THE EFFECTS OF PHYSICAL ACTIVITY AND DIETARY INTAKE ON WEIGHT GAIN AMONG CHINESE MEN: ESTIMATION OF A DYNAMIC MODEL [1]

DRAFT. PLEASE DO NOT CITE

Shu Wen Ng

David K. Guilkey

Edward C. Norton

Barry M. Popkin

Authors Affliations:

Shu Wen Ng, Department of Health Policy and Management, University of North Carolina at Chapel Hill

David K. Guilkey, Department of Economics, University of North Carolina at Chapel Hill.

Edward C. Norton, Department of Health Management and Policy and Department of Economics, University of Michigan at Ann Arbor.

Barry M. Popkin, Department of Nutrition, University of North Carolina at Chapel Hill.

DRAFT. PLEASE DO NOT CITE

[1] This research uses data from China Health and Nutrition Survey (CHNS). We thank the Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention, Carolina Population Center, at the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, and R01-HD38700) and the Fogarty International Center for financial support for the CHNS data collection and analysis files, and for partial support of Ng and Popkin. None of the authors has conflict of interests of any type with respect to this manuscript. All errors are of the authors' alone.

ABSTRACT

To date, economic analysis on weight can only speculate on the differential effects of diet and physical activity. This paper considers the neo-classical economics of obesity model and uses time and spatially varying macro-level urbanization and price measures as instruments to correct for the endogenous and auto-correlated choices of diet, physical activity, smoking and drinking on weight. We apply a dynamic panel system GMM estimation model on longitudinal (1991–2006) data from China and found that among adult men in China, 30% of weight gain was due to declines in physical activity, while 20% was due to higher fat intake.

Keywords: Dynamic panel model, weight, physical activity, diet, urbanization, price, men, China

Corresponding author: Shu Wen Ng
University of North Carolina at Chapel Hill
CB # 8120 University Square
Chapel Hill, NC 27516-3997
shuwen@unc.edu / +1-919-966-1734

I. INTRODUCTION

In the last decade or so, economists have tried to understand the economics of obesity (or overweight), and have come up with a few theoretical models to do so. One of the main approaches is the neo-classical model (Lakdawalla & Philipson, 2006; Rashad & Grossman, 2004). This model is based on the idea of maximizing utility under a set of changing constraints due to changes in relative prices of goods, income, and time allocation. As such, most of the empirical work tends to look at variation in supply, demand, and prices of food or energy, income, state laws or the opportunity cost of time (Cawley, 2004; Chou et al., 2004; Drewnowski & Darmon, 2005; Kuchler et al., 2004; Lakdawalla et al., 2005; Schroeter et al., 2008; Sturm, 2004). Moreover, most of the studies have been done in the context of developed countries, particularly the United States, often using aggregate level data, or are not truly longitudinal (e.g., Lakdawalla and colleagues used a number of different data sets in order to piece together a panel dataset for their analyses). These do not allow for a proper empirical examination of the forces contributing to the long-run weight gain over time. Hence, it is unclear from the current literature what factors are the most important causes of overweight or obesity.

At first glance, modeling an individual's weight change appears straightforward: define the number of calories consumed and expended, and determine the resulting trends in weight gain. However, the layers of complexity in the relationships among physical activity, diet, and weight change are lost in a simplistic formulation of weight change (Moore, 2000).

For one, despite a consistent relationship between low physical activity, high caloric intake and weight in cross-sectional studies, there might be endogeniety -- it is possible that lower physical activity and increased sedentary behaviors among those who are heavy may be the consequence of being heavy (social stigma, excluded from sports, etc.). Similarly, it is possible that people consume more (as a coping mechanism) in reaction to being marginalized. Moreover, diet and activity combined can interact to affect weight (Astrup, 1999). Therefore, there is a need to untangle the possible reverse causalities and confounding often discussed in the diet, physical activity and weight relationships. Past studies typically only look at the relationships between physical activity and weight, and diet and weight separately, without considering the endogenous decisions of contemporaneous and lagged diet and physical activities on weight, and the serial correlations of these decisions over time (Klesges et al., 1992). This is problematic because if endogeneity does exist but is not corrected for, the result will be inconsistent parameter estimates. Meanwhile, if the correlation of weight over time is not controlled for, then the effect of past weight will tend to be over-estimated.

Second, even if one is able to properly remove reverse causality and confounding, there might also be a different effect of diet on weight depending on the type of food consumed, since calories from fat versus calories from carbohydrates or proteins might affect weight change differently (Miller et al., 1990; Tryon, 1987). For example, it has been found that people with low physical activity levels but high fat intake have slower metabolisms, which results in greater weight gain (Bray & Popkin, 1998), particularly for people in developing countries who might have experienced undernutrition during prenatal and postnatal growth (Frisancho, 2003; James & Ralph, 1999).

In this paper, we handle these complexities and determine the directions and relative importance of the diet, physical activity, and health behavior with regards to weight gain among adult men over a period of rapid economic growth in China by employing two strategies. First, we use the neo-classical model and explicitly include time and spatially varying macro-level factors such as urbanization and prices to be used as valid instruments to correct for the endogenous micro-level choices of diet, physical activity and other health behaviors to affect weight change over time. Second, we apply a dynamic panel system GMM model estimation model, which allows current weight to depend on prior weight and endogenous decisions about physical activity, diet and health behaviors and use statistical methods that control for the simultaneity problem and for temporal autocorrelation.

We estimate our dynamic panel model using a generalized method of moments (GMM) estimator developed by (Blundell & Bond, 1998) that exploits a large set of moment conditions and combines in a system, the regression in differences with the regression in levels. We provide a comparison of these results to those derived from usual reduced form and instrumental variable (IV) models to show how the failure to correct for simultaneity bias and autocorrelation can affect the findings.

To our knowledge, a dynamic panel analysis of these decisions that controls for endogeneity and autocorrelation in weight, diet, physical activity and health behaviors has not yet been done due to the requirement of good longitudinal data with at least two waves of data and large sample sizes. We will test these hypotheses by using six waves of the longitudinal China Health and Nutrition Surveys (CHNS). The data contain detailed individual data on anthropometrics, dietary consumption, energy expenditure, as well as time-varying community measures of urbanization and prices that can be used as instrumental variables for potentially endogenous variables.

China is an interesting country in which to carry out the research because its growing epidemic of overweight and obesity, risk markers for a large number of chronic diseases, will have severe consequences on its economic productivity and will become a significant healthcare burden. Currently, it is estimated that the cost of overweight and related diseases will be almost

9% of China's gross national product (GNP) by 2025 (Popkin et al., 2006). In addition, we focus on Chinese men in this study for two main reasons. First, men and women have very different biological processes and social or economic roles that can influence their weight gain rather diversely. In terms of biological processes, prior studies have found that men and women expend energy at different rates (Ferraro et al., 1992; Kennedy et al., 1997) and gain weight differently (Lovejoy & Sainsbury, 2009; Westerterp & Goran, 1997). Socio-economic roles mean that women often have the triple burden of work, children and domestic chores, which have competing effect on diet and physical activity choices and limit the variance observed in data. Second, there is better variance in the data for men because men are more likely to own technology like motorized vehicles as well as experience occupational change than women.

II. NEO-CLASSICAL MODEL OF THE DYNAMICS AND DETERMINANTS OF WEIGHT CHANGE

We use the standard neo-classical model on the economics of obesity (Cawley, 2004; Chou et al., 2004; Drewnowski & Darmon, 2005; Lakdawalla et al., 2005; Rashad & Grossman, 2004). It is useful to understanding the theory behind the model as it helps provide guidance on instruments and what factors might be endogenous.

II.A. Dynamics of weight change

We follow Lakdawalla et al. (2005), where an individual's utility in current period, t, depends on food consumption, F, other consumption, C, and current weight, W. We can write this as $U_t\ (F_t,\ C_t,\ W_t)$, where U rises in food consumption and other consumption, but is only monotonically increasing in weight if current weight is less than ideal weight, $\ddot{W}$, otherwise U declines in W. To summarize, the partial derivations of utility with respect to each argument are: $U_F \geq 0$; $U_C \geq 0$; $U_W \geq 0$ if $W \leq \ddot{W}$; and $U_W < 0$ if $W > \ddot{W}$.

In reality, food intake is multi-dimensional, comprising of calorics, the nutritional content of foods eaten, and the types of foods eaten. For simplicity, in this modeling section we only consider food intake as a scalar. We will assume that food consumption and other consumption are not substitutes in the sense that $U_{FC} \geq 0$. We also assume that there is no direct utility from activity, A. Instead, A is a determinant of weight and affects utility indirectly through its effect on weight. This is because it not clear if this effect is positive or negative since U_A could be > 0 for people who enjoy exercise, and < 0 for people who do not.

Let's consider an individual who manages weight according to a dynamic problem where his weight, W is the state variable. Weight is a capital stock that depreciates over time (where δ can be thought of as basal metabolism, because holding food intake and activity constant, there is some metabolic cost of living to the next period), and can be accumulated by food consumption, F, or decreased by activity, A. Moreover, an individuals' activity level depends on how developed a place he or she lives and works is, D, such that $A_t = A(D_t)$. Thus, an individual's weight at time t, depends on prior period's weight, food consumption and activity level:

(1) $$W_t = (1-\delta)W_{t-1} + g(F_{t-1}, A(D_{t-1})),$$

where $\delta < 1$ and g is continuous, concave, increasing in food consumption and decreasing in activity level ($g_F \geq 0$ and $g_A \leq 0$).

Here we see that an individual makes choices on F_{t-1} and A_{t-1} simultaneously. In addition, F_{t-1} and A_{t-1} affects W_t, and also is strongly related to F_t and A_t. Hence F_t and A_t will be correlated to W_t both via W_{t-1} and through F_{t-1} and A_{t-1}. This suggests that it is important to control for the simultaneity bias from diet and physical activities choices on weight, and the serial correlations of these decisions over time.

Moreover, individuals are subject to a budget constraint at each time period: $p_F F + p_c C \leq I$, where p_F and p_c are the prices of food and other consumption goods respectively. Standardizing by p_c, we can write the budget constraint as:

(2) $$C \leq I - p_F F$$

Calculations of the comparative statics[2] show that these will yield a steady-state in food consumption, activity and weight as long as the marginal utility of food is falling in weight[3]. The optimality condition will be:

(3) $$v'\{W + g[F_t, A(D_t)]\} = (p_F U_C - U_F)/g_F ,$$

which is that the marginal benefit of weight in the future equals the marginal cost of spending on weight change.

II.B. Steady State Determinants of Weight, Diet and Physical Activity

The steady state choices of weight, food consumption and physical activity can be determined by income, I, food prices, p_F, and urbanization, D, such that $W^*(I, p_F, D)$, $F^*(I, p_F, D)$, $A^*(I, p_F, D)$. If these factors are exogenous to weight, diet and physical activity, and vary spatially and over time, then they would make ideal instruments for use in correcting for the simultaneity bias and autocorrelation found in estimating weight dynamics.

Increases in income will initially raise weight, but at high levels of income, further increases could actually lower weight such that $U_{WC} > 0$ for the underweight but $U_{WC} < 0$ for the overweight (i.e., W_I^* can have an inverted U-shape). This is because an increase in income lowers the marginal cost of spending on weight gain (food consumption), and also affects the

---

[2] See the Appendix for complete calculations and details on the dynamics of weight change.

[3] It would be possible to incorporate addictive eating preferences, under which increases in weight raise the marginal utility of eating. This would result in multiple unstable equilibria in food, activity and weight. To focus on the core relationships among food, activity and weight, I will not consider the cases of addictions here.

marginal value of weight, *v'*. Income can also affect one's activity level, such that A is a function of job characteristics which are also reflected in earned income. This is like a substitution effect (i.e., the quantity of activity changes but the individual derives the same level of utility). Hence, the total effect of income on weight includes the direct income effect along with effect of earned income on activity levels, $dW^{*}/dI = W_I^{*} + W_A A_I (p_F, D, I)$. We assume that $W_A < 0$ since holding all else constant, increasing activity levels should lead to decreasing weight. For a developing country with a large rural population, such as in China, those who are poorer generally work in areas that require greater physical activity. Thus, we also assume that A_I <0. This means the total effect of income on weight will be positive when an individual is underweight, or when an individual is overweight and the negative direct effect of income on weight is less than the indirect effect of activity level on weight. The total effect of income on weight will only be negative when an individual is overweight and the negative direct effect of income on weight is greater than the indirect effect of activity level on weight.

Increasing the price of food, p_F, raises the marginal cost of spending on food, but the marginal benefit of weight tomorrow remains the same. An increase in the price of food increases the marginal "joy of eating", so food consumption decreases, so that $F_{pF}^{*}(p_F, D, I) < 0$. The decrease in food consumption will have a negative effect on weight, so that $W_{pF}^{*}(p_F, D, I) < 0$. Hence, prices are important determinants of weight change and are exogenous factors that need to be included in any structural modeling of weight change.

Community level urbanization, D if exogenous to individual choice assuming that people who move to other communities do not do so based primarily on these community-level characteristics. It can certainly affect prices of food, other consumption goods, and income, such that increased development will lower prices and raise incomes. Hence it can be thought of as an argument for p_F, and I. Using chain rule, the effect of urbanization on food consumption and activity levels are $F_D^* > 0$ and $A_D^* < 0$. The effect of urbanization on weight will depend on the relationships between income, prices and weight[4]. The direction of these effects should not be surprising. Urbanization would certainly lower physical activity at work due to shifts in labor market needs and job functions, access to technologies that aid work and domestic activities, and availability of motorized transportation, for example. Also, one would expect urbanization at the community level to reduce food prices such as through lowering transportation costs, and lessening the time costs involved in purchasing food.

III. EMPIRICAL MODELING

The dynamic empirical model that relates weight to its own lagged value along with lagged diet and lagged physical activity takes the following form:

(4) $$W_{it} = \alpha W_{i,t-1} + \beta F_{i,t-1} + \gamma A_{i,t-1} + \theta HB_{i,t-1} + \pi X_{it} + \eta_i + \mu_{ti},$$

where W_{it} denotes weight in the current wave t for individual i; $W_{i,t-1}$ denotes weight in the prior wave for individual i; $F_{i,t-1}$ denote lagged values of two dietary consumption variables: total caloric intake and energy from dietary fat; $A_{i,t-1}$ denote lagged total physical activities; $HB_{i,t-1}$ denote lagged smoking and drinking status; X_{it} denotes other control variables such as age, marital status, educational attainment, predicted household income and time dummies; α, β, γ, θ and π denote the vectors of coefficients for the explanatory respective variables; η_i denotes unobserved time-invariant individual characteristics, and μ_{it} denotes a time varying disturbance term.

[4] $W_D^* > 0$ if $dW/dI > 0$
> 0 if $dW/dI < 0$ and $(dp_F/dD)\cdot(dW/dp_F) > |(dI/dD)\cdot(dW/dI)|$
< 0 if $dW/dI < 0$ and $(dp_F/dD)\cdot(dW/dp_F) < |(dI/dD)\cdot(dW/dI)|$

We expect β_{kcal} *and* β_{efat} (the coefficients for lagged caloric intake and lagged energy from dietary fat) to be positively related to $W_{it.}$, and γ (the coefficient for lagged physical activity) to be negatively related to W_{it}. If we find that β_{kcal} , β_{efat} and γ to be statistically significant, then it we can determine the contribution of caloric intake, dietary fat intake and physical activity in determining weight gain and from that tell whether diet or physical activity are more important in affecting weight gain. The type of estimation method used depends on assumptions about:

1) The correlation between explanatory variables and η_i.
2) Autocorrelation: correlation in the time varying error terms over time (e.g., corr($\mu_{i,t-1}$, μ_{it}))
3) The type of correlation between the explanatory variables and μ_{it} , μ_{it-1} or μ_{it+1}

It is clear from the dynamic form of the stochastic model that at a minimum lagged weight will be correlated with η_i, the time invariant error term and it is also highly likely that there will be overlap in the set of unobserved fixed characteristics of the individuals that affect weight, diet, physical activity, smoking and drinking which will cause correlation between η_i and these variables as well. It is well known that first differencing will drop η_i along all time invariant observed variables from the model:

(5) $$\Delta W_{it} = \alpha \Delta W_{i,t-1} + \beta \Delta F_{i,t-1} + \gamma \Delta A_{i,t-1} + \theta \Delta HB_{i,t-1} + \pi \Delta X_{it} + \Delta \mu_{ti}$$

A frequently made assumption is that the time varying error is not correlated with the explanatory variables, which means that in differenced form diet, physical activity, smoking, and drinking will be uncorrelated with the error term in equation (5). However, it will still be the case that differenced weight will be correlated with the differenced error term. However, $W_{i,t-2}$ will be not and can be used as an instrument.

This instrumental variables estimation in differences tends to yield imprecise parameter estimates if α is large (Alonso-Borrego & Arellano, 1999; Blundell & Bond, 1998). An alternative (Blundell & Bond, 1998) is to estimate the model in levels with $\Delta W_{i,t-2}$ used as an instrument for $W_{i,t-1}$ in equation (4). This method, of course, must assume that there is no correlation between the other explanatory variables and either the time invariant or time varying error term. A more efficient method estimator, (Blundell & Bond, 1998) would jointly estimate equations (4) and (5) using a system generalized method of moments (GMM) approach.

As already noted, it is highly likely that there will be correlation between diet, physical activity, smoking, and drinking and the time invariant error and so instruments are needed for these variables in addition to lagged weight in equation (4). It is also possible that there will be correlation between these variables and the time varying error term, meaning that instruments may be needed for these variables even in differenced form in equation (5). Autocorrelation in the time varying error could also invalidate $W_{i,t-2}$ as an instrument in equation (5).

It is clear that separate instrument sets must be specified for equations (4) and (5) in the system GMM approach and we discuss these sets further below. Cameron and Trivedi (2005) provide a discussion of the large set of instruments that are potentially available in dynamic panel models and a series of papers provide information on efficient estimation strategies for these models (Arellano & Bond, 1998; Blundell & Bond, 2000; Blundell et al., 2000; Bond, 2002). Fortunately, our data set includes lagged measures of various dimensions of urbanizations and real price of consumption items that can be used to help provide identification, which will be discussed later.

We estimate robust standard errors using the two-step version of the Arellano-Bond system estimator with a finite-sample correction[5] (Windmeijer, 2005) using the *xtabond2* procedure in Stata (Roodman, 2003). We also perform two specification tests. First, we test for the presence of second-order autocorrelation in the differenced equation. Note that first-order autocorrelation in the differenced equation is expected and does not signify an improper model specification. Second, we test for the exogeneity of the instruments using Sargan-Hansen's *J*-test, which is robust to heteroskedasticity and autocorrelation, and is asymptotically distributed as χ^2 in the number of restrictions.

This dynamic panel approach has been used in financial/investment (Carstensen & Toubal, 2004; Horioka & Wan, 2006), environmental economics (Arbués et al., 2004), healthcare organization modeling (Brown et al., 2006; Mark et al., 2004) and in untangling the health-wealth relationship (Michaud & van Soest, 2008),where autocorrelation and endogeneity are of potential concern. For the empirical question of weight over time, the system GMM dynamic panel approach is ideal. This is the first paper to our knowledge to use it because it requires at least two consecutive waves of panel data (depending on the exact specification) and a large number of observations in each wave. We have six waves of 4,180 unique men, or 1,380 to 2,010 observations per wave. The results are also straightforward to interpret (in the same manner as with usual regression results), where a one-unit increase in the explanatory variable causes a coefficient unit change in weight.

Finally, we compare our two-step system GMM estimator to simple estimators to see how the results differ. First of all, we estimate a random effects model that does not control for

[5] The one-step version uses a weighted matrix that does not depend on estimated parameters. The two-step estimator may result in efficiency gains (smaller standard errors), but the asymptotical distribution approximations may be less reliable due to the dependence of the two-step weighted matrix on estimated parameters.

the correlation between the explanatory variables and the disturbance terms. We refer to this method as exogenous regressor model – we expect the results of this estimation to be badly biased for the reasons laid out above. Second, we apply an instrumental variables (IV) estimator to equation (4). We expect this estimator to provide consistent parameter estimates but this estimator should be less efficient than the two-step system GMM estimator.

IV. DATA

This paper used comprehensive longitudinal data from the six most recent waves (1991, 1993, 1997, 2000, 2004 and 2006) of the China Health and Nutrition Survey (CHNS) on all adults (18 to 55 years old) interviewed during any of the survey waves. The CHNS were conducted in nine diverse provinces (Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, and Shandong) of China, and contains detailed individual-level information on income, diet, health and demography for all members of sampled households as well as detailed community level data on infrastructure, public services and facilities. A multistage, random cluster process was used to draw the sample surveyed in each of the provinces. Counties in the nine provinces were stratified by income and a weighted sampling scheme was used to randomly select four counties in each province. Villages and townships within the counties and urban and suburban neighborhoods within the cities were selected randomly into primary sampling units (PSUs). The same households were surveyed over time as best possible and newly formed households began to be surveyed in 1993. After we limited the data to men between 18 and 55 years old and who were not disabled during a particular wave there were 16,883 person-wave observations made up by 10,935 men.

Of these, only 4,180 men had at least two consecutive waves of data, making up 8,645 observations (Table I). There is concern as to the loss of many observations over time due to this criterion. Some of the loss of observations was due to the fact that those whose first survey was conducted in 2006 were not included in the analytic sample (850 observations made up of 643 men). In addition, Liaoning province was dropped from the survey and replaced by Heilongjiang province in 1997 (Heilongjiang was kept in henceforth) due to flooding in Liaoning that year. This meant that observations from adult men in Liaoning who were first collected in 1993 would not have made it to the analytic sample due to the missing data for 1997. Also, the 1991 and 1993 Heilongjiang sample did not exist.

Beyond these factors, there was also loss to follow up. To ensure that attrition was not systematic, we ran a Heckman selection model (Heckman, 1979) on the individual level data. This two-stage estimation was based on whether an individual had two or more consecutive waves of data using observed exogenous characteristics[6] in their first wave, and the last observed weight of individuals using exogenous characteristics from the last observed wave among those with two or more consecutive waves of data. The Wald test of independence in the errors between these two stages produced a χ^2-statistic of 1.63, meaning that we cannot reject the null that there is no correlation between the errors of these two equations (i.e., selection is not a problem). We also ran a Hausman specification test (Hausman, 1978) between the coefficients from the second equation and from basic OLS and found that we could not reject the null hypothesis that the difference in coefficients are not systematic ($\chi^2(35) = 1.43$).

[6] These exogenous baseline and last observed characteristics were: community urbanization measures, prices, province, time, age, marital status, education attainment and predicted household income.

IV.A. Dependent Variable

Anthropometric data was collected by trained health workers during a comprehensive physical exam at a local clinic or at the respondent's home. Height was measured without shoes to the nearest 0.2 cm using a portable stadiometer; weight was measured without shoes and in light clothing to the nearest 0.1 kg on a calibrated beam scale. Figure I shows that from 1991 to 2006, both weight and body mass index (BMI= weight in kg/(height in m)^2) rose significantly among adult men in China. In this analysis, we used weight as the dependent variable (but control for height) because it is easier to interpret the results in terms of weight, and because height does not change much within an adult (18 to 55 year old) population[7].

IV.B. Key Explanatory Variables

The key explanatory variables were lagged total physical activities, total caloric intake, the proportion of energy from fat, smoking and drinking status.

Total physical activity was based on self- reported information on activity levels and times spent for up to two occupations, and time spent on four types of domestic activities (buying food, preparing food, doing laundry and childcare). These were combined with specific metabolic equivalent (MET) values based on the Compendium of Physical Activities (Ainsworth et al., 2000) to derive MET-hours per week to account for both intensity of activities and time spent on activities. A unit of MET, is defined as the ratio of a person's working metabolic rate relative to his/her resting (basal) metabolic rate. There is additional information about leisure activities and travel activities, which was only available in the last four waves of the CHNS. However, limiting the analysis to only these last four waves would have severely compromised sample sizes, so only activities from occupations and domestic chores were included. Moreover, among men in China,

[7] We also will run the two-step system GMM with height as an endogenous variable.

these two domains made up the bulk of physical activities based on 1997-2006 data). Additional information on the creation of the physical activity measures are in a recent paper by Ng (2008). In our analytic sample, physical activity levels among Chinese men fell significantly by 37 percentage points in a span of 15 years (Figure II).

Detailed consumption data at both the household and individual level were collected over three consecutive days, which were randomly allocated from Monday to Sunday in order to determine average daily caloric intake from various macronutrient sources for each individual. Household food consumption was determined by examining changes in inventory from the beginning to the end of each day. Individual dietary intake for the same three consecutive days was surveyed all individuals from 1991 onwards based on daily self-reported 24-hour recalls on all food consumed away from home and at-home. The collection of both household and individual dietary intake allowed for quality checks. 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 (Wang et al., 2000).

The 1991 Food Composition Table (FCT) for China was utilized to calculate macronutrient intake values for the dietary data of 2000 and previous years (Institute of Nutrition and Food Hygiene, 1991). The UNC-CH group has worked with the National Institute of Nutrition and Food Safety to update and improve this FCT, which represents a significant advancement over the earlier China FCT, due to higher quality chemical analyses, improved techniques of developing average nutrient values for foods whose nutrient value varies over the country in a geographic context, and updating for new foods at each wave. A new 2000 version of the FCT (Institute of Nutrition and Food Hygiene, 2002) was used for 2004 and 2006 surveys, and update for new foods each year by the UNC-CH team.

We decided to use measure of proportion of energy from dietary fat (%) and total caloric intake because they are able to better capture the role of dietary fat in explaining weight gain. From the consumption data, we were able to get individual level measures of caloric fat, carbohydrate, protein and alcohol intake to allow us to determine these measures. Figure II shows that over the 15 year period, total caloric intake fell by about 17%, but the proportion of energy from dietary fat rose by 6% points.

Dummy variables for being a smoker or drinker were included in the analyses as controls because past research has shown the smoking is negatively associated with caloric intake and weight gain as nicotine increases energy expenditure and could reduce appetite in the short term (Hofstetter et al., 1986; Williamson et al., 1991), but heavy smoking can increases insulin resistance and is associated with central fat accumulation (Chiolero et al., 2008; Xu et al., 2007). The relationships between drinking and weight gain are not so clear (Hu et al., 2000; Maskarinec et al., 1998), as drinking might increase total caloric intake, but can also have a protective effect on weight gain or loss (Liu et al., 1994). The CHNS also asked respondents whether they drank any beer or other alcohol beverage in the past year; and if they are current smokers. In this sample, the smoking and drinking prevalence among Chinese men declined with age (Table II), likely due to the both the aging effect and mortality effect (smokers and drinkers might have higher mortality rates than non-smokers and non-drinkers). Other important explanatory variables we included were: a) lagged weight (based on weight in the prior wave) since weight is a stock concept that accumulates or de-cumulates over time; and b) height in centimeters, since weight will vary depending on one's height. These variables were discussed earlier.

Also, as mentioned before, a problem with these key explanatory variables is the endogeneity of decisions concerning dietary intake, physical activity and health behaviors with relation to weight due to the introduction of lagged weight as an explanatory variable. In order to correct for the possible endogeneity, we need to identify instrumental variables, and conduct specification tests to ensure their validity.

IV.C. Potential Instruments

Potential instruments are time-varying and arguably exogenous dimensions of urbanization and prices of food items of each community. We conducted specification tests on various sets of these variables to determine the final set of instruments for use in both the instrumental variables estimation and the dynamic panel estimation.

We used ten community-level measures of various dimensions of urbanization: population, density, market accessibility, economic wellbeing, transportation, communications, educational institutions, health facilities, sanitation and housing infrastructures. These reflect changes in the various dimensions of urbanization over time and reflect the environment in which people function. Each of these dimensions was given a score from zero to ten and was comprised of data collected from local area administrators or official records. Additional detail as to how these scores were created and their distributions over time can be found in Ng (2008).

The CHNS community-level measures of urbanization have also been previously used in papers by Monda et al. (2007), and Zimmer et al. (2007). Figure III shows that over time, the communities on average had improvements in these dimensions. Note that urbanization was not uniform across communities, with some communities experiencing declines in certain dimensions even though in general the average community might have seen improvements. China's household registration (*hukou*) system and the longitudinal nature of the CHNS data ensure that selection into communities and inclusion in the data was as independent of individual or household choices and behavior as best possible.

Prices may affect weight gain via consumption of various types of food items. We included prices of food items that may be particularly important in the context of China, such as rice, flour, pork and oil. Community price surveys conducted on a set of sample stores and markets were used to provide price data[8]. Variations in food prices across communities are due to both supply and demand side factors. On the supply side, agricultural production, transportation, marketing and distribution costs, imports of specific foods, and availability of substitutes and complements can affect prices across communities. On the demand side, preferences or food fads may vary by communities. Most price changes in China are driven by supply factors and exogenous economic decisions made at the provincial level by price commissions and other macroeconomic government decisions; hence, they can be considered exogenous variables that vary greatly over time and across communities as shown in the lack of a clear pattern in price changes among the ten most populous communities in the CHNS in Figure IV. In addition, there are also variations in inflation, measured by the Consumer Price Indices (CPI)[9] across communities. Price and income variables were deflated by these.

[8] There were three sources of price information for a representative basket of goods. These include state store prices, free market prices collected from visits to stores in the communities surveyed, and authority price records published by the State Statistical Bureau (SSB) of China, which provides the provincial average. The state store prices were no longer used after the 1991–1992 price reform in China. Therefore, in almost all situations, the free market prices will be used as the basis, except when the goods studied were not sold in the free market, in which case, prices from the state stores will be used, followed by SSB recorded prices if the other two sources do not have the information. Farmers both produce and consume food, which adds complexity to the price issue. However, we would argue that the free market prices for the food can be seen as the opportunity cost of consuming instead of selling the produce. Hence, using free market prices (when available) is appropriate.

[9] There is no published absolute consumer price index for China that provides a way to compare provinces or urban and rural areas. Rather, the State Statistical Bureau (SSB) publishes annually a consumer price index ratio that shows for urban and rural areas in each province the shift in the cost of living. Thus, the CHNS team created its own cost of living index by using a consumer goods basket of 57 items created by the government and urban price data. The CHNS urban and rural price data were then used to create a ratio of urban and rural costs for elements of this consumer goods basket. These were used to create costs of the consumer basket for each time period for urban and rural areas in each province in the CHNS. Lastly, they set China's food costs for urban Liaoning province for 2006 equal to 1.00 (or 100%) and made all other prices relative to this (CPC, online).

VI.D. Testing the validity of instruments

We tested the null hypothesis that there is exogeneity by conducting a Hausman test between the models assuming exogeneity and the instrumental variables model. Results from a Hausman test showed that we can reject the null hypothesis that there is exogeneity $\chi^2(21) = 52.76$. This suggests that we should use instrumental variables in the estimation. Otherwise, the results will be biased and inconsistent (Hausman, 1978). Valid instrumental variables need to satisfy two criteria: they must be correlated with the included endogenous variables and orthogonal to the error process.

The former condition may be readily tested by examining the first-stage regression results from the IV model, which used lagged values of the instruments and other control variables as explanatory variables for the endogenous variables of lagged weight, physical activity, dietary intake, smoking and drinking. Table III presents the results. We found that all but one of the lagged community urbanization measures was statistically predictive of physical activity in the prior wave. In particular, the scores for educational institutions, sanitation, economic wellbeing and housing infrastructures were highly associated with declines in physical activity. The community urbanization variables were also highly predictive of the dietary intake outcomes.

The lagged community price variables were most predictive of lagged weight and energy from fat. There are a few interesting results of note. For example, two key drivers of dietary change in China were particularly sensitive to price changes. A one percentage point increase in the price of pork appears to be related to a 1.36 percentage point decrease in energy from fat, and a one percentage point increase in the price of oil appears to be related to a 1 percentage point decrease in energy from fat. For endogenous variables of lagged weight, physical activity, total caloric intake and energy from fat, it appears that the ten community urbanization measures and eight price variables satisfied the requirement that these instruments are correlated with them. In general, the models did not perform as well for the health behaviors, particularly smoking. The relatively poor results for smoking are not too surprising since we only had dichotomous variables for whether an individual was a current smoker or not. Since smoking is an addictive habit (or lifetime choice), it is unlikely that there is much variation over time to one's smoking status once an individual picks up smoking.

The latter criteria (the instruments need to be independent from an unobservable error process) can be ascertained by using tests of overidentification since there are more potential instruments then there are endogenous variables. In GMM models such as what is used in the dynamic panel approach, a Hansen's *J*-statistic is used to test for this. The *J*-statistic follows a χ^2 distribution with degrees of freedom equal to the number of over identifying restrictions rather than the total number of moment conditions. A rejection of the null hypothesis can imply that the instruments do not satisfy the orthogonality conditions (either because they are not truly exogenous, or because they are being incorrectly excluded from the regression), or that the model specification is incorrect. In standard IV models, the Sargan statistic is calculated instead. The Sargan's statistic is a special case of Hansen's *J*-statistic, which uses an estimate of the error variance from the IV regression estimated with the full set of over identifying restrictions, and will generate a consistent estimator of the error variance under the null of instrument validity. We found that the Sargan test of over identification cannot be rejected (χ^2 (12) = 11.57), meaning that the set of instruments appear to satisfy the requirement that the instruments be independent from unobserved error (Table IV, column 2).

V. RESULTS

V.A. Exogenous Regressors

Our results show that the estimates when assuming exogeneity are indeed lower for caloric intake, energy from fat and caloric intake compared to the respective estimates from the IV and the GMM system approaches (Table IV, column 1). This suggests that these results are biased and inconsistent due to endogeneity. Moreover, the coefficient estimate for lagged weight is higher in this model compared to the IV and GMM system approach, suggesting that autocorrelation is also a problem.

V.B. Instrumental Variables

As expected, we found that the coefficients for the GMM system model (Table IV, column 3) and the IV model (Table IV, column 2) be closer than without correcting for the endogeneity, but the standard errors were much smaller for the dynamic panel model because the GMM system estimator is asymptotically more efficient than the IV estimator.

V.C. Two-step system GMM

The dynamic panel two-step procedure combines in a system GMM, a regression in differences over time, and a regression in levels (i.e., within the same wave). Recall that the consistency of the GMM estimation relies on the autocorrelation of the residuals and the validity of the instrumental variables. The *xtabond2* procedure in Stata (Roodman, 2003) performs validation tests for these. In our estimation, the rejection of the presence of a second-order autocorrelation (i.e., the AR(2) *z*-statistic is not significant) satisfies the first criterion, and the rejection of the *J*-test of overidentification satisfies the second criterion (Table IV, column 3).

The difference-in-Sargan test allows us to test a subset of the original set of orthogonality conditions by computing the difference in the two Sargan statistics from the entire set of overidentifying restrictions and from the model using a smaller set of restrictions. We see that the Hansen test excluding the set of instruments is χ^2 (37) = 41.28. So, the difference-in-Sargan test of the set of 18 instruments is χ^2 (18) = 20.72. The fact that we can reject the difference-in-Sargan test of exogeneity supports the validity of these instruments.

The coefficient estimates from the GMM system dynamic panel estimation are interpreted as in a standard linear model. We found that unsurprisingly, height is positively related to weight ($p<0.05$), while prior weight is positively related to current weight ($p<0.01$). An increase of 10 MET-hours/week of physical activity in the prior wave is associated with a weight loss of 0.03kg ($p<0.1$), a larger coefficient than what was found in the reduced-form. Also, a one percentage point increase in energy from dietary fat in the prior wave was associated with a 0.05 kg weight gain [10].

We can also tell from these coefficients and the noted change in physical activity levels and dietary fat intake over the 1991 and 2006 period, how each of these factors may have contributed to weight gain on average. Table V shows the breakdown of these figures based on results from this analysis. 30 percent of the increase in weight among adult Chinese men was due to declines in physical activity, while 20 percent was due to increases in fat in their diets, and the remaining 50 percent was due to other factors.

VI. DISCUSSION

It is critical to combine both measures of physical activity and dietary intake in trying to understand the dynamics of these on weight. This paper uses a dynamic panel system GMM estimation model and adds to the structural understanding of relationships among macro-level

[10] When height was considered an endogenous variable in the two-step system GMM model, the results were virtually the same. Results are not presented but available upon request.

factors on micro-level behavior instead of the typical reduced-form modeling approach. We found that declines in physical activities and increases in fat as a proportion of people's diet are positively associated with weight gain among adult men in China. Of these two factors, the declines in physical activity seem to a larger contributor to weight gain, although dietary fat intake is also important.

Physical activity has been found to be a critical factor in body weight regulation in lean and obese individuals due to its protective role over time through both direct energy expenditure, improved physical fitness and resultant metabolic effects on lipid mobilization and oxidation and biochemical changes in the muscle fiber that contribute to improved regulation of body weight (Saris, 1998). Previous work have also hinted that physical activity may be a more successful strategy than dietary approaches to weight loss and maintenance among men (King et al., 1989).

VI.A. Policy implications

The findings from this paper highlight the importance of promoting physical activities in order to both reduce weight gain. We do not, however, know if there are effective programmatic and policy options in China for increasing physical activity that will work to reduce the prevalence of overweight and obesity, risk markers for a large number of chronic diseases. Indeed, lack of physical activity has shown to be significantly associated with increased cardio-vascular problems (Hong et al., 1994; Sundquist et al., 2005), type II diabetes (Hu et al., 1999; Meisinger et al., 2005), incidence of stroke (West et al., 2006), cholesterol levels (Durstine et al., 2001) and mortality from cardio-vascular diseases, cancer and all causes (Hu et al., 2005). These diseases will create new health and financial challenges for developing countries such as China.

In China, declines in physical activity have occurred mostly at work, therefore intervention strategies to increase physical activity levels at the workplace are one possible strategy (Bell et al., 2001). To increase physical activity via active leisure and travel, activities such as walking and bicycling can be promoted by designing built environments that are safe and conducive for such transit or exercise modes (Bell et al., 2002; Forsyth et al., 2008; Nagel et al., 2008). It is also possible that improved access to public transit can help promote walking (Rodriguez et al., 2008) or bicycling – a diminishing activity in urban and rural China. Policies in the form of higher taxes on automobiles, lower entry fees to parks and government run health facilities can also help promote physical activity. Moreover, disincentives for automobile ownership can discourage motorized transportation and help reduce air pollution and provide more pleasant environments for outdoor exercise. These policies can be province or community specific depending on geographical, cultural and socio-economic circumstances.

The other side of the weight equilibrium has to do with dietary intake. In particular, there is some controversy regarding the role of fat intake on weight gain or the prevalence of obesity (Bray & Popkin, 1998; Willett, 1998). The debate concerns whether it is fat itself that increases weight or, that it is the fact that fat per gram is more energy dense, which is why both epidemiological studies and animal studies that also control for total energy intake find little significance of fat intake (Bray et al., 2004) and a recent study randomized controlled study found that it is total caloric change that affects weight dynamics and a lower-fat diet works as well as others to reduce weight (Sacks et al., 2009). We found that the proportion of energy from fat was important, but did not find a significant relationship between total caloric intake and weight gain, which seems to suggest that the latter may be the case in this particular population. This may be due to greater accuracy in measuring the structure of dietary intake than total caloric intake or that edible oils and fatty pork have higher price elasticities and are major drivers of the shifts in the composition of the diet in China (Ng, Zhai et al., 2008; Popkin, 2007). Nonetheless, regardless of whether it is the excess calories or the chemical nature of fat, it has been found that lower fat diets have been able to prevent overweight/obesity, hypertension, high total serum cholesterol, high serum triglycerides and high low density lipoprotein cholesterol (LDL-C) (Chen et al., 2008). Certainly, fat intake is only one environmental factor that affects the genetic expression of overweight and obesity. There is a need to tackle this growing epidemic from various angles, creating policies that target both physical activity and diets.

VI.B. Contributions

This paper addresses the questions of the roles of physical activity and dietary intake on weight, while controlling for the complex relationships among these variables by using a dynamic panel, using time-varying macro-level factors such as urbanization measures and prices as instruments. The dynamic panel estimates are unbiased, asymptotically efficient, and allows for autocorrelation in the error terms. Hence, the GMM system dynamic panel approach is ideal for the empirical model in question. However, it requires at least two waves and data, a large sample size, detailed individual data on anthropometrics, dietary intake, measures of physical activity, and availability of valid instruments. To our knowledge this is the first paper that has used this estimation strategy in the context of weight.

VI.C. Limitations and Future Direction

While this paper has provided insights into the dynamics among physical activity, diet and weight, there are a number of limitations with this analysis. First, this paper focused on a specific population— adult (18-55 year old) men in China, which limits its applicability to the general population. While gender disparities would be interesting to uncover, the CHNS data did not contain unmeasured predictors such as metabolic rate that are more operative for women. In addition, women often have the triple burden of work, children and domestic chores, which have competing effect on diet and physical activity choices and limits the variance observed in data. Second, there is better variance in the data for men because men are more likely to own technology like motorized vehicles as well as experience occupational change than women (Bell et al., 2002).

Second, we only looked at weight as the dependent variable. Weight in and of itself is not necessarily a good indicator of health status, since for someone who is underweight, weight gain may actually be a health benefit. Diagnoses of chronic diseases may be more appropriate outcomes to use. However, in order to conduct analysis on health outcomes will first require a good understanding of the dynamic among physical activity, diet and weight.

There are some data limitations that may compromise our findings. For example, the use of MET-hours per week to quantify energy expenditure does not take into account individual differences that may alter the energy cost of movements (such as basal metabolic rates). Nonetheless, this approach is the best available way to systematically apply average energy cost estimates in self-reported measures (Matthews, 2002). Moreover, we included multiple sources of occupational and domestic activity to allow a more complete assessment of physical activity.

Technological changes related to purchases of home assets, such as refrigerators, rice cookers, microwaves, vacuum cleaners and washing machines could be examined as other endogenous determinants of both diet and physical activity change.

Lastly, this analysis was unable to directly estimate the interacted effect of physical activity and dietary intake. However, it does so implicitly in the dynamic panel system GMM estimation approach by including lagged weight with controls for endogeneity in the physical activity and diet variables that affect weight.

Future work should consider looking at women, or study gender disparities in the dynamics of weight gain, and perhaps to parse out the various biological processes and social or economic roles that can influence their weight gain differently. Future work should also be considered regarding joint decisions about time and energy allocation among household members, instead of just considering individuals alone.

APPENDIX: DYNAMICS OF WEIGHT CHANGE

An individual's utility in current period, t, depends on food consumption, F, other consumption, C, and current weight, W. We can write this as $U_t\ (F_t,\ C_t,\ W_t)$, where U rises in food consumption and other consumption, but is only monotonically increasing in weight if current weight is less than ideal weight, $\ddot{W}$, otherwise U declines in W. Notationally, $U_F \geq 0$; $U_C \geq 0$; $U_W \geq 0$ if $W \leq \ddot{W}$; and $U_W < 0$ if $W > \ddot{W}$.

In reality, food intake is multi-dimensional, comprising of calorics, the nutritional content of foods eaten, and the types of foods eaten. For simplicity, in this modeling section we only consider food intake as a scalar. We will assume that food consumption and other consumption are not substitutes in the sense that $U_{FC} \geq 0$. We also assume that there is no direct utility from activity, A. Instead, A is a determinant of weight and affects utility indirectly through its effect on weight. This is because it not clear if this effect is positive or negative since U_A could be > 0 for people who enjoy exercise, and < 0 for people who do not.

Let's consider an individual who manages weight according to a dynamic problem where his weight, W is the state variable. Weight is a capital stock that depreciates over time (where δ can be thought of as basal metabolism, because holding food intake and activity constant, there is some metabolic cost of living to the next period), and can be accumulated by food consumption, F, or decreased by activity, A. Moreover, an individuals' activity level depends on how developed a place he or she lives and works is, D, such that $A_t = A(D_t)$. Thus, an individual's weight at time t, depends on prior period's weight, food consumption and activity level:

(A1) $$W_t = (1-\delta)W_{t-1} + g(F_{t-1}, A(D_{t-1})),$$

where $\delta < 1$ and g is continuous, concave, increasing in food consumption and decreasing in activity level ($g_F \geq 0$ and $g_A \leq 0$). This is also known as the transition equation.

Over multiple time periods, an individual's value function (or lifetime-indirect utility) depends on the current period's utility and the value function from future time periods, such that:

(A2) $$v(W_t) = \max_{F,C,W}\{U_t(F_t, C_t, W_t) + \beta v(W_{t+1})\},$$

where β is the discount factor. This is subject to the transition equation mentioned above, and a budget constraint: $p_F F + p_c C \leq I$, where p_F, and p_c are the prices of food and other consumption goods respectively. Standardizing by p_c, we can write the budget constraint as:

(A3) $$C \leq I - p_F F.$$

Combining Eq (A2) and Eq (A3), and taking the first order conditions with respect to F_t and C_t, and setting them to zero so that one is maximizing their utility, gives:

(A4) $$U_F(F_t, W_t) + \beta v'(W_{t+1}) g_F = U_C[F_t, (I_t - p_F F_t), W_t]$$

That is: Marginal utility of eating plus discounted marginal utility of weight in future period due to eating equals the marginal utility of consuming other goods.

Taking first order conditions of Eq (A2) with respect to W, we can get the envelope theorem:

(A5) $$v'(W_t) = U_W[F_t, (I_t - p_F F), W_t] + \beta(1-\delta)v'(W_{t+1}),$$

which shows that the long run marginal value of weight is equal to the marginal utility of weight in the current period plus the discounted marginal utility of weight.

These will yield a steady-state in food consumption, activity and weight as long as the marginal utility of food consumption is falling in weight, and that the marginal utility of activity is rising in weight. Rewriting the optimality condition,

(A6) $$v'\{W + g[F_t, A(D_t)]\} = \left[(p_F U_C - U_F)/g_F\right],$$

which is the marginal benefit of weight in the future equaling the marginal cost of spending on weight change.

References

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., et al. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc, 32*(9 Suppl), S498-504.

Alonso-Borrego, C., & Arellano, M. (1999). Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data. *Journal of Business & Economic Statistics, 17*(1), 36-49.

Arbués, F., Barberán, R., & Villanúa, I. (2004). Price impact on urban residential water demand: A dynamic panel data approach. *Water Resrouce Research, 40*(W11402).

Arellano, M., & Bond, S. (1998). *Dynamioc Panel Data Estimation Using DPD98 for Gauss: A Guide for Users*.Unpublished manuscript, London.

Astrup, A. (1999). Macronutrient balances and obesity: the role of diet and physical activity. *Public Health Nutr, 2*, 341-347.

Bell, A. C., Ge, K., & Popkin, B. M. (2001). Weight gain and its predictors in Chinese adults. *Int J Obes Relat Metab Disord, 25*(7), 1079-1086.

Bell, A. C., Ge, K., & Popkin, B. M. (2002). The road to obesity or the path to prevention: motorized transportation and obesity in China. *Obes Res, 10*(4), 277-283.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115-143.

Blundell, R., & Bond, S. (2000). GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews, 19*(3), 321 - 340.

Blundell, R., Bond, S., & Windmeijer, F. (2000). *Estimation in dynamic panel data models: improving on the performance of the standard GMM estimator*: Institute for Fiscal Studies.

Bond, S. R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal, 1*(2), 141.

Bray, G. A., Paeratakul, S., & Popkin, B. M. (2004). Dietary fat and obesity: a review of animal, clinical and epidemiological studies. *Physiol Behav, 83*(4), 549-555.

Bray, G. A., & Popkin, B. M. (1998). Dietary fat intake does affect obesity! *Am. J. Clinical Nutrition, 68*(6), 1157-1173.

Brown, T. T., Coffman, J. M., Quinn, B. C., Scheffler, R. M., & Schwalm, D. D. (2006). Do Physicians Always Flee from HMOs? New Results Using Dynamic Panel Estimation Methods. *Health Services Research, 41*(2), 357-373.

Cameron, A. C., & Trivedi, P., K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Carstensen, K., & Toubal, F. (2004). Foreign direct investment in Central and Eastern European countries: a dynamic panel analysis. *Journal of Comparative Economics, 32*(1), 3-22.

Cawley, J. (2004). An economic framework for understanding physical activity and eating behaviors. *American Journal of Preventive Medicine, 27*(3), 117-125.

Chen, C. M., Zhao, W., Yang, Z., Zhai, Y., Wu, Y., & Kong, L. (2008). The role of dietary factors in chronic disease control in China. *Obes Rev, 9 Suppl 1*, 100-103.

Chiolero, A., Faeh, D., Paccaud, F., & Cornuz, J. (2008). Consequences of smoking for body weight, body fat distribution, and insulin resistance. *Am J Clin Nutr, 87*(4), 801-809.

Chou, S. Y., Grossman, M., & Saffer, H. (2004). An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *Journal of Health Economics, 23*(3), 565-587.

CPC. (online). Creation of our CHNS Consumer Price Index for equating Income in the CHNS. 2006, from http://www.cpc.unc.edu/projects/china/data/documentation/income/cprindex.0800.pdf

Drewnowski, A., & Darmon, N. (2005). The economics of obesity: dietary energy density and energy cost. *American Journal of Clinical Nutrition, 82*(1), 265s-273s.

Durstine, J. L., Grandjean, P. W., Davis, P. G., Ferguson, M. A., Alderson, N. L., & DuBose, K. D. (2001). Blood lipid and lipoprotein adaptations to exercise: a quantitative analysis. *Sports Med, 31*(15), 1033-1062.

Ferraro, R., Lillioja, S., Fontvieille, A. M. R., R., Bogardus, C., & Ravussin, E. (1992). Lower sedentary metabolic rate in women compared with men. *J Clin Invest 90*(3), 780-784.

Forsyth, A., Hearst, M., Oakes, J. M., & Schmitz, K. H. (2008). Design and Destinations: Factors Influencing Walking and Total Physical Activity. *Urban Stud, 45*(9), 1973-1996.

Frisancho, A. R. (2003). Reduced rate of fat oxidation: A metabolic pathway to obesity in the developing nations. *American Journal of Human Biology, 15*(4), 522-532.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica, 46*(6), 1251-1271.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161.

Hofstetter, A., Schutz, Y., Jequier, E., & Wahren, J. (1986). Increased 24-hour energy expenditure in cigarette smokers. *N Engl J Med, 314*(2), 79-82.

Hong, Y., Bots, M. L., Pan, X., Wang, H., Jing, H., Hofman, A., et al. (1994). Physical Activity and Cardiovascular Risk Factors in Rural Shanghai, China. *Int. J. Epidemiol., 23*(6), 1154-1158.

Horioka, C. Y.-J., & Wan, J.-M. (2006). *The Determinants of Household Saving in China: A Dynamic Panel Analysis of Provincial Data*: National Bureau of Economic Research, Inc.

Hu, F. B., Sigal, R. J., Rich-Edwards, J. W., Colditz, G. A., Solomon, C. G., Willett, W. C., et al. (1999). Walking compared with vigorous physical activity and risk of type 2 diabetes in women: a prospective study. *JAMA, %20;282*, 1433 - 1439.

Hu, F. B., Wang, B., Chen, C., Jin, Y., Yang, J., Stampfer, M. J., et al. (2000). Body Mass Index and Cardiovascular Risk Factors in a Rural Chinese Population. *Am. J. Epidemiol., 151*(1), 88-97.

Hu, G., Tuomilehto, J., Silventoinen, K., Barengo, N. C., Peltonen, M., & Jousilahti, P. (2005). The effects of physical activity and body mass index on cardiovascular, cancer and all-cause mortality among 47 212 middle-aged Finnish men and women. *Int J Obes (Lond), 29*, 894 - 902.

Institute of Nutrition and Food Hygiene. (1991). *Chinese Food Composition Tables 1991* Beijing, PRC: People's Health Publishing House

Institute of Nutrition and Food Hygiene. (2002). *China Food Composition Table 2002*. Beijing, PRC: Beijing Medical University Publishing House.

James, W. P., & Ralph, A. (1999). New understanding in obesity research. *Proc Nutr Soc, 58*(2), 385-393.

Kennedy, A., Gettys, T. W., Watson, P., Wallace, P., Ganaway, E., Pan, Q., et al. (1997). The Metabolic Significance of Leptin in Humans: Gender-Based Differences in Relationship to Adiposity, Insulin Sensitivity, and Energy Expenditure. *J Clin Endocrinol Metab, 82*(4), 1293-1300.

King, A. C., Frey-Hewitt, B., Dreon, D. M., & Wood, P. D. (1989). Diet vs exercise in weight maintenance. The effects of minimal intervention strategies on long-term outcomes in men. *Arch Intern Med, 149*(12), 2741-2746.

Klesges, R. C., Klesges, L. M., Haddock, C. K., & Eck, L. H. (1992). A longitudinal analysis of the impact of dietary intake and physical activity on weight change in adults. *Am J Clin Nutr, 55*(4), 818-822.

Kuchler, F., Tegene, A., Harris, J. M., & United States. Dept. of Agriculture. Economic Research Service. (2004). *Taxing snack foods what to expect for diet and tax revenues*. Washington, D.C.: Economic Research Serivce U.S. Dept. of Agriculture.

Lakdawalla, D., & Philipson, T. J. (2006). Economics of Obesity. In A. M. Jones (Ed.), *The Elgar companion to health economics* (pp. 72-82). Massachusetts: Edward Elgar Publishing Inc.

Lakdawalla, D., Philipson, T. J., & Bhattacharya, J. (2005). Welfare-enhancing technological change and the growth of obesity. *American Economic Review, 95*(2), 253-257.

Liu, S., Serdula, M. K., Williamson, D. F., Mokdad, A. H., & Byers, T. (1994). A prospective study of alcohol intake and change in body weight among US adults. *Am J Epidemiol, 140*(10), 912-920.

Lovejoy, J. C., & Sainsbury, A. (2009). Sex differences in obesity and the regulation of energy homeostasis. *Obesity Reviews, 10*(2), 154-167.

Mark, B. A., Harless, D. W., McCue, M., & Xu, Y. (2004). A Longitudinal Examination of Hospital Registered Nurse Staffing and Quality of Care. *Health Services Research, 39*(2), 279-300.

Maskarinec, G., Meng, L. X., & Kolonel, L. N. (1998). Alcohol intake, body weight, and mortality in a multiethnic prospective cohort. *Epidemiology, 9*(6), 654-661.

Matthews, C. E. (2002). Use of self-report instruments to assess physical activity. In G. Welk (Ed.), *Physical activity assessments for health-related research* (pp. x, 269). Champaign, IL: Human Kinetics.

Meisinger, C., Lowel, H., Thorand, B., & Doring, A. (2005). Leisure time physical activity and the risk of type 2 diabetes in men and women from the general population. The MONICA/KORA Augsburg Cohort Study. *Diabetologia, 48*, 27 - 34.

Michaud, P.-C., & van Soest, A. (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics, 27*(5), 1312-1325.

Miller, W. C., Lindeman, A. K., Wallace, J., & Niederpruem, M. (1990). Diet composition, energy intake, and exercise in relation to body fat in men and women. *Am. J. Clinical Nutrition, 52*(3), 426-430.

Monda, K. L., Gordon-Larsen, P., Stevens, J., & Popkin, B. M. (2007). China's transition: The effect of rapid urbanization on adult occupational physical activity. *Soc Sci Med, 64*(4), 858-870.

Moore, M. S. (2000). Interactions between physical activity and diet in the regulation of body weight. *Proc Nutr Soc, 59*(2), 193-198.

Nagel, C. L., Carlson, N. E., Bosworth, M., & Michael, Y. L. (2008). The Relation between Neighborhood Built Environment and Walking Activity among Older Adults. *Am. J. Epidemiol., 168*(4), 461-468.

Ng, S. W., Norton, E. C., & Popkin, B. M. (2008). Why have physical activity levels declined among Chinese adults? . *Social Science and Medicine, Under review*.

Ng, S. W., Zhai, F., & Popkin, B. M. (2008). Impacts of China's edible oil pricing policy on nutrition. *Social Science & Medicine, 66*(2), 414-426.

Popkin, B. M. (2007). Understanding Global Nutrition Dynamics as a step toward controlling cancer morbidity and mortality. *Nature Reviews Cancer 247*, 61-67

Popkin, B. M., Kim, S., Rusev, E. R., Du, S., & Zizza, C. (2006). Measuring the full economic costs of diet, physical activity and obesity-related chronic diseases. *Obesity Reviews, 7*(3), 271-293.

Rashad, I., & Grossman, M. (2004). The economics of obesity. *Public Interest*(156), 104-112.

Rodriguez, D. A., Aytur, S., Forsyth, A., Oakes, J. M., & Clifton, K. J. (2008). Relation of modifiable neighborhood attributes to walking. *Prev Med*.

Roodman, D. (2003). XTABOND2: Stata module to extend xtabond dynamic panel data estimator (Version S435901): Boston College Department of Economics.

Sacks, F. M., Bray, G. A., Carey, V. J., Smith, S. R., Ryan, D. H., Anton, S. D., et al. (2009). Comparison of Weight-Loss Diets with Different Compositions of Fat, Protein, and Carbohydrates. *N Engl J Med, 360*(9), 859-873.

Saris, W. H. (1998). Fit, fat and fat free: the metabolic aspects of weight control. *Int J Obes Relat Metab Disord, 22 Suppl 2*, S15-21.

Schroeter, C., Lusk, J., & Tyner, W. (2008). Determining the impact of food price and income changes on body weight. *Journal of Health Economics, 27*(1), 45-68.

Sturm, R. (2004). The economics of physical activity - Societal trends and rationales for interventions. *American Journal of Preventive Medicine, 27*(3), 126-135.

Sundquist, K., Qvist, J., Johansson, S. E., & Sundquist, J. (2005). The long-term effect of physical activity on incidence of coronary heart disease: a 12-year follow-up study. *Prev Med, 41*, 219 - 225.

Tryon, W. W. (1987). Activity as a function of body weight. *Am. J. Clinical Nutrition, 46*(3), 451-455.

Wang, Y., Ge, K., & Popkin, B. M. (2000). Tracking of body mass index from childhood to adolescence: a 6-y follow-up study in China. *Am J Clin Nutr, 72*(4), 1018-1024.

West, R., Gilsenan, A., Coste, F., Zhou, X. L., Brouard, R., Nonnemaker, J., et al. (2006). The ATTEMPT cohort: a multi-national longitudinal study of predictors, patterns and consequences of smoking cessation; introduction and evaluation of internet recruitment and data collection methods. *Addiction, 101*(9), 1352-1361.

Westerterp, K. R., & Goran, M. I. (1997). Relationship between physical activity related energy expenditure and body composition: a gender difference. *International Journal of Obesity & Related Metabolic Disorders, 21*(3), 184.

Willett, W. C. (1998). Dietary fat and obesity: an unconvincing relation. *Am J Clin Nutr, 68*(6), 1149-1150.

Williamson, D. F., Madans, J., Anda, R. F., Kleinman, J. C., Giovino, G. A., & Byers, T. (1991). Smoking cessation and severity of weight gain in a national cohort. *N Engl J Med, 324*(11), 739-745.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics, 126*(1), 25-51.

Xu, F., Yin, X. M., & Wang, Y. F. (2007). The association between amount of cigarettes smoked and overweight, central obesity among Chinese adults in Nanjing, China. *Asia Pacific Journal of Clinical Nutrition, 16*(2), 240-247.

Zimmer, Z., Kaneda, T., & Spess, L. (2007). An examination of urban versus rural mortality in China using community and individual data. *J Gerontol B Psychol Sci Soc Sci, 62*(5), S349-357.

TABLE I

SAMPLE SIZE OF MEN FROM THE CHNS 1991-2006

Consecutive waves	**Unit of observation**	**Wave**						**Total**
		1991	**1993**	**1997**	**2000**	**2004**	**2006**	
T ≥ 2	Individuals		1,138	542	779	728	993	4,180
	Person-wave		2,014	1,381	1,756	1,743	1,751	8,645
T =1	Individuals	2,398	558	1,280	984	892	643	6,755
	Person-wave	2,873	654	1,487	1,314	1,060	850	8,238
Total	Individuals	2,398	1,696	1,822	1,763	1,620	1,636	10,935
	Person-wave	2,873	2,668	2,868	3,070	2,803	2,601	16,883

TABLE II

WEIGHT, PHYSICAL ACTIVITY LEVELS, DIETARY INTAKE, AND OTHER DEMOGRAPHIC CHARACTERISTICS AMONG ADULT MEN IN CHINA (CHNS 1991-2006)

	Year						Change
	1991	**1993**	**1997**	**2000**	**2004**	**2006**	**(1991-2006)**
Weight (kg)	59.25 (8.46)	59.71 (8.49)	61.65 (9.56)	63.70 (10.28)	65.93 (13.72)	66.43 (12.22)	7.18 **
Height (cm)	166.1 (6.26)	165.95 (6.17)	166.60 (6.36)	167.31 (6.38)	167.66 (6.65)	167.88 (6.88)	1.78 **
Work & Domestic physical activity level (MET-hrs/week)	389.84 (220.37)	356.92 (217.01)	346.40 (215.97)	305.97 (202.97)	246.00 (180.18)	247.09 (177.90)	-142.75 **
Total caloric Intake (kcal)	2972.73 (826.07)	2872.65 (922.26)	2612.81 (717.89)	2618.09 (807.06)	2530.91 (804.02)	2458.11 (774.49)	-514.62 **
Energy from dietary fat (%)	21.72	23.06	25.50	28.22	28.03	27.79	6.07 **
Smoker (%)	68.57	65.99	63.11	61.28	60.07	57.52	-11.05 **
Drinker (%)	67.77	64.14	67.12	65.27	64.15	63.32	-4.45 *
Age (year)	35.18 (9.93)	35.83 (10.08)	36.73 (10.19)	38.08 (10.10)	39.72 (9.95)	40.41 (9.78)	5.23 **
Married (%)	79.69	78.60	78.70	80.0	82.08	83.93	4.24 *
Live alone (%)	0.21	0.19	0.21	0.23	0.32	0.31	0.10 **
No education (%)	15.12	11.12	9.75	6.44	4.37	6.50	-8.62 **
Highest education is primary school (%)	62.68	65.61	64.66	62.79	61.75	54.94	-7.74 **
Highest education is secondary school (%)	15.26	16.50	17.46	18.49	19.75	20.57	5.31 **
Highest education is technical school (%)	3.19	3.98	4.12	5.77	7.62	9.00	5.81 **
Has university degree or higher (%)	3.75	2.79	4.01	6.51	6.51	8.99	5.24 **
Predicted household income (2006 yuan)	2228.2 (1443.17)	2261 (3746.15)	2545.71 (2362.72)	3573.33 (3432.76)	4431.98 (4419.80)	5243.65 (8047.30)	3015.45 **
Number of observations	2851	2649	2841	2980	2795	2601	

Standard Deviations in parentheses

* difference between 1991 and 2006 is significant at 5%; ** significant at 1%

TABLE III

URBANIZATION AND PRICES ON WEIGHT, PHYSICAL ACTIVITY, DIETARY INTAKE, SMOKING AND DRINKING AMONG CHINESE MEN

Community Factors	**Lagged Weight (kg)**	**Lagged Activity (MET-hrs/wk)**	**Lagged Total caloric intake (kcal)**	**Lagged % energy from Fat**	**Lagged P(Smoker)**	**Lagged P(Drinker)**
Population score (0-10)	-0.032	0.33	**6.916**	-0.061	0.026	-0.009
	(0.041)	(0.737)	(3.591) *	(0.041)	(0.019)	(0.015)
Density score (0-10)	0.02	**-3.141**	**-7.733**	**0.138**	-0.017	**0.058**
	(0.044)	(0.802) **	(3.609) *	(0.042) **	(0.02)	(0.016) **
Market Accessibility score (0-10)	0.002	**-3.013**	**-11.56**	-0.036	**-0.031**	**-0.039**
	(0.025)	(0.624) **	(2.884) **	(0.035)	(0.013) *	(0.011) **
Transportation score (0-10)	0.024	**-3.471**	**-16.435**	**0.142**	0.002	0.015
	(0.036)	(0.853) **	(4.055) **	(0.046) **	(0.018)	(0.016)
Communications score (0-10)	**0.079**	**-2.513**	-5.84	**0.268**	-0.02	0.009
	(0.042) +	(1.08) *	(4.81)	(0.056) **	(0.022)	(0.019)
Economy score (0-10)	0.063	**-7.465**	**12.924**	**0.183**	0.012	0.012
	(0.054)	(1.078) **	(5.175) **	(0.068) **	(0.025)	(0.021)
Educational Institution score (0-10)	**0.176 ****	**-10.403**	**-35.684**	**0.339**	0.001	0.018
	(0.049)	(1.106) **	(4.672) **	(0.056) **	(0.025)	(0.02)
Health Facilities score (0-10)	0.07	**-3.918**	-5.407	**0.204**	**0.039**	0.015
	(0.049)	(1.05) **	(4.813)	(0.054) **	(0.022) +	(0.018)
Sanitation infrastructure score (0-10)	**0.048**	**-9.085**	1.042	**0.197**	0.024	-0.008
	(0.029) +	(0.775) **	(4.007)	(0.039) **	(0.016)	(0.014)
Housing infrastructure score (0-10)	**0.102**	**-6.594**	-5.224	**0.523**	-0.015	-0.031
	(0.05) *	(1.208) **	(6.197)	(0.068) **	(0.027)	(0.023)
Log real price of Rice (yuan/kg)	**0.662**	-0.709	-13.521	**-0.894**	**0.39**	-0.069
	(0.382) +	(8.935)	(43.193)	(0.497) +	(0.202) *	(0.171)
Log real price of Flour (yuan/kg)	**-1.543**	-3.734	-13.81	**1.663**	-0.012	**-0.268**
	(0.326) **	(8.411)	(39.251)	(0.475) **	(0.181)	(0.156) +
Log real price of Pork (yuan/kg)	**-1.308**	**17.387**	-2.27	**-1.367**	0.082	-0.047
	(0.332) **	(6.425) **	(32.506)	(0.365) **	(0.151)	(0.126)
Log real price of Chicken (yuan/kg)	**-2.495**	-11.037	**156.631**	**1.837**	**0.575**	-0.031
	(0.331)	(7.599)	(38.648) **	(0.461) **	(0.176) **	(0.148)
Log real price of Oil (yuan/liter)	**-0.478**	2.847	**-54.151**	**-0.998**	-0.036	-0.134
	(0.208) *	(4.476)	(22.623) *	(0.244) **	(0.099)	(0.082)
Log real price of Beer (yuan/bottle)	-0.364	-7.385	-20.696	0.232	-0.086	**-0.243**
	(0.357)	(6.99)	(34.429)	(0.412)	(0.171)	(0.142) +
Log real price of Cigarettes (yuan/box of 20)	0.067	**-13.128**	**-33.014**	**0.486**	**-0.172**	-0.065
	(0.154)	(4.072) **	(19.263) *	(0.213) *	(0.082) *	(0.069)
Consumer Price Index (100=urban Liaoning)	**-8.768**	18.71	79.594	**14.256**	**1.865**	0.409
	(1.694) **	(31.864)	(138.207)	(1.742) **	(0.741) *	(0.611)
Observations	8645	8645	8645	8645	8645	8645
Number of Individuals	4120	4120	4120	4120	4120	4120
Overall Statistic $\chi^2(35)$	2064 **	5970.48 **	737.72 **	3179.27 **	238.72 **	253.15 **
Joint test of significance for all community variables $\chi^2(18)$	223.34 **	1856.63 **	233.72 **	1414.28 **	37.76 **	50.45 **
Joint test of significance for urbanization measures $\chi^2(10)$	67.59 **	1402 **	176.49 **	651.03 **	13.08	28.04 **
Joint test of significance for price variables $\chi^2(8)$	156.57 **	21.28 **	27.26 **	157.36 **	23.79 **	17.84 *

Controlling for height, time, age, marital status, living situation, education, and predicted household income. Robust standard errors in parentheses. + significant at 10%;* significant at 5%; ** significant at 1%

TABLE IV

RESULTS FROM DIFFERENT APPROACHES TO ESTIMATE DETERMINANTS OF WEIGHT AMONG CHINESE MEN

	(1) **Exogenous regressors**	(2) **Instrumental Variables**	(3) **Two-step system GMM**
Key Explanatory Variables			
Current			
Height (cm)	**0.429** (0.031) **	**0.528** (0.216)*	**0.711** (0.276)*
Lagged (t-1)			
Weight (kg)	**0.512** (0.035) **	0.424 (0.266)	**0.313** (0.072)**
Occupational & Domestic Physical activity level (METs-hours/week)	**-0.002** (0.0004) **	0.008 (0.008)	**-0.003** (0.002) *
Total Caloric Intake (kcal)	0.0002 (0.0001)	-0.0004 (0.001)	-0.0001 (0.0002)
Energy from Fat (%)	0.009 (0.008)	0.067 (0.086)	**0.047** (0.022)*
Smoker	**-0.490** (0.194) **	-0.363 (4.633)	0.299 (0.597)
Drinker	0.375 (0.204)	1.456 (4.616)	0.037 (0.525)
Instruments	None	Δ community variables used as instruments for lagged endogenous variables	Δ community variables for first difference equation. Lagged difference for weight, physical activity, dietary intake, smoking and drinking status for t-1 and prior for level equation.
Observations	8645	8645	8645
Number of Unique Individuals	4120	4120	4120
Number of Instruments used	None	35	79
Overall Statistic	$\chi^2(23)$= 6027.44 **	$\chi^2(23)$= 5175.05 **	$\chi^2(23)$ = 206.34 **
Difference-in-Hansen tests of exogeneity of Community variables as instruments			
Hansen test excluding group			$\chi^2(37)$ = 41.28
Difference (H_0 = exogenous)			$\chi^2(18)$ = 20.72
Sargan/Hansen's test of overidentification		$\chi^2(12)$= 11.57	$\chi^2(55)$ = 62.00
Test for Autocorrelation			
AR(1) in first differences (*z*-statistic)			-4.06 **
AR(2) in first differences (*z*-statistic)			1.38

Among those with 2 or more consecutive waves of data. Controlling for time, age, marital status, living situation, education, and predicted household income. Robust standard errors in parentheses.
* significant at 5%; ** significant at 1%

TABLE V

CONTRIBUTION OF PHYSICAL ACTIVITY AND ENERGY FROM DIETARY FAT ON WEIGHT GAIN AMONG CHINESE MEN

	(1) Change (1991 – 2006)	**(2) Coefficient from Table 4, Column 3**	**(3) Absolute Contribution (kg) (1)x(2)x5 intervals[‡]**	**(4) Percentage Contribution (%)**
Occupational & Domestic Physical activity level (METs-hours/week)	-142.75	-0.003	2.14	29.81
Energy from Fat (%)	6.07	0.047	1.43	19.92
Weight (kg)	7.18		3.57	49.73
Unexplained			3.61	50.27

[‡] Since the coefficients are for the change in the explanatory variable from t-1 to t, we need to multiply the coefficient by the 5 intervals.

FIGURE I

WEIGHT AND BODY MASS INDEX OF ADULT MEN IN CHINA (CHNS 1991-2006)

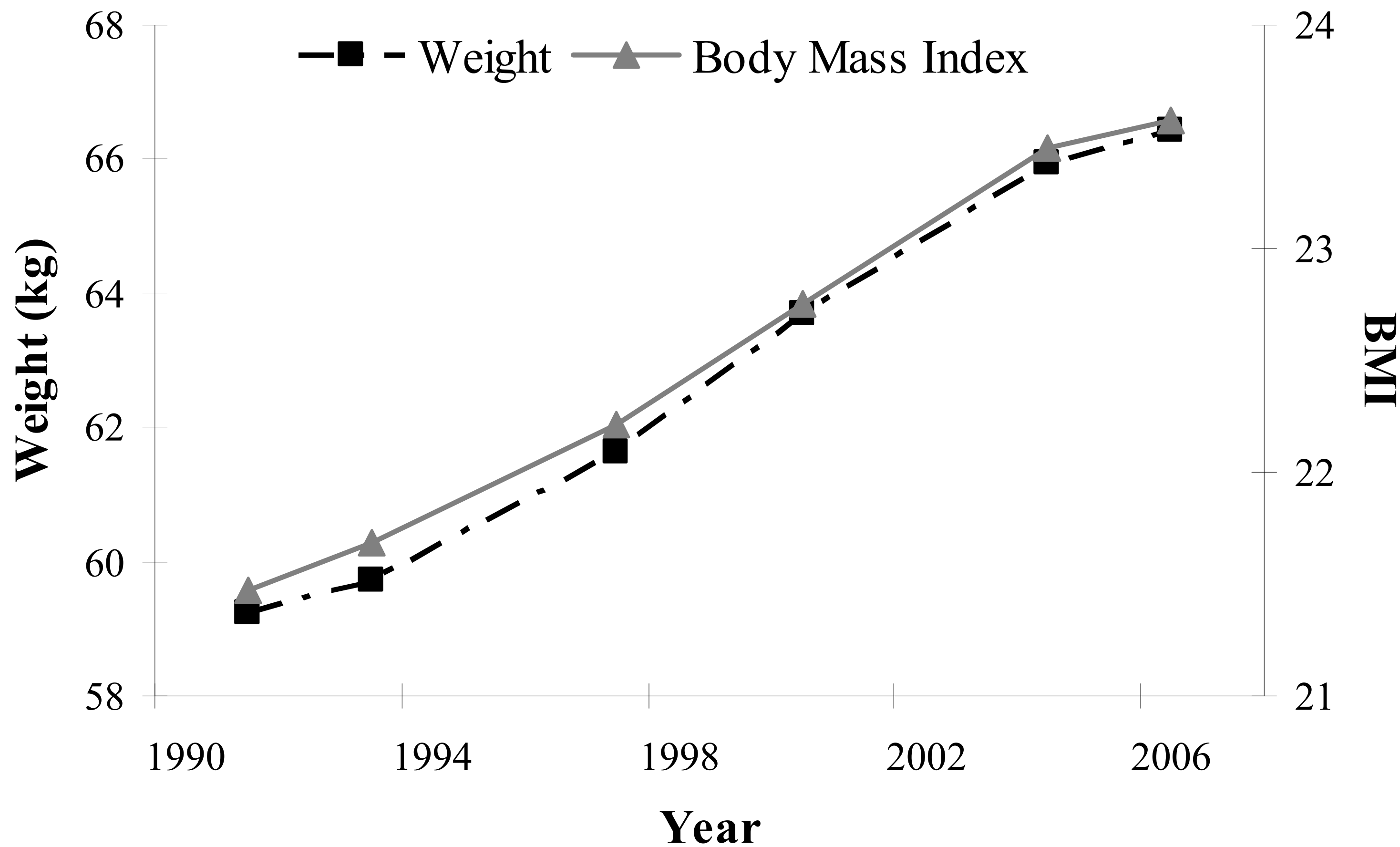
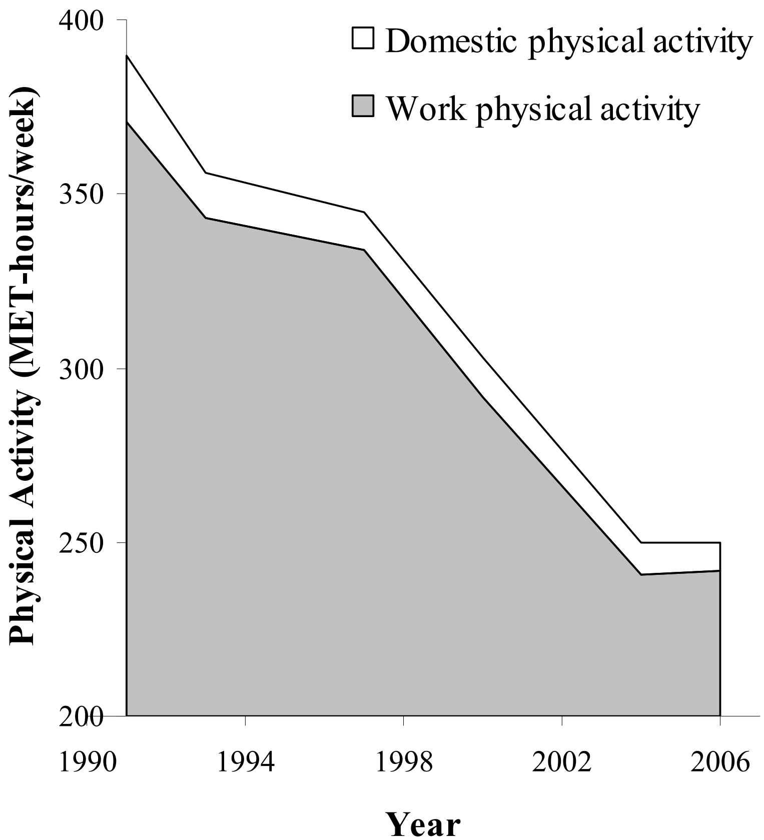


FIGURE II

PHYSICAL ACTIVITY LEVELS, DAILY CALORIC INTAKE AND PROPORTION OF ENERGY FROM DIETARY FAT AMONG ADULT MEN IN CHINA (CHNS 1991-2006)

II.a. Change in work and domestic activity



II.b. Change in total caloric intake and energy from fat

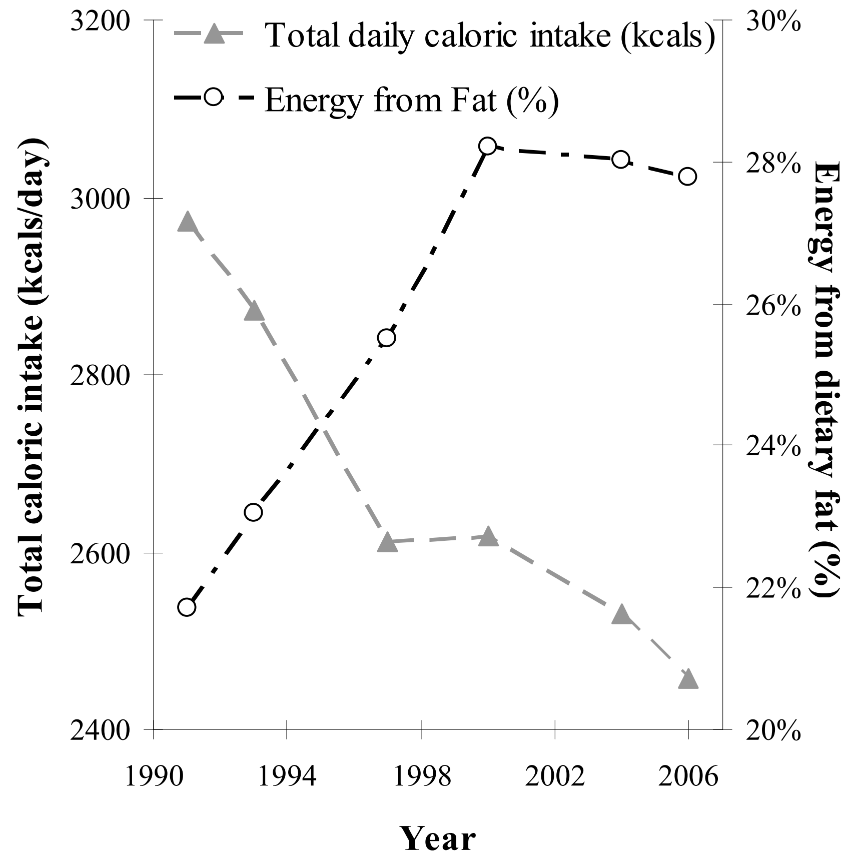


FIGURE III

AVERAGE CHANGE IN SELECT URBANIZATION DOMAINS AMONG COMMUNITIES IN CHINA (CHNS 1991-2006)

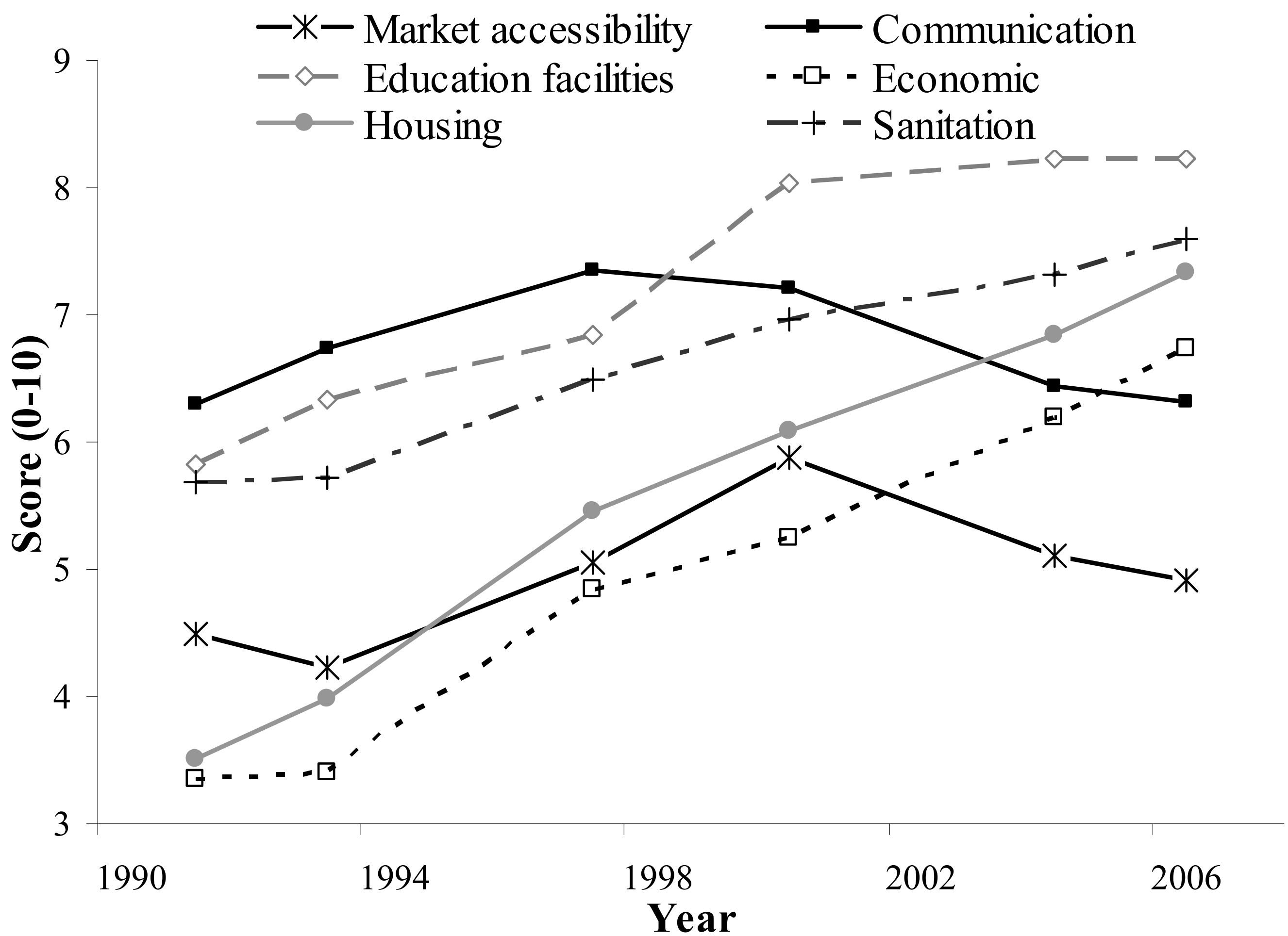
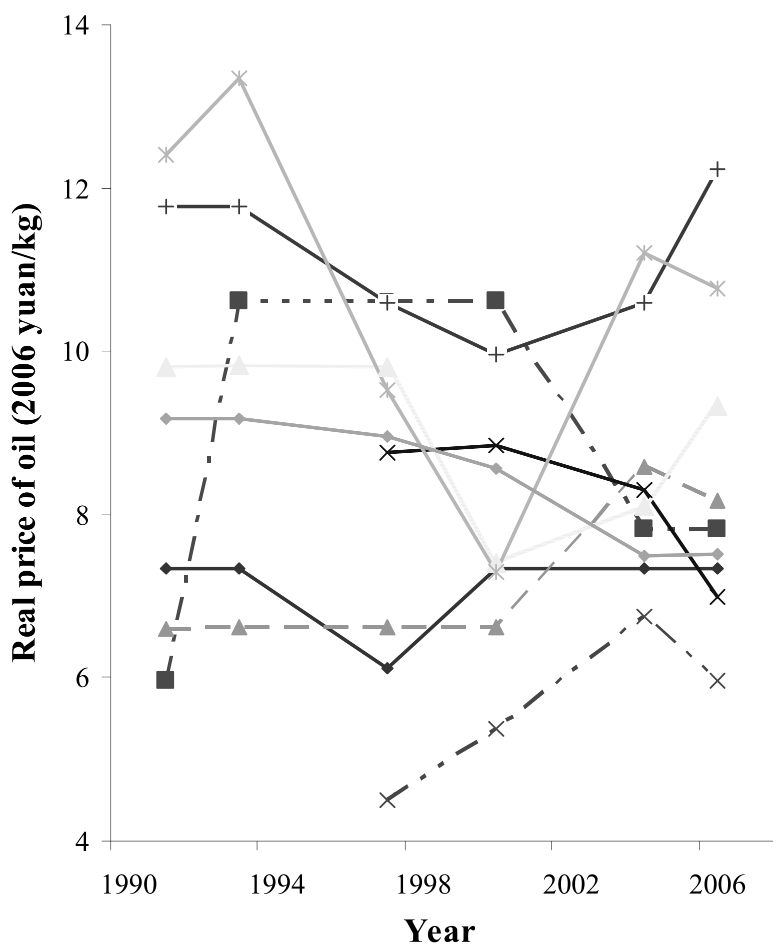


FIGURE IV

EXAMPLES OF PRICE VARIATION OVER TIME AND ACROSS THE TEN MOST POPULOUS COMMUNITIES (CHNS 1991-2006)

IV.a. Change in real price of oil



IV.b. Change in real price of flour