (0.003)**	0.010 (0.004)**	0.010 (0.004)**	-0.014 (0.009)	-0.013 (0.009)
Land	-0.001 (0.001)	-0.001 (0.001)	-0.003 (0.002)**	-0.003 (0.002)**	-0.002 (0.002)	-0.002 (0.002)
Married		0.029 (0.012)**		0.042 (0.041)		-0.041 (0.071)
Divorced Separated		-0.002 (0.044)		0.028 (0.093)		-0.142 (0.103)
Widowed		-0.031 (0.026)		-0.017 (0.048)		-0.054 (0.077)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	6678	6678	4292	4292	1790	1790
R-squared	0.197	0.198	0.291	0.292	0.359	0.359
P-value for testing coefficients equal to zero						
Education	0.5130	0.4869	0.7246	0.6880	0.3056	0.2993
Assets	0.0221	0.0271	0.0115	0.0118	0.2615	0.2931
Age dummies	0.5810	0.3719	0.7032	0.6745	0.6329	0.6341
Cohort dummies	0.2476	0.2674	0.2590	0.2754	0.5621	0.5740
Marital Status		0.0050		0.1691		0.5386
Community dummies	0.9595	0.9697	0.0000	0.0000	0.0000	0.0000

Table continues

Table C.2 (cont'd)

Variables	Age 20-39		Female Age 40-59		Age 60+	
Some primary education	-0.016 (0.019)	-0.016 (0.019)	0.010 (0.020)	0.009 (0.020)	-0.013 (0.041)	-0.014 (0.041)
Primary	-0.016 (0.021)	-0.016 (0.021)	0.029 (0.025)	0.028 (0.025)	-0.099 (0.053)*	-0.104 (0.054)*
Middle school	-0.029 (0.020)	-0.029 (0.020)	0.006 (0.030)	0.005 (0.030)	-0.110 (0.083)	-0.116 (0.083)
High school	-0.064 (0.024)**	-0.063 (0.024)**	0.099 (0.049)**	0.099 (0.049)**	0.177 (0.179)	0.183 (0.178)
Tech/College+	-0.074 (0.030)**	-0.072 (0.030)**	0.007 (0.055)	0.006 (0.055)	0.047 (0.138)	0.039 (0.138)
Log real prod assets 1 ^a	0 (0.003)	0 (0.003)	0 (0.005)	0 (0.005)	0.015 (0.008)*	0.015 (0.008)*
Log real prod assets 2 ^a	-0.002 (0.003)	-0.002 (0.003)	0.008 (0.005)*	0.008 (0.005)*	-0.015 (0.008)*	-0.014 (0.008)*
Land	0.003 (0.001)**	0.003 (0.001)**	-0.003 (0.002)	-0.003 (0.002)	0.002 (0.003)	0.002 (0.003)
Married		0.017 (0.014)		0.022 (0.058)		-0.037 (0.088)
Divorced Separated		-0.037 (0.076)		-0.028 (0.100)		0.015 (0.143)
Widowed		0.156 (0.099)		-0.014 (0.065)		-0.064 (0.089)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	7177	7177	4578	4578	1897	1897
R-squared	0.180	0.181	0.284	0.284	0.356	0.356
P-value for testing coefficients equal to zero						
Education	0.0272	0.0413	0.3465	0.3435	0.2562	0.2189
Assets	0.6709	0.6564	0.0499	0.0573	0.1341	0.1538
Age dummies	0.0512	0.0442	0.5502	0.5524	0.6055	0.6125
Cohort dummies	0.0092	0.0088	0.5018	0.4928	0.1057	0.0972
Marital Status		0.2459		0.5738		0.6732
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table C.3: Probability Model of Undernourishment: Overall, Urban and Rural Areas

Variables	Overall		Male		Rural	
			Urban			
Some primary education	0.034 (0.014)**	0.036 (0.014)**	0.061 (0.020)**	0.063 (0.020)**	0.041 (0.017)**	0.042 (0.017)**
Primary	0.024 (0.014)*	0.027 (0.013)**	0.054 (0.022)**	0.057 (0.021)**	0.034 (0.017)**	0.036 (0.017)**
Middle school	0.018 (0.014)	0.022 (0.014)	0.037 (0.019)*	0.040 (0.019)**	0.031 (0.017)*	0.033 (0.017)*
High school	0.020 (0.015)	0.024 (0.015)	0.028 (0.021)	0.031 (0.021)	0.030 (0.020)	0.033 (0.020)*
Tech/College+	0.015 (0.018)	0.018 (0.017)	0.023 (0.022)	0.026 (0.022)	0.041 (0.024)*	0.043 (0.024)*
Log real prod assets 1 ^a	0.000 (0.002)	0.000 (0.002)	0.002 (0.003)	0.002 (0.003)	-0.002 (0.002)	-0.002 (0.002)
Log real prod assets 2 ^a	-0.001 (0.002)	-0.001 (0.002)	-0.002 (0.003)	-0.002 (0.003)	0.000 (0.002)	0.000 (0.002)
Land	0.000 (0.001)	0.000 (0.001)	-0.002 (0.003)	-0.002 (0.003)	0.000 (0.001)	0.000 (0.001)
Married		-0.018 (0.013)		-0.016 (0.023)		-0.007 (0.015)
Divorced Separated		0.093 (0.051)*		0.037 (0.081)		0.149 (0.067)**
Widowed		0.028 (0.031)		0.066 (0.054)		0.013 (0.039)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	12760	12760	4087	4087	8673	8673
R-squared	0.12	0.13	0.14	0.15	0.138	0.140
P-value for testing coefficients equal to zero						
Education	0.1990	0.1596	0.0480	0.0305	0.2672	0.2481
Assets	0.4417	0.4441	0.7186	0.6842	0.3596	0.3623
Age dummies	0.5040	0.6021	0.0042	0.0016	0.2572	0.2900
Cohort dummies	0.1535	0.1795	0.2674	0.2954	0.6166	0.6243
Marital Status		0.0302		0.2733		0.1184
Community dummies	0.0004	0.0003	0.0494	0.0561	0.0011	0.0010

Table continues

Table C.3 (cont'd)

Table 3.5 (cont'd)						
Variables	Overall		Female		Rural	
			Urban			
Some primary education	-0.018 (0.010)*	-0.017 (0.010)*	-0.017 (0.019)	-0.015 (0.019)	-0.010 (0.011)	-0.010 (0.011)
Primary	-0.038 (0.010)**	-0.037 (0.010)**	-0.038 (0.019)*	-0.035 (0.019)*	-0.030 (0.012)**	-0.030 (0.012)**
Middle school	-0.018 (0.011)*	-0.018 (0.011)*	-0.045 (0.019)**	-0.043 (0.019)**	-0.006 (0.013)	-0.006 (0.013)
High school	-0.002 (0.013)	-0.002 (0.013)	-0.038 (0.021)*	-0.036 (0.021)*	0.007 (0.018)	0.007 (0.018)
Tech/College+	-0.002 (0.017)	-0.002 (0.017)	-0.033 (0.022)	-0.031 (0.022)	-0.005 (0.030)	-0.005 (0.030)
Log real prod assets 1 ^a	0.000 (0.002)	0.000 (0.002)	-0.003 (0.003)	-0.003 (0.003)	0.001 (0.002)	0.001 (0.002)
Log real prod assets 2 ^a	0.000 (0.002)	0.000 (0.002)	0.003 (0.003)	0.003 (0.003)	0.000 (0.002)	0.000 (0.002)
Land	0.000 (0.001)	0.000 (0.001)	-0.003 (0.004)	-0.004 (0.004)	0.000 (0.001)	0.000 (0.001)
Married		-0.012 (0.016)		-0.012 (0.029)		-0.002 (0.019)
Divorced Separated		0.021 (0.036)		0.021 (0.046)		0.044 (0.068)
Widowed		0.015 (0.023)		0.017 (0.037)		0.012 (0.030)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	13652	13652	4493	4493	9159	9159
R-squared	0.12	0.125	0.113	0.11	0.144	0.144
P-value for testing coefficients equal to zero						
Education	0.0014	0.0021	0.2239	0.2743	0.0566	0.0594
Assets	0.9680	0.9631	0.5520	0.5579	0.8849	0.8798
Age dummies	0.1619	0.1850	0.1648	0.1649	0.3480	0.3510
Cohort dummies	0.5156	0.5233	0.4815	0.4799	0.3047	0.3106
Marital Status		0.2978		0.5250		0.8205
Community dummies	0.0000	0.0000	0.0176	0.0184	0.0000	0.0000

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Table C.4: Probability Model of Undernourishment: By Age Groups

Variables	Male					
	Age 20-39		Age 40-59		Age 60+	
Some primary education	-0.010 (0.030)	-0.008 (0.030)	0.037 (0.020)*	0.045 (0.020)**	0.066 (0.028)**	0.065 (0.028)**
Primary	-0.021 (0.031)	-0.020 (0.031)	0.050 (0.019)**	0.057 (0.019)**	0.021 (0.032)	0.020 (0.032)
Middle school	-0.020 (0.030)	-0.018 (0.030)	0.024 (0.019)	0.034 (0.019)*	0.036 (0.035)	0.037 (0.035)
High school	-0.030 (0.031)	-0.028 (0.031)	0.034 (0.025)	0.042 (0.025)*	0.001 (0.053)	-0.004 (0.053)
Tech/College+	-0.044 (0.035)	-0.042 (0.035)	0.023 (0.028)	0.036 (0.029)	0.026 (0.045)	0.026 (0.045)
Log real prod assets 1 ^a	0.000 (0.002)	0.000 (0.002)	-0.002 (0.002)	-0.002 (0.002)	0.005 (0.007)	0.005 (0.007)
Log real prod assets 2 ^a	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.008)	-0.003 (0.008)
Land	0.001 (0.001)	0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.008 (0.004)**	0.008 (0.004)**
Married		-0.002 (0.014)		-0.070 (0.038)*		0.043 (0.066)
Divorced Separated		0.044 (0.066)		-0.020 (0.087)		0.293 (0.162)*
Widowed		-0.010 (0.026)		0.097 (0.066)		0.028 (0.073)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	6678	6678	4292	4292	1790	1790
R-squared	0.162	0.162	0.224	0.232	0.413	0.417
P-value for testing coefficients equal to zero						
Education	0.5802	0.5848	0.1526	0.0836	0.3047	0.2960
Assets	0.5008	0.5076	0.1157	0.1277	0.7938	0.7907
Age dummies	0.7063	0.7658	0.6773	0.6583	0.6089	0.6328
Cohort dummies	0.6464	0.6479	0.0986	0.0885	0.8436	0.8418
Marital Status		0.8886		0.0140		0.3030
Community dummies	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table continues

Table C.4 (cont'd)

Variables	Female					
	Age 20-39		Age 40-59		Age 60+	
Some primary education	-0.026 (0.016)*	-0.026 (0.016)*	0.002 (0.013)	0.003 (0.013)	0.020 (0.031)	0.021 (0.031)
Primary	-0.030 (0.016)*	-0.030 (0.016)*	-0.032 (0.014)**	-0.031 (0.014)**	-0.021 (0.034)	-0.021 (0.034)
Middle school	-0.026 (0.016)*	-0.026 (0.016)*	0.003 (0.017)	0.004 (0.017)	-0.018 (0.037)	-0.019 (0.038)
High school	-0.010 (0.019)	-0.011 (0.019)	-0.044 (0.022)**	-0.044 (0.022)**	-0.110 (0.062)*	-0.113 (0.062)*
Tech/College+	0.003 (0.024)	0.001 (0.024)	-0.064 (0.023)**	-0.064 (0.023)**	-0.065 (0.049)	-0.063 (0.050)
Log real prod assets 1 ^a	-0.001 (0.002)	-0.001 (0.002)	-0.002 (0.003)	-0.002 (0.003)	0.001 (0.006)	0.001 (0.006)
Log real prod assets 2 ^a	0.002 (0.002)	0.002 (0.002)	0.000 (0.003)	0.000 (0.003)	-0.003 (0.005)	-0.003 (0.005)
Land	-0.001 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.001)	-0.001 (0.004)	-0.001 (0.004)
Married		-0.016 (0.018)		0.027 (0.030)		0.012 (0.054)
Divorced Separated		0.007 (0.057)		0.014 (0.045)		0.072 (0.111)
Widowed		-0.003 (0.063)		0.056 (0.040)		0.021 (0.056)
Community *year dummy	Yes	Yes	Yes	Yes	Yes	Yes
No. of Observations	7177	7177	4578	4578	1897	1897
R-squared	0.146	0.147	0.222	0.222	0.388	0.389
P-value for testing coefficients equal to zero						
Education	0.2009	0.2306	0.0037	0.0037	0.3050	0.2979
Assets	0.5701	0.5425	0.5372	0.5489	0.8340	0.8604
Age dummies	0.2308	0.1956	0.6336	0.6656	0.2185	0.1975
Cohort dummies	0.8244	0.8265	0.0493	0.0528	0.3416	0.3656
Marital Status		0.7889		0.5292		0.9071
Community dummies	0.9775	0.9729	0.0000	0.0000	0.0000	0.0000

Note: Also included in the regressions are age dummies and five-year cohort dummies. Person level cluster robust standard errors are in parentheses. * indicates statistical significance at 0.1 level and ** at 0.05 level.

Appendix D: AR1 Model of Determinants of BMI

Table D.1: AR1 Models for BMI

	Male		Female	
Some primary schooling	0.066 (0.119)	0.043 (0.119)	0.058 (0.105)	0.054 (0.105)
Primary	-0.013 (0.124)	-0.044 (0.124)	0.270 (0.119)**	0.264 (0.119)**
Middle school	0.003 (0.127)	-0.036 (0.127)	-0.018 (0.129)	-0.024 (0.129)
High school	0.131 (0.148)	0.098 (0.148)	-0.345 (0.167)**	-0.357 (0.167)**
Tech/College+	-0.035 (0.170)	-0.059 (0.170)	-0.493 (0.211)**	-0.515 (0.212)**
Log real prod asset 1	0.023 (0.017)	0.021 (0.017)	0.044 (0.019)**	0.045 (0.019)**
Log real prod asset 2	-0.002 (0.017)	0 (0.017)	-0.021 (0.019)	-0.021 (0.019)
Land farmed	-0.002 (0.007)	-0.002 (0.007)	0.002 (0.009)	0.002 (0.009)
Married		0.388 (0.118)**		-0.196 (0.169)
Divorced Separated		-0.349 (0.350)		-0.222 (0.425)
Widowed		-0.005 (0.239)		-0.409 (0.235)*
Number of obs	12760	12760	13652	13652
Community*year dummies	Yes	Yes	Yes	Yes
Wald chi2	2549.10	2571.00	2032.62	2036.13
Rho estimated autocorrelation coefficient	0.434	0.434	0.495	0.495
fraction of variance due to u_i	0.295	0.293	0.444	0.443
modified Bhargava et al. Durbin-Watson	1.438	1.438	1.361	1.361
Baltagi-Wu LBI	2.365	2.365	2.293	2.294

Appendix E: Descriptive Statistics of Nutrient Intakes in CHNS89-93

Table E.1: Patterns and Trends of Individual Daily Nutrient Intakes in CHNS 1989-1993

		Male			Female		
		1989	1991	1993	1989	1991	1993
Overall							
	No. Obs	2074	3413	3208	2242	3752	3476
Calorie (kcal)	Median	2724.65	2726.46	2603.09	2359.46	2330.36	2245.11
	Mean	2816.94	2794.99	2688.56	2441.09	2403.25	2313.70
	(std)	(870.10)	(816.16)	(815.98)	(712.18)	(705.54)	(672.66)
Protein (g)	Median	95.20	97.18	95.11	82.65	84.85	84.72
	Mean	101.18	102.06	100.44	88.94	89.48	88.23
	(std)	(39.56)	(36.57)	(36.73)	(34.83)	(32.32)	(32.04)
Fat (g)	Median	38.40	44.01	45.69	31.62	36.16	39.06
	Mean	47.65	51.76	53.86	40.57	44.31	45.87
	(std)	(38.79)	(39.65)	(43.01)	(34.04)	(36.43)	(36.48)
Carbohydrates (g)	Median	462.01	443.60	410.88	408.96	389.65	368.87
	Mean	482.57	462.36	434.09	427.91	408.72	385.06
	(std)	(159.05)	(151.25)	(151.14)	(135.80)	(134.18)	(128.35)
% Calorie from fat	Median	13.51	15.59	17.14	12.43	15.11	16.80
	Mean	14.61	16.15	17.45	14.39	16.01	17.28
	(std)	(9.62)	(10.09)	(10.72)	(9.80)	(10.32)	(10.76)
% Calorie from protein	Median	13.87	14.13	14.34	14.02	14.34	14.60
	Mean	14.37	14.70	15.04	14.57	14.98	14.60
	(std)	(3.33)	(3.41)	(3.59)	(3.56)	(3.47)	14.60
% Calorie from carbohydrates	Median	70.43	67.27	65.79	71.95	68.91	67.26
	Mean	69.42	66.97	65.38	70.70	68.52	67.06
	(std)	(12.54)	(13.03)	(13.61)	(11.72)	(12.17)	(12.49)
Urban Areas							
	No. Obs	633	1194	1015	694	1332	1110
Calorie (kcal)	Median	2520.74	2523.25	2443.17	2151.21	2134.54	2101.66
	Mean	2598.40	2577.65	2500.23	2214.54	2214.52	2149.33
	(std)	(763.02)	(728.01)	(749.07)	(609.99)	(648.88)	(626.39)
Protein (g)	Median	98.45	98.43	96.30	84.01	85.85	85.47
	Mean	101.84	103.34	101.92	89.26	90.22	89.52
	(std)	(36.75)	(36.76)	(37.63)	(33.15)	(32.42)	(31.79)
Fat (g)	Median	49.65	53.69	57.57	40.23	45.76	49.88
	Mean	55.43	59.57	63.10	46.41	51.63	54.51
	(std)	(38.62)	(37.77)	(39.38)	(31.72)	(36.20)	(35.81)
Carbohydrates (g)	Median	398.33	379.71	352.77	345.49	326.90	315.30
	Mean	410.51	392.87	368.06	357.74	343.44	323.73
	(std)	(112.62)	(116.13)	(115.63)	(102.17)	(104.42)	(99.56)
% Calorie from fat	Median	18.14	19.74	21.95	17.70	19.86	21.64
	Mean	18.27	20.01	21.86	18.09	20.02	21.92
	(std)	(8.92)	(9.55)	(9.39)	(9.36)	(9.72)	(9.84)
% Calorie from protein	Median	15.25	15.62	15.89	15.81	15.85	16.24
	Mean	15.63	16.06	16.33	16.02	16.33	16.70
	(std)	(3.16)	(3.37)	(3.63)	(3.44)	(3.45)	(3.68)
% Calorie from carbohydrates	Median	64.51	62.25	60.08	64.97	63.21	61.00
	Mean	64.49	61.98	59.93	65.50	62.99	61.11
	(std)	(11.26)	(11.48)	(11.47)	(10.85)	(10.99)	(10.96)

Table continues

Table continued		Male			Female		
		1989	1991	1993	1989	1991	1993
Age 20-39							
	No. Obs	1686	1682	1435	1838	1924	1579
Calorie (kcal)	Median	2719.27	2830.43	2699.58	2356.48	2423.88	2336.42
	Mean	2815.52	2913.59	2802.60	2439.61	2509.81	2405.95
	(std)	(871.61)	(843.48)	(804.05)	(711.81)	(711.15)	(677.22)
Protein (g)	Median	95.47	100.73	99.02	83.21	88.84	88.42
	Mean	101.75	106.31	104.92	89.51	93.41	92.01
	(std)	(40.27)	(37.79)	(37.08)	(35.47)	(32.83)	(32.18)
Fat (g)	Median	38.46	45.54	47.45	32.74	37.66	40.82
	Mean	47.57	53.17	55.92	41.36	46.52	47.50
	(std)	(16.28)	(15.46)	(14.98)	(13.35)	(13.42)	(13.01)
Carbohydrates (g)	Median	462.40	467.36	440.39	406.71	408.85	384.56
	Mean	482.25	486.19	457.07	425.49	426.99	400.89
	(std)	(160.28)	(154.46)	(149.80)	(135.35)	(135.42)	(133.01)
% Calorie from fat	Median	13.56	15.55	16.98	13.14	15.18	16.70
	Mean	14.61	15.88	17.43	14.69	16.10	17.34
	(std)	(9.66)	(9.74)	(10.83)	(9.92)	(10.25)	(10.76)
% Calorie from protein	Median	13.92	14.11	14.38	14.07	14.32	14.57
	Mean	14.45	14.69	15.04	14.66	14.98	15.39
	(std)	(3.35)	(3.38)	(3.49)	(3.61)	(3.50)	(3.71)
% Calorie from carbohydrates	Median	70.38	67.69	66.22	71.33	68.98	67.34
	Mean	69.38	67.57	65.97	70.36	68.54	66.99
	(std)	(12.62)	(12.71)	(13.26)	(11.82)	(11.94)	(12.48)
Age 40-59							
	No. Obs	388	1185	1198	404	1262	1327
Calorie (kcal)	Median	2773.12	2747.12	2635.58	2371.04	2347.64	2290.76
	Mean	2823.10	2811.20	2727.54	2447.86	2431.95	2349.42
	(std)	(864.63)	(764.96)	(812.15)	(714.74)	(674.80)	(651.01)
Protein (g)	Median	93.44	97.00	95.97	80.75	84.62	85.33
	Mean	98.74	100.86	101.26	86.34	89.05	88.67
	(std)	(36.27)	(33.80)	(36.17)	(31.69)	(30.98)	(31.34)
Fat (g)	Median	37.44	43.63	44.89	26.32	35.69	38.57
	Mean	47.99	51.62	53.30	36.98	43.58	45.99
	(std)	(38.53)	(39.45)	(43.17)	(32.60)	(35.67)	(37.53)
Carbohydrates (g)	Median	457.26	444.53	417.61	423.05	395.59	382.56
	Mean	483.99	464.77	441.14	438.93	417.21	393.03
	(std)	(153.74)	(145.96)	(152.21)	(137.46)	(131.00)	(123.65)
% Calorie from fat	Median	13.31	15.54	16.70	10.60	14.55	16.49
	Mean	14.61	16.02	17.04	13.07	15.55	16.96
	(std)	(9.42)	(10.23)	(10.52)	(9.16)	(10.29)	(10.87)
% Calorie from protein	Median	13.62	13.95	14.23	13.92	14.15	14.54
	Mean	14.03	14.46	14.98	14.17	14.72	15.15
	(std)	(3.25)	(3.29)	(3.64)	(3.28)	(3.35)	(3.63)
% Calorie from carbohydrates	Median	70.45	67.58	65.91	74.41	69.72	67.73
	Mean	69.58	66.95	65.50	72.24	69.17	67.53
	(std)	(12.20)	(13.14)	(13.79)	(11.10)	(12.17)	(12.51)

Table continues on next page

Table continued		Male			Female		
		1989	1991	1993	1989	1991	1993
No Formal Education							
	No. Obs	75	378	327	350	1196	1073
Calorie (kcal)	Median	3129.63	2569.97	2468.10	2635.49	2280.42	2187.89
	Mean	3133.97	2618.60	2602.30	2687.51	2344.65	2275.60
	(std)	(983.27)	(814.04)	(952.81)	(768.87)	(720.97)	(705.67)
Protein (g)	Median	102.55	88.04	83.30	82.65	79.36	77.27
	Mean	108.18	91.27	89.60	89.25	83.38	81.92
	(std)	(45.26)	(32.37)	(35.96)	(35.19)	(29.38)	(30.73)
Fat (g)	Median	26.17	29.44	30.64	19.93	26.63	27.57
	Mean	41.56	40.86	42.34	33.95	37.54	37.27
	(std)	(40.34)	(35.33)	(46.19)	(35.84)	(34.05)	(33.77)
Carbohydrates (g)	Median	563.48	445.64	410.53	489.65	400.62	381.45
	Mean	576.47	460.10	449.20	503.91	415.31	400.99
	(std)	(181.29)	(157.02)	(191.60)	(142.95)	(142.14)	(139.24)
% Calorie from fat	Median	8.56	11.35	12.07	6.89	11.38	11.46
	Mean	11.26	13.66	14.18	10.47	13.94	14.32
	(std)	(8.96)	(10.18)	(10.85)	(8.97)	(10.40)	(10.34)
% Calorie from protein	Median	13.34	13.51	13.35	12.88	13.72	13.70
	Mean	13.65	14.07	13.90	13.20	14.37	14.44
	(std)	(3.16)	(3.11)	(3.11)	(2.97)	(3.25)	(3.28)
% Calorie from carbohydrates	Median	76.24	72.31	71.89	79.36	72.90	72.76
	Mean	74.54	70.87	69.70	75.98	71.13	70.83
	(std)	(11.26)	(12.48)	(13.91)	(10.71)	(12.26)	(12.05)
Some Primary Schooling							
	No. Obs	272	595	758	390	575	779
Calorie (kcal)	Median	2938.05	2662.15	2587.53	2390.16	2370.51	2314.79
	Mean	3001.04	2786.70	2657.82	2475.20	2467.40	2378.96
	(std)	(907.35)	(854.34)	(813.28)	(678.10)	(718.48)	(686.44)
Protein (g)	Median	95.12	92.99	90.93	76.85	84.87	85.06
	Mean	101.46	97.74	96.40	84.13	88.59	88.36
	(std)	(38.33)	(35.21)	(34.91)	(30.19)	(31.60)	(31.96)
Fat (g)	Median	32.30	39.72	42.66	22.29	34.03	38.48
	Mean	43.22	47.62	51.52	32.12	41.43	45.14
	No. Obs	(35.33)	(39.38)	(44.38)	(28.09)	(34.15)	(35.62)
Carbohydrates (g)	Median	508.08	450.64	409.88	436.28	408.89	393.44
	Mean	538.34	471.13	431.77	460.59	431.49	403.12
	(std)	(178.73)	(154.53)	(147.59)	(142.66)	(142.02)	(134.13)
% Calorie from fat	Median	9.91	13.88	16.67	8.34	13.86	16.26
	Mean	12.74	14.85	16.84	11.48	14.67	16.63
	No. Obs	(9.27)	(9.89)	(10.71)	(8.65)	(9.72)	(10.72)
% Calorie from protein	Median	13.10	13.49	13.82	13.03	13.77	14.27
	Mean	13.56	14.16	14.60	13.70	14.45	14.94
	(std)	(2.89)	(3.33)	(3.45)	(3.40)	(3.30)	(3.58)
% Calorie from carbohydrates	Median	74.53	69.20	66.68	76.86	70.67	68.52
	Mean	72.16	68.53	65.95	74.54	70.36	68.13
	(std)	(11.68)	(12.71)	(13.80)	(10.47)	(11.57)	(12.32)

Table continues

Table continued		Male			Female		
		1989	1991	1993	1989	1991	1993
Middle School							
	No. Obs	766	1024	996	659	832	764
Calorie (kcal)	Median	2683.71	2772.08	2659.74	2303.54	2316.80	2272.34
	Mean	2808.16	2854.06	2735.91	2360.76	2406.16	2330.88
	(std)	(884.38)	(818.02)	(759.66)	(667.80)	(685.69)	(659.49)
Protein (g)	Median	95.99	98.88	98.27	84.73	87.38	89.08
	Mean	101.46	104.48	104.52	89.08	92.78	91.92
	(std)	(38.22)	(36.54)	(37.35)	(32.42)	(33.18)	(32.12)
Fat (g)	Median	35.78	44.70	50.88	35.36	40.38	45.32
	Mean	47.12	52.29	57.46	43.72	47.66	50.51
	No. Obs	(40.10)	(39.55)	(43.16)	(32.72)	(36.40)	(35.18)
Carbohydrates (g)	Median	462.73	453.89	419.27	381.28	382.05	363.01
	Mean	483.29	471.68	437.13	401.04	399.26	375.39
	(std)	(164.50)	(154.54)	(141.31)	(124.54)	(127.60)	(121.65)
% Calorie from fat	Median	13.04	15.29	18.04	15.28	16.52	18.39
	Mean	14.41	16.03	18.34	16.10	17.19	18.95
	No. Obs	(9.68)	(9.81)	(10.79)	(9.76)	(10.13)	(10.27)
% Calorie from protein	Median	14.00	14.12	14.58	14.54	14.93	15.20
	Mean	14.47	14.74	15.33	15.07	15.47	15.84
	(std)	(3.28)	(3.31)	(3.62)	(3.47)	(3.57)	(3.63)
% Calorie from carbohydrates	Median	71.36	67.82	65.20	69.12	67.13	65.18
	Mean	69.70	66.87	64.57	68.55	66.96	64.92
	(std)	(12.64)	(13.17)	(13.58)	(11.40)	(11.81)	(11.84)
High School							
	No. Obs	324	422	368	263	336	280
Calorie (kcal)	Median	2639.68	2749.21	2644.53	2178.72	2209.19	2241.90
	Mean	2721.10	2801.24	2720.45	2290.17	2307.26	2322.62
	(std)	(837.45)	(791.93)	(757.54)	(697.17)	(625.33)	(689.44)
Protein (g)	Median	97.04	100.74	101.98	83.48	87.08	93.63
	Mean	102.69	107.12	107.11	91.13	92.11	97.07
	(std)	(40.94)	(39.69)	(34.86)	(39.97)	(32.31)	(32.93)
Fat (g)	Median	46.22	53.43	52.89	38.21	46.26	50.35
	Mean	52.28	59.36	59.85	46.98	53.88	59.72
	(std)	(39.29)	(42.48)	(37.97)	(37.51)	(35.28)	(48.25)
Carbohydrates (g)	Median	431.03	426.71	407.58	374.73	348.21	336.61
	Mean	444.89	442.88	421.43	373.93	360.67	347.69
	(std)	(137.50)	(130.79)	(135.51)	(104.61)	(101.91)	(105.51)
% Calorie from fat	Median	16.54	18.38	19.49	17.19	19.84	21.72
	Mean	16.48	18.26	19.37	17.30	20.18	21.89
	(std)	(9.49)	(9.97)	(9.77)	(9.97)	(9.44)	(11.01)
% Calorie from protein	Median	14.67	14.88	15.42	15.45	15.39	16.31
	Mean	15.06	15.32	15.91	15.68	16.03	16.85
	(std)	(3.17)	(3.49)	(3.58)	(3.52)	(3.59)	(3.83)
% Calorie from carbohydrates	Median	67.05	64.04	62.91	66.38	63.37	60.66
	Mean	66.63	64.38	62.62	66.74	63.28	61.00
	(std)	(12.12)	(12.76)	(12.82)	(11.73)	(10.77)	(12.33)

Note: The daily intake of energy, fat, protein and carbohydrates contents are based on the 1991 China Food Composition Tables (FCT). The energy content of fat, protein and carbohydrates is calculated based on the formulae that one gram of fat yields 9 kcal of energy; one gram of protein and carbohydrate each yields 4 kcal of energy.

Appendix F: First Stage

Regressions for section 2.6

A simple health production function analysis was carried out in chapter two in which the lagged health (lagged log weight and height, or lagged log BMI in 1991), current physical activity levels and nutrient intakes are included and treated as endogenous. Also included in the regression are individual year age dummies. The identifying instruments are education dummies, real productive asset splines, land cultivated and community characteristics including prices, water and sanitation conditions in 1991.

The first stage regressions for these endogenous variables on all exogenous factors are listed here.

Table F.1: First Stage Regression for Health Production Function Analysis

	Male									
	Log weight 1991	Log height 1991	Log BMI 1991	Activities 1993	Calories (100 Kcal)	Protein (g)	Fat (g)	% Calories From Fat	% Calories From Protein	
Some primary education	0.010 (0.011)	0.007 (0.003)**	-0.003 (0.009)	-0.038 (0.028)	-0.462 (0.599)	-0.932 (2.725)	3.739 (3.236)	1.484 (0.761)*	0.169 (0.250)	
Primary	0.014 (0.012)	0.008 (0.003)**	-0.002 (0.010)	-0.028 (0.030)	-0.308 (0.645)	-0.863 (2.934)	2.126 (3.485)	0.899 (0.819)	0.176 (0.269)	
Middle school	0.028 (0.011)**	0.012 (0.003)**	0.003 (0.010)	-0.069 (0.029)**	-0.820 (0.632)	1.296 (2.875)	7.467 (3.415)**	2.976 (0.803)**	0.693 (0.263)**	
High school	0.034 (0.014)**	0.015 (0.004)**	0.005 (0.012)	-0.127 (0.035)**	-0.683 (0.750)	3.197 (3.412)	8.754 (4.053)**	3.402 (0.953)**	1.026 (0.312)**	
Tech/College+	0.044 (0.016)**	0.020 (0.004)**	0.004 (0.014)	-0.502 (0.040)**	-1.002 (0.868)	-0.027 (3.953)	6.007 (4.694)	2.299 (1.104)**	0.762 (0.362)**	
Log real prod assets 1	0.000 (0.002)	0.000 (0.000)	0 (0.002)	-0.002 (0.005)	0.095 (0.103)	0.961 (0.471)**	1.240 (0.559)**	0.347 (0.131)**	0.095 (0.043)**	
Log real prod assets 2	0.004 (0.002)**	0.001 (0.001)*	0.002 (0.002)	0.001 (0.005)	-0.068 (0.106)	-0.443 (0.485)	-0.594 (0.576)	-0.139 (0.135)	-0.048 (0.044)	
Land	0.000 (0.001)	0.000 (0.000)	0 (0.001)	0.015 (0.003)**	0.180 (0.056)**	-0.126 (0.254)	-1.508 (0.302)**	-0.503 (0.071)**	-0.121 (0.023)**	
Price of Rice	0.006 (0.010)	0.002 (0.003)	0.003 (0.009)	-0.034 (0.026)	0.785 (0.553)	10.864 (2.516)**	12.807 (2.988)**	2.868 (0.703)**	1.065 (0.230)**	
Price of Pork	-0.001 (0.001)	0.000 (0.000)	-0.001 (0.001)	-0.002 (0.002)	-0.122 (0.045)**	-0.490 (0.203)**	-0.532 (0.242)**	-0.124 (0.057)**	-0.011 (0.019)	
Price of Eggs	-0.004 (0.001)**	0.000 (0.000)	-0.004 (0.001)**	0.005 (0.004)	-0.030 (0.076)	0.659 (0.347)*	0.636 (0.412)	0.225 (0.097)**	0.105 (0.032)**	
Price of Oil	0.003 (0.002)**	-0.001 (0.000)**	0.005 (0.001)**	-0.003 (0.004)	0.067 (0.087)	-1.435 (0.394)**	-0.848 (0.468)*	-0.231 (0.110)**	-0.245 (0.036)**	
Water Source 2	-0.084 (0.013)**	-0.019 (0.003)**	-0.046 (0.011)**	0.081 (0.033)**	-1.595 (0.718)**	-10.285 (3.266)**	8.306 (3.879)**	1.955 (0.912)**	-0.734 (0.299)**	

Table continues

Table F.1 (cont'd)

	Male									
	Log weight 1991	Log height 1991	Log BMI 1991	Activities 1993	Calories (100 Kcal)	Protein (g)	Fat (g)	% Calories From Fat	% Calories From Protein	
Water Source 3	-0.075 (0.016)**	-0.015 (0.004)**	-0.046 (0.013)**	0.039 (0.040)	-0.477 (0.852)	-6.242 (3.877)	10.292 (4.605)**	2.889 (1.082)**	-0.806 (0.355)**	
Water Source 4	-0.072 (0.009)**	-0.018 (0.003)**	-0.036 (0.008)**	-0.032 (0.024)	-1.810 (0.521)**	-1.961 (2.372)	7.773 (2.817)**	3.086 (0.662)**	0.664 (0.217)**	
In house no flush	-0.048 (0.023)**	-0.007 (0.006)	-0.035 (0.020)*	0.110 (0.058)*	1.577 (1.246)	10.847 (5.670)*	-0.590 (6.734)	-1.402 (1.583)	0.503 (0.519)	
Outside toilets	-0.025 (0.013)*	-0.005 (0.004)	-0.015 (0.011)	0.013 (0.034)	0.279 (0.730)	0.786 (3.323)	6.049 (3.947)	0.705 (0.928)	-0.175 (0.304)	
Open pit	-0.112 (0.015)**	-0.026 (0.004)**	-0.060 (0.013)**	0.220 (0.037)**	1.095 (0.808)	-1.521 (3.676)	-3.834 (4.366)	-2.899 (1.026)**	-0.793 (0.337)**	
No toilets	-0.280 (0.070)**	-0.081 (0.018)**	-0.118 (0.060)**	-0.197 (0.177)	-4.635 (3.821)	-39.008 (17.393)**	-21.285 (20.657)	-2.715 (4.856)	-1.353 (1.593)	
Very little excreta	-0.048 (0.014)**	-0.010 (0.004)**	-0.027 (0.012)**	0.131 (0.036)**	-1.345 (0.785)*	-16.309 (3.572)**	-19.245 (4.242)**	-5.886 (0.997)**	-1.697 (0.327)**	
Some excreta	-0.041 (0.014)**	-0.010 (0.004)**	-0.021 (0.012)*	0.160 (0.036)**	2.016 (0.783)**	-2.002 (3.564)	6.445 (4.232)	0.307 (0.995)	-1.374 (0.326)**	
Much excreta	-0.398 (0.082)**	-0.088 (0.022)**	-0.222 (0.071)**	0.313 (0.209)	-7.734 (4.502)*	-66.126 (20.496)**	-48.475 (24.341)**	-13.264 (5.722)**	-7.465 (1.877)**	
Constant	4.137 (0.033)**	5.144 (0.009)**	3.060 (0.029)**	0.747 (0.085)**	28.598 (1.835)**	117.655 (8.355)**	58.065 (9.923)**	19.674 (2.333)**	16.582 (0.765)**	
Number of obs	2556	2556	2556	2556	2556	2556	2556	2556	2556	
R-squared	0.160	0.174	0.075	0.404	0.095	0.083	0.074	0.181	0.226	
F(All)	6.64	7.28	3.42	21.17	4.12	3.68	3.38	7.56	9.69	
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
F(Identifying)	18.67	18.09	7.99	62.45	5.96	7.03	13.74	31.52	38.10	
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Table continues

	Male				Female				
	Log weight 1991	Log height 1991	Log BMI 1991	Activities 1993	Calories (100 Kcal)	Protein (g)	Fat (g)	% Calories From Fat	% Calories From Protein
Some primary education	0.017 (0.009)*	0.001 (0.002)	0.016 (0.008)*	-0.055 (0.020)**	-0.372 (0.355)	1.591 (1.740)	4.711 (2.004)**	2.018 (0.544)**	0.571 (0.181)**
Primary	0.024 (0.011)**	0.004 (0.003)*	0.016 (0.010)	-0.054 (0.024)**	-1.216 (0.427)**	-0.455 (2.091)	5.130 (2.408)**	2.721 (0.654)**	0.688 (0.218)**
Middle school	0.023 (0.011)**	0.009 (0.002)**	0.006 (0.010)	-0.103 (0.024)**	-1.089 (0.418)**	1.926 (2.045)	8.399 (2.355)**	3.765 (0.640)**	1.074 (0.213)**
High school	0.029 (0.014)**	0.013 (0.003)**	0.003 (0.013)	-0.179 (0.032)**	-0.827 (0.558)	5.131 (2.734)*	15.123 (3.149)**	5.128 (0.856)**	1.467 (0.285)**
Tech/College+	0.007 (0.019)	0.015 (0.004)**	-0.022 (0.017)	-0.309 (0.043)**	-1.176 (0.745)	3.979 (3.648)	13.679 (4.201)**	5.556 (1.141)**	1.475 (0.380)**
Log real prod assets 1	0.006 (0.002)**	0.001 (0)	0.005 (0.002)**	-0.001 (0.005)	0.029 (0.079)	0.908 (0.385)**	0.953 (0.444)**	0.361 (0.121)**	0.123 (0.040)**
Log real prod assets 2	0.000 (0.002)	0.001 (0.000)	-0.001 (0.002)	-0.001 (0.005)	-0.084 (0.081)	-0.892 (0.398)**	-0.869 (0.458)*	-0.213 (0.125)*	-0.088 (0.041)**
Land	0.000 (0.001)	0.000 (0.000)	0 (0.001)	0.022 (0.003)**	0.214 (0.045)**	0.237 (0.223)	-0.629 (0.256)**	-0.335 (0.070)**	-0.102 (0.023)**
Price of Rice	0.030 (0.011)**	0.008 (0.002)**	0.014 (0.009)	-0.034 (0.024)	0.492 (0.415)	8.782 (2.030)**	8.203 (2.338)**	2.176 (0.635)**	0.958 (0.212)**
Price of Pork	-0.001 (0.001)	0.000 (0.000)	-0.001 (0.001)	-0.002 (0.002)	-0.057 (0.036)	-0.229 (0.174)	-0.261 (0.200)	-0.062 (0.054)	-0.006 (0.018)
Price of Eggs	-0.009 (0.001)**	-0.001 (0.000)**	-0.006 (0.001)**	0.010 (0.003)**	-0.032 (0.059)	0.514 (0.289)*	0.344 (0.333)	0.177 (0.090)**	0.102 (0.030)**
Price of Oil	0.002 (0.002)	-0.001 (0.000)**	0.005 (0.001)**	0.002 (0.004)	-0.016 (0.066)	-0.892 (0.321)**	-0.743 (0.370)**	-0.168 (0.100)*	-0.141 (0.033)**
Water Source 2	-0.090 (0.014)**	-0.019 (0.003)**	-0.052 (0.012)**	0.124 (0.031)**	-1.915 (0.547)**	-8.146 (2.680)**	1.403 (3.086)	0.916 (0.839)	-0.276 (0.279)

Table continues

Table F.1 (cont'd)

	Log weight 1991	Log height 1991	Log BMI 1991	Activities 1993	Calories (100 Kcal)	Protein (g)	Fat (g)	% Calories From Fat	% Calories From Protein
Female									
Water Source 3	-0.087 (0.016)**	-0.023 (0.004)**	-0.041 (0.015)**	0.042 (0.037)	-1.693 (0.644)**	-9.277 (3.156)**	5.290 (3.634)	2.173 (0.987)**	-0.724 (0.329)**
Water Source 4	-0.043 (0.010)**	-0.014 (0.002)**	-0.015 (0.009)	-0.045 (0.023)**	-1.612 (0.401)**	-0.553 (1.962)	5.892 (2.259)**	3.209 (0.614)**	0.899 (0.204)**
In house no flush	-0.020 (0.023)	-0.008 (0.005)	-0.004 (0.021)	0.274 (0.053)**	-0.599 (0.922)	-0.889 (4.514)	-5.713 (5.199)	-2.928 (1.412)**	0.031 (0.470)
Outside toilets	-0.002 (0.015)	-0.006 (0.003)*	0.010 (0.013)	0.123 (0.033)**	-0.535 (0.580)	-2.557 (2.840)	-3.082 (3.270)	-1.474 (0.889)*	-0.437 (0.296)
Open pit	-0.062 (0.016)**	-0.019 (0.004)**	-0.025 (0.014)*	0.364 (0.036)**	-0.271 (0.627)	-4.786 (3.070)	-8.920 (3.535)**	-3.963 (0.960)**	-0.873 (0.320)**
No toilets	-0.295 (0.071)**	-0.061 (0.017)**	-0.172 (0.064)**	0.094 (0.160)	-0.263 (2.799)	-34.477 (13.709)**	11.782 (15.787)	5.720 (4.289)	-4.464 (1.428)**
Very little excreta	-0.037 (0.015)**	-0.004 (0.004)	-0.029 (0.014)**	0.129 (0.035)**	-0.558 (0.607)	-11.087 (2.974)**	-12.206 (3.425)**	-5.082 (0.931)**	-1.629 (0.310)**
Some excreta	-0.062 (0.015)**	-0.005 (0.004)	-0.052 (0.014)**	0.181 (0.035)**	3.266 (0.604)**	4.083 (2.958)	2.179 (3.407)	-0.996 (0.926)	-1.263 (0.308)**
Much excreta	-0.389 (0.085)**	-0.109 (0.020)**	-0.170 (0.077)**	0.184 (0.193)	-6.346 (3.364)*	-63.882 (16.473)**	-47.433 (18.970)**	-15.193 (5.154)**	-8.727 (1.716)**
Constant	3.984 (0.036)**	5.063 (0.008)**	3.068 (0.032)**	0.441 (0.080)**	24.249 (1.397)**	88.335 (6.842)**	39.350 (7.880)**	15.012 (2.141)**	14.952 (0.713)**
Number of obs	2897	2897	2897	2897	2897	2897	2897	2897	2897
R-squared	0.132	0.165	0.086	0.416	0.120	0.074	0.088	0.209	0.236
F(All)	6.10	7.66	4.16	25.01	5.61	3.70	4.24	9.89	11.37
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
F(Identifying)	16.31	15.58	8.44	73.57	10.26	6.67	13.42	38.03	46.36
P-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Appendix G: The Design Matrix and Constraints Needed

1. Consecutive Years

Example (G1.1). Age: 20, 21, 22, 23; Year: 1989, 1990, 1991; Cohort: 1966, 67, 68, 69, 70, 71

$$X = \begin{bmatrix} c0 & c1 & c2 & c3 & c4 & c5 & c6 & a1 & a2 & a3 & a4 & t1 & t2 & t3 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\text{Rank}(X)=10, k=1+5+3+2=11$$

Linear dependency:

$$t3 = 0.5c2 + c3 + 1.5c4 + 2c5 + 2.5c6 + 0.5a2 + a3 + 1.5a4 - 0.5t2 - 1.5$$

Hence the number of constraints necessary is one after setting

$$\{\alpha_1 = 0, \beta_1 = 0, \gamma_1 = 0\}.$$

Example (G1.2). Age: 20, 21, 22, 23, 24; Year: 1989, 1990, 1991; Cohort: 1965, 66, 67, 68, 69, 70, 71

$$X = \begin{bmatrix} c0 & c1 & c2 & c3 & c4 & c5 & c6 & c7 & a1 & a2 & a3 & a4 & a5 & t1 & t2 & t3 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\text{Rank}(X)=12, k=1+6+4+2=13$$

Linear Dependency:

$$t3 = 0.5c2 + c3 + 1.5c4 + 2c5 + 2.5c6 + 3c7 + 0.5a2 + a3 + 1.5a4 + 2a5 - 0.5t2 - 2$$

Hence the number of constraints necessary is one after setting

$$\{\alpha_1 = 0, \beta_1 = 0, \gamma_1 = 0\}.$$

2. Biannual Data

Example (G2.1). Age: 20, 21, 22, 23; Year: 1989, 1991, 1993; Cohort: 1966, 67, 68, 69, 70, 71, 72, 73

$$X = \begin{bmatrix} c0 & c1 & c2 & c3 & c4 & c5 & c6 & c7 & c8 & a1 & a2 & a3 & a4 & t1 & t2 & t3 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$\text{Rank}(X)=11, k=1+7+3+2=13$$

Linear dependency:

$$a4 = 1 - c2 - c4 - c6 - c8 - a2$$

$$t3 = -0.5 * c2 + 0.5 * c3 + c5 + 0.5 * c6 + 1.5 * c7 + c8 - 0.5 * a2 + 0.5 * a3 - 0.5 * t2$$

Hence the number of constraints necessary is two after setting

$$\{\alpha_1 = 0, \beta_1 = 0, \gamma_1 = 0\}.$$

Bibliography

- Attanasio, O. P. and Hoynes, H. W. (2000), 'Differential mortality and wealth accumulation', *THE JOURNAL OF HUMAN RESOURCES* **35**, 1–29.
- Baltagi, B. H. and Wu, P. X. (1999), 'Unequally spaced panel data regressions with ar(1) disturbances', *Econometric Theory* **15**(6), 814–23.
- Baum, C. F., Schaffer, M. E. and Stillman, S. (2003), 'Instrumental variables and gmm: estimation and testing', *The Stata Journal* **3**(1), 1–31.
- Behrman, J. R. and Deolalikar, A. B. (1988), Health and nutrition, in H. Chenery and T. N. Srinivasan, eds, 'Handbook of Development Economics', Vol. 1, Elsevier Science Publishers B.V.
- Behrman, J. R., Foster, A. and Rosenzweig, M. R. (1997), 'The dynamics of agricultural production and the calorie-income relationship: evidence from pakistan', *Journal of Econometrics* **77**, 187–207.
- Behrman, J. R. and Wolfe, B. L. (1984), 'More evidence on nutrition demand: income seems overrated and women's schooling underemphasized', *Journal of Development Economics* **14**(1-2), 105–28.
- Behrman, J. R. and Wolfe, B. L. (1989), 'Does more schooling make women better nourished and healthier?', *Journal of Human Resources* **24**(4), 644–63.
- Bell, A. C., Ge, K. and Popkin, B. M. (2001), 'Weight gain and its predictors in chinese adults', *International Journal of Obesity* **25**, 1079–86.
- Bhandari, R. (1998), Education and food consumption behavior in China: household analysis and policy implications, PhD thesis, North Carolina State University, A Bell & Howell Company, Ann Arbor, MI.
- Bouis, H. E. (1994), 'The effect of income on demand for food in poor countries: are our food consumption databases giving us reliable estimates?', *Journal of Development Economics* **44**(1), 199–226.
- Bound, J., Jaeger, D. A. and Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and endogenous explanatory variable is weak', *Journal of the American Statistical Association* **90**, 443–50.
- Card, D. (2001), 'Estimating the return to schooling: progress on some persistent econometric problems', *Econometrica* **69**(5), 1127–60.
- Cebu-Study-Team (1992), 'A child health production function estimated from longitudinal data', *Journal of Development Economics* **38**(2), 323–51.
- Chen, X. (1990), 'Dietary standards and goals for the people's republic of china with special emphasis on nutritional requirements of chinese infants', *Progress in Food and Nutrition Science* **14**, 83–92.

- Costa, D. L. and Steckel, R. H. (1995), Long-term trends in health, welfare, and economic growth in the united states. National Bureau of Economic Research, Historical Paper no. 76.
- Deaton, A. (1988), 'Quality, quantity and spatial variation in price', *American Economic Review* **78**(3), 418–30.
- Deaton, A. (1997), *The Analysis of Household Survey: a Microeconometric Approach to Development Policy*, The Johns Hopkins University Press: Baltimore.
- Dowler, E. A. and Seo, Y. O. (1985), 'Assessment of energy intake: estimates of food supply vs measurement of food consumption', *Food Policy* **10**(3), 278–88.
- Drewnowski, A. and Popkin, B. M. (1997), 'The nutrition transition: new trends in the global diet', *Nutrition Reviews* **55**(2), 31.
- Du, S., Lu, B., Zhai, F. and Popkin, B. M. (2002), 'A new stage of the nutrition transition in china', *Public Health Nutrition* **5**(1A), 169–74.
- Evenson, R. E. and Pray, C. E. (1994), 'Measuring food production: with reference to south asia', *Journal of Development Economics* **44**(1), 173–97.
- Fogel, R. W. (1986), Nutrition and the decline in mortality since 1700: some preliminary findings, in S. L. Engerman and R. E. Gallman, eds, 'Long-Term Factors in American Economic Growth', The University of Chicago Press.
- Fogel, R. W. (1994), 'Economic growth, population theory and physiology: the bearing of long-term processes on the making of economic policy', *The American Economic Review* **84**(3), 369–95.
- Foster, A. D. (1995), 'Prices, credit markets and child growth in low-income rural areas', *The Economic Journal* **105**, 551–70.
- Ge, K., Chen, C. and Shen, T. (1991), 'Food consumption and nutritional status in china: Achievements, problems and policy implications, part i', *Food, Nutrition and Agriculture* **1**(2/3), 54–61.
- Ge, K., Chen, C., Shen, T. and Zhang, S. (1992), 'Food consumption and nutritional status in china: Achievements, problems and policy implications, part ii', *Food, Nutrition and Agriculture* **2**(4), 10–17.
- Ge, K., Weisell, R., Guo, X., Cheng, L., Ma, H., Zhai, F. and Popkin, B. M. (1994), 'The body mass index of chinese adults in the 1980s', *European Journal of Clinical Nutrition* **48**(Suppl 3), S148–54.
- Grossman, M. (1972), 'On the concept of health capital and the demand for health', *The Journal of Political Economy* **80**, 223–55.

- Grossman, M. (2000), The human capital model, in A. J. Culyer and J. P. Newhouse, eds, 'Handbook of Health Economics', Vol. 1a, Elsevier Science Publishers B.V., chapter 7.
- Gu, E. (2001), 'Market transition and the transformation of the health care systems in urban china', *Policy Studies* **22**(3-4), 197-215.
- Guo, X. (1998), The impact of household income and food prices on food consumption and dietary fat intake in China, 1989-1993, PhD thesis, University of North Carolina, UMI Dissertation Services: A Bell & Howell Company, Ann Arbor, MI.
- Guo, X., Mroz, T. A., Popkin, B. M. and Zhai, F. (2000), 'Structural change in the impact of income on food consumption in china, 1989-1993', *Economic Development and Cultural Change* **48**, 737-60.
- Guo, X., Popkin, B. M., Mroz, T. A. and Zhai, F. (1999), 'Food price policy can significantly reduce the negative effects of the nutrition transition in china', *Journal of Nutrition* **129**, 994-1001.
- Guo, X., Popkin, B. M. and Zhai, F. (1999), 'Patterns of change in food consumption and dietary fat intake in chinese adults, 1989-1993', *Food and Nutrition Bulletin* **20**(3), 344-53.
- Heckman, J. and Robb, R. (1985), Using longitudinal data to estimate age, period and cohort effects in earnings equation, in W. M. Mason and S. E. Fienberg, eds, 'Cohort Analysis in Social Research: Beyond the Identification Problem', New York: Springer-Verlag, pp. 137-50.
- Heilig, G. K. (1999), *China Food: Can China Feed Itself?*, IIASA, Laxenburg.
- Holford, T. R. (1983), 'The estimation of age, period and cohort effects for vital rates', *Biometrics* **39**, 311-24.
- Johnson, D. G. (2003), 'Provincial migration in china in the 1990s', *China Economic Review* **14**, 22-31.
- Ke, B. (1999), Policy and institutional change for agriculture in china: production, consumption and trade implications, in G. H. Peters and J. won Braun, eds, 'Food Security, Diversificaiton and Resource Management: Refocusing the Role of Agriculture'. Proceedings of the twenty-thrid international conference of agricultural economists.
- Kuczmarski, R. J., Flegal, K. M., Campbell, S. M. and Johnson, C. L. (1994), 'Increasing prevalence of overweight among us adults: the national health and nutrition examination surveys', *The Journal of the American Medical Association* **272**, 205-11.

- Lakdawalla, D. and Philipson, T. (2002), The growth of obesity and technological change: a theoretical and empirical examination, working paper 8946, National Bureau of Economic Research.
- Lin, J. Y. (1988), The household responsibility system in china's rural reform, in A. Maunier and A. Valdés, eds, 'Agriculture and Governments In An Interdependent World'. Proceedings of the 20th international conference of agricultural economists.
- Luo, Z. (2003a), *Age, cohort and year analysis in the socioeconomic determinants of adult BMI in the 1990s*, Michigan State University, Department of Economics.
- Luo, Z. (2003b), *Socioeconomic Determinants of Body Mass Index of Adult Chinese in the 1990s*, Michigan State University, Department of Economics.
- Ma, H. and Popkin, B. M. (1995), 'Income and food-consumption behaviour in china: a structural-shift analysis', *Food and Nutrition Bulletin* **16**(2), 155–65.
- McKenzie, D. J. (2002), Disentangling age, cohort, and time effects in the additive model, working paper 009, Stanford University Economics Department.
- NIH (1998), *Clinical Guidelines On The Identification, Evaluation, and Treatment of Overweight and Obesity in Adults*. NIH Publication No 98-4083.
- Paeratakul, S., Popkin, B. M., Ge, K., Adair, L. S. and Stevens, J. (1998), 'Changes in diet and physical activity affect the body mass index of chinese adults', *International Journal of Obesity* **22**, 424–31.
- Piazza, A. (1983), Trends in food and nutrient availability in china, 1950-81, Working Paper 60, World Bank, World Bank, Washington, DC.
- Pitt, M. M. (1983), 'Food preferences and nutrition in rural bangladesh', *The Review of Economics and Statistics* **65**(1), 105–14.
- Pitt, M. M. and Rosenzweig, M. R. (1985), 'Health and nutrient consumption across and within farm households', *The Review of Economics and Statistics* **67**, 212–23.
- Pitt, M. M., Rosenzweig, M. R. and Gibbons, D. M. (1993), 'The determinants and consequences of the placement of government programs in indonesia', *World Bank Economic Review* **7**(3), 319–48.
- Poleman, T. T. (1981), 'A reappraisal of the extent of world hunger', *Food Policy* **6**(4), 236–52.
- Popkin, B. M. (1999), 'Urbanization, lifestyle changes and the nutrition transition', *World Development* **27**(11), 1905–16.
- Popkin, B. M. (2002), 'The shift in stages of the nutrition transition in the developing world differs from past experiences!', *Public Health Nutrition* **5**(1A), 205–14.

- Popkin, B. M. and Doak, C. (1998), 'The obesity epidemic is a worldwide phenomenon', *Nutrition Reviews* **56**(4), 106–14.
- Popkin, B. M., Ge, K., Zhai, F., Guo, X., Ma, H. and Zohoori, N. (1993), 'The nutrition transition in china: a cross-sectional analysis', *European Journal of Clinical Nutrition* **47**, 333–46.
- Popkin, B. M., Paeratakul, S., Zhai, F. and Ge, K. (1995a), 'Dietary and environmental correlates of obesity in a population study in china', *Obesity Research* **3**(Suppl 2), S135–43.
- Popkin, B. M., Paeratakul, S., Zhai, F. and Ge, K. (1995b), 'A review of dietary and environmental correlates of obesity with emphasis on developing countries', *Obesity Research* **3**(Suppl 2), S145–53.
- Popkin, B. M., Paeratakul, S. and Zhai, K. G. F. (1995), 'Body weight patterns among the chinese: results from the 1989 and 1991 china health and nutrition surveys', *American Journal of Public Health* **85**(5), 690–4.
- Reddy, K. S. (2002), 'Cardiovascular diseases in the developing countries: dimensions, determinants, dynamics and directions for public health action', *Public Health Nutrition* **5**(1A), 231–7.
- Rosenzweig, M. R. and Schultz, T. P. (1983), 'Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight', *The Journal of Political Economy* **91**(5), 723–46.
- Ruhm, C. (2000), 'Are recessions good for your health?', *The Quarterly Journal of Economics* **115**(2), 617–50.
- Schultz, T. P. (1984), 'Studying the impact of household economic and community variables on child mortality', *Population and Development Review* **10**(suppl), 215–35.
- Sickles, R. C. and Taubman, P. (1997), *Mortality and morbidity among adults and the elderly*, Elsevier Science B. V.
- Smil, V. (1981), 'China's food: availability, requirements, composition, and prospects', *Food Policy* **6**(2), 67–77.
- Smith, J. P. (1998), 'Socioeconomic status and health', *The American Economic Review* **88**(2), 192–96.
- Smith, J. P. and Kington, R. (1997), 'Demographic and economic correlates of health in old age', *Demography* **34**(1), 159–70.
- Srinivasan, T. N. (1992), *Undernutrition: concepts, measurements, and policy implications*, Clarendon Press, Oxford, chapter 4, pp. 97–120.

- Srinivasan, T. N. (1994), 'Data base for development analysis: an overview', *Journal of Development Economics* **44**(1), 3–27.
- SSB (1988-1998), *China Statistical Yearbook*, State Statistics Bureau, China Statistical Publishing House.
- Stookey, J. D., Zhai, F., Zohoori, N. and Popkin, B. M. (2000), 'Nutrition of elderly people in china', *Asia Pacific Journal of Clinical Nutrition* **9**(4), 243–51.
- Strauss, J. (1984), 'Joint determination of food consumption and production in rural sierra leone, estimation of a household-firm model', *Journal of Development Economics* **14**(1-2), 77–103.
- Strauss, J., Beegle, K., Dwiyanto, A., Herawati, Y., Pattinasarany, D., Satriawan, E., Sikoki, B., Sukamdi and Witoelar, F. (2004), *Indonesian living standards three years after the crisis: evidence from the Indonesia Family Life Survey*, Institute of South East Asia Study, Singapore.
- Strauss, J. and Thomas, D. (1995), Human resources: empirical modeling of household and family decisions, in J. Behrman and T. N. Srinivasan, eds, 'Handbook of Development Economics', Vol. 3, Elsevier Science Publishers B.V.
- Strauss, J. and Thomas, D. (1996), 'Measurement and mismeasurement of social indicators', *The American Economic Review* **86**(2), 30–34.
- Strauss, J. and Thomas, D. (1998), 'Health, nutrition, and economic development', *JOURNAL OF ECONOMIC LITERATURE* **36**(2), 766–817.
- Subramanian, S. and Deaton, A. (1996), 'The demand for food and calories', *Journal of Political Economy* **104**(1), 133–62.
- Sundquist, J. and Johansson, S.-E. (1998), 'The influence of socioeconomic status, ethnicity and lifestyle on body mass index in a longitudinal study', *International Journal of Epidemiology* **27**, 57–63.
- Thomas, D., Lavy, V. and Strauss, J. (1996), 'Public policy and anthropometric outcomes in the Côte d'Ivoire', *Journal of Public Economics* **61**, 155–92.
- Wang, Q., Jensen, H. H. and Johnson, S. R. (1993), 'China's nutrient availability and sources, 1950-91', *Food Policy* pp. 403–13.
- Welch, F. (1970), 'Education in production', *Journal of Political Economy* **78**(1), 35–59.
- Whitewell, G. and Nicholas, S. (2001), 'Weight and welfare of australians, 1890-1940', *Australian Economic History Review* **41**(2), 159–75.
- Whitney, E. N. and Rolfes, S. R. (1996), *Understanding Nutrition*, 7 edn, West Publishing Company.

- Willett, W. (1998), *Nutritional Epidemiology*, New York: Oxford University Press.
- Wolpin, K. (1997), Determinants and consequences of the mortality and health of infants and children, in M. R. Rosenzweig and O. Stark, eds, 'Handbook of Population and Family Economics', Elsevier Science B. V.
- Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.
- World Development Indicators* (1999), World Bank.
- Wu, Y. (1997), 'China's health care sector in transition: resources, demand and reforms', *Health Policy* **39**, 137–52.
- Yaohui, Z. (1999), 'Leaving the countryside: rural-to-urban migration decisions in china', *American Economic Review* **89**(2), 281–86.
- Zhai, F., Fu, D., Du, S., Ge, K., Chen, C. and Popkin, B. M. (1996), 'Evaluation of the 24-hour individual recall method in china', *Food and Nutrition Bulletin* **17**(2), 154–61.
- Zhai, F., Fu, D., Du, S., Ge, K., Chen, C. and Popkin, B. M. (2002), 'What is china doing in policy-making to push back in the negative aspects of the nutrition transition', *Public Health Nutrition* **5**(1A), 269–73.