Health, Econometrics and Data Group

THE UNIVERSITY of York

HEDG Working Paper 09/19

Lose Weight for Money Only if Over-Weight: Marginal Integration for Semi-Linear Panel Models

Kamhon Kan

Myoung-jae Lee

July 2009

ISSN 1751-1976

york.ac.uk/res/herc/hedgwp

Lose Weight for Money Only if Over-Weight: Marginal Integration for Semi-Linear Panel Models

(May 29, 2009)

Kamhon Kan
Institute of Economics
Academia Sinica
128, Academia Rd., Sec 2
Taipei 115, Taiwan, R.O.C
kk@sinica.edu.tw

Myoung-jae Lee*
Department of Economics
Korea University
Anam-dong, Sungbuk-gu
Seoul 136-701, South Korea
myoungjae@korea.ac.kr

Body mass index (BMI), weight(kg)/ height(m)^2, is a widely used measure for obesity in medical science. In economics, there appeared studies (e.g., Cawley (2004) and Burkhauser and Cawley (2008)) showing that BMI has a negative (or no) effect on wage. But BMI is a tightly specified function of weight and height, and there is no priori reason to believe why the particular function is the best to combine weight and height. In this paper, we address the question of *weight effect on wage*, employing two-wave panel data for white females; the same panel data with more waves were used originally in Cawley (2004). We posit a semi-linear model consisting of a nonparametric function of height and weight and a linear function of the other regressors. The model is differenced to get rid of the unit specific effect, which results in a difference of two nonparametric functions with the same shape. We estimate each nonparametric function with a 'marginal integration method', and then combine the two estimated functions using the same shape restriction. We find that there is *no weight effect on wage up to the average weight, beyond which a large negative effect* kicks in. The effect magnitude is greater than that in Cawley (2004) who used a linear BMI model. The linear model gives the false impression that there would be a wage gain by becoming slimmer than the average and that the 'obesity penalty' is less that what it actually is.

JEL Classification Numbers: C14, C33, I10, J30.

Key Words: BMI, weight effect on wage, panel data, semi-linear model, marginal integration.

* Corresponding Author: myoungjae@korea.ac.kr; 82-2-3290-2229 (phone); 82-2-926-3601.

1 Introduction

Consider a nonparametric "related-effect" panel data model:

$$Y_{it} = \rho(C_i, X_{it}) + \alpha_i + U_{it}, \qquad i = 1, \ldots, N, \quad t = 1, 2 \tag{1.1}$$

where Y_{it} is a response variable, C_i is a time-constant regressor, X_{it} is a time-variant regressor, $\rho(C_i, X_{it})$ is an unknown function of C_i and X_{it}, α_i is an unobserved time-constant error possibly related to C_i and X_{it}, and U_{it} is a time-variant error such that

$$E(U_{it}|C_i, X_{i1}, X_{i2}, \alpha_i) = \text{a time-constant function of } C_i \text{ and } \alpha_i, \quad t = 1, 2; \tag{1.2}$$

i indexes individuals and t indexes time periods. Assume iid (independent and identically distributed) across i to often omit the subscript i in the rest of this paper. This moment condition includes the moment being zero as a special case.

The expression "related-effect" refers to α_i being possibly related to regressors. In the panel data literature, related-effect is usually called "fixed-effect," which is, however, also used for cases where α_i is estimated (along with the model parameters) regardless of its relationship with regressors. In (1.2), all period regressors are in the conditioning set ('strict exogeneity'), which is typically invoked in the panel related-effect literature as can be seen in Manski (1987), Honoré (1992), Kyriazidou (1997) and Lee (1999), although there are exceptions as in Holtz-Eakin et al. (1988), Chamberlain (1992) and Wooldridge (1997).

A standard way to deal with the 'unit-specific term' α_i is first-differencing across the two periods. For instance, when

$$\rho(C_i, X_{it}) = \beta_1 + \beta_c C_i + \beta_x X_{it} + \beta_{xc} X_{it} C_i + \beta_{xx} X_{it}^2, \tag{1.3}$$

first-differencing yields

$$Y_{i2} - Y_{i1} = \beta_x (X_{i2} - X_{i1}) + \beta_{xc}(X_{i2} - X_{i1}) C_i + \beta_{xx}(X_{i2}^2 - X_{i1}^2) + U_{i2} - U_{i1}. \tag{1.4}$$

While first-differencing is straightforward with a parameterized regression function as (1.3), a misspecified parametric function in general leads to inconsistent estimators. The goal of this paper is to explore a kernel nonparametric estimation for the semi-linear regression function, using the idea of '*marginal integration*' in Linton and Nielsen (1995) and Newey (1994).

The linear model (1.3) suggests that, if a series-approximation is used for the nonparametric model, then we may not need a set-up fancier than the usual linear model to handle the related-effect. But series approximation, as a global nonparametric method, has properties different from kernel method which is a local approximation. The two methods have pros and cons. A difficulty with series approximation relative to kernel method is that, if the regression function is high-dimensional only in a small area, then series approximation will force this feature into the whole support of the regression function. Another difficulty is that, while choosing the order of series approximation can be done automatically, say with cross validation (CV), the order taking integers is too rough a measure for the degrees of smoothing, while the degree of smoothing can be chosen as finely as desired in kernel methods. An advantage of series approximation is that a series-approximated model can be fit as a familiar linear model, and thus many familiar estimation/test techniques for linear models are applicable. But this linear model analogy is "deceiving" in theory, as the convergence rate of a series-approximation estimator is slower than the usual N^{-12} rate.

The nonparametric model (1.1) is relevant, e.g., for nonparametric growth curve estimation (see Müller (1988) and references therein) where Y_{it} is the height of a child, C_i is gender, X_{it} is nutrition, and α_i captures the genetic factors. The prime example we have in mind in this paper is the effect of body mass index index (BMI) on wage where BMI is defined as weight in kg divided by squared height in meters. As height C_i is fixed for adults while weight X_{it} changes, the real *effect of interest is the effect of weight* X_{it} *on* $Y_{it} = \ln(wage_{it})$. While BMI specifies $\rho(c, x)$ as proportional to x/c^2, from modeling viewpoint, this is a very tight specification—think of all possible functional forms of c and x. It will be a remarkable luck of draw if the functional form x/c^2 holds up well in reality.

Since wage depends on variables other than weight and height, we will generalize the nonparametric model into a semi-linear model

$$Y_{it} = \rho(C_i, X_{it}) + W_{it}'\beta + \alpha_i + U_{it} \tag{1.5}$$

where W_{it} is the other regressors possibly affecting wage. Not to get distracted by W_{it}, however, we will examine (1.1) without W_{it} in detail first, and then (1.5) later; the generalization with W_{it} does not take much extra work.

Li and Stengos (1996) extended Robinson's (1988) two-stage approach for cross-section semi-linear models to panel data without much concern on α_i. Lin and Carroll (2006) examined semi-linear panel data models where ρ is a function of time-varying regressors. Although a differenced model was considered briefly for matched-pair samples, Lin and Carroll (2006) primarily looked at 'unrelated-effect' (or random-effect) models. Henderson et al. (2008) applied the Lin and Carroll (2006) approach to panel semi-linear models to estimate differenced semi-linear models with 'profile-based' kernel methods. Despite these studies, models of the form (1.5) with a time-constant C_i non-separable from X_{it} in the nonparametric part have not been specifically examined in the literature as far as we are aware of. Also our marginal integration approach to be explained in the next section is much simpler than the profile-based kernel methods.

The rest of this paper is organized as follows. Section 2 presents our kernel estimator for a normalized version of $\rho(c, x)$ in (1.1) using the marginal integration idea, assuming that C_i and X_{it} are continuously distributed. Section 3 examines the augmented model (1.5) using the two-stage method in Robinson (1988). Section 4 presents our empirical analyses for the same data as used in Cawley (2004) that were originally drawn from NLSY (National Longitudinal Survey of Youth). Finally, Section 5 draws conclusions.

Showing our main conclusion in advance, we find no weight effect on wage up to the average weight, beyond which there is a fairly large negative effect. That is, for each given height, the log wage function is flat over the under-weight range and then declines rapidly over the over-weight range. This is in contrast to what Cawley (2004) found using a model linear in BMI (and the other regressors) where the effect magnitude of BMI is smaller than our effect over the over-weight range. This seems to be due to combining zero effect over the under-weight range and the relatively greater effect in absolute value over the over-weight range. Hence our main finding can be summed up as follows: *there is no wage gain by becoming slimmer than normal, but there is a higher wage gain than suggested in the linear model if over-weight.*

2 Marginal Integration

2.1 Main Idea

First-difference (1.1) to get

$$\Delta Y_i = \mu(C_i, X_{i1}, X_{i2}) + \Delta U_i \qquad \text{where } \Delta Y_i \equiv Y_{i2} - Y_{i1}, \quad \Delta U_i \equiv U_{i2} - U_{i1} \text{ and}$$

$$\mu(C_i, X_{i1}, X_{i2}) \equiv \rho(C_i, X_{i2}) - \rho(C_i, X_{i1}).$$

Let $\hat{\mu}(c, x_1, x_2)$ denote a kernel nonparametric estimator for $\mu(c, x_1, x_2)$. Two marginally integrated versions of $\hat{\mu}(c, x_1, x_2)$ are

$$\begin{aligned}\hat{\mu}_{c1}(c, x_1) &\equiv \frac{1}{N}\sum_i \hat{\mu}(c, x_1, X_{i2}) \to^p \int \rho(c, x_2) f_2(x_2) dx_2 - \rho(c, x_1) \equiv \mu_{c1}(c, x_1) \\ \hat{\mu}_{c2}(c, x_2) &\equiv \frac{1}{N}\sum_i \hat{\mu}(c, X_{i1}, x_2) \to^p \rho(c, x_2) - \int \rho(c, x_1) f_1(x_1) dx_1 \equiv \mu_{c2}(c, x_2)\end{aligned}$$

where f_t denotes the X-density for time t. There is no difficulty in estimating μ; the question is how to take the advantage of the additive structure of μ in ρ.

Observe

$$\begin{aligned}\hat{m}(c, x) &\equiv \frac{\hat{\mu}_{c2}(c, x) - \hat{\mu}_{c1}(c, x)}{2} \to^p m(c, x) \quad \text{where} \\ m(c, x) &\equiv \frac{\mu_{c2}(c, x) - \mu_{c1}(c, x)}{2} = \rho(c, x) - \int \rho(c, x) \frac{f_1(x) + f_2(x)}{2} dx.\end{aligned}$$

Hence $\hat{m}(c, x)$ *is a consistent estimator for* $\rho(c, x)$ *up to a function of of* c. Each of $-\hat{\mu}_{c1}(c, x)$ and $\hat{\mu}_{c2}(c, x)$ is a valid estimator for $\rho(c, x)$ up to a function of c. But, by combining the two estimators as in this display, we are taking advantage of the information/restriction that the probability limits of $-\hat{\mu}_{c1}(c, x)$ and $\hat{\mu}_{c2}(c, x)$ are the same. Due to the constant 1/2, the resulting estimator has the standard deviation (SD) twice smaller than when only a single estimator is used.

If $\rho(c, x) = x/c^2$ as in BMI, then

$$\begin{aligned}&\int \rho(c, x) \frac{f_1(x) + f_2(x)}{2} dx = \frac{1}{c^2} \int x \frac{f_1(x) + f_2(x)}{2} dx \\ \implies \quad & m(c, x) = \frac{x - E_{12}(x)}{c^2} \quad \text{where } E_{12}(x) \equiv \int x \frac{f_1(x) + f_2(x)}{2} da.\end{aligned}$$

Hence the normalized version $m(c, x)$ of ρ is just a 'X-mean'-centered version of ρ where the X-mean is obtained using the simple-averaged marginal densities.

Before we proceed further, we make two remarks. Firstly, although we assumed that the same functional form ρ holds in the two periods, we can in fact easily allow a time-varying intercept, say τ_t:

$$Y_{it} = \tau_t + \rho(C_i, X_{it}) + \alpha_i + U_{it} \implies \mu(C_i, X_{i1}, X_{i2}) \equiv \Delta\tau + \rho(C_i, X_{i2}) - \rho(C_i, X_{i1})$$

where $\Delta\tau \equiv \tau_2 - \tau_1$. But $\Delta\tau$ will get cancelled in the difference $\hat{\mu}_{c2}(c, x) - \hat{\mu}_{c1}(c, x)$. This shows that a time-varying intercept is allowed in the model. Secondly, over a short period of time, $f_1 = f_2$ can happen; for the BMI example, the marginal distribution of weight may not change although some people gain weight while some people lose. With $f_1 = f_2$,

$$m(c,x) = \rho(c,x) - \int \rho(c,x) f_0(x) dx \quad \text{where } f_1 = f_2 \equiv f_0;$$

$\int m(c,x) f_0(x) dx = 0$ by construction. With $\rho(c,x) = x/c^2$ as in BMI,

$$m(c,x) = \frac{x - E(x)}{c^2} \quad \text{where } E(x) = \int x f_0(x) dx.$$

2.2 Estimation Details

Define

$$Z_i \equiv (C_i, X_{i1}, X_{i2})'.$$

Let $f_z(z)$ be the density function for $Z = z$; the components of z will be denoted also as z_1, z_2, and z_3. For a three-dimensional product kernel $K(z) = L(z_1)L(z_2)L(z_3)$ and a bandwidth h, define

$$\hat{f}_z(z) \equiv \frac{1}{Nh^3} \sum_{i=1}^{N} K(\frac{Z_i - z}{h}), \quad \hat{g}_z(z) \equiv \frac{1}{Nh^3} \sum_{i=1}^{N} K(\frac{Z_i - z}{h}) \Delta Y_i, \quad \hat{\mu}(z) \equiv \frac{\hat{g}_z(z)}{\hat{f}_z(z)};$$

e.g., we may use the product of three $N(0,1)$ densities for K: $K(z) = \phi(z_1)\phi(z_2)\phi(z_3)$.

In practice, to account for the scale differences among the regressors, it is necessary to use a different bandwidth for each regressor proportional to its SD as in

$$\hat{f}_z(z) = \frac{1}{N\hat{\sigma}_c \hat{\sigma}_{x1} \hat{\sigma}_{x2} h_0^3} \sum_i \phi(\frac{C_i - c}{\hat{\sigma}_c h_0}) \cdot \phi(\frac{X_{i1} - x_1}{\hat{\sigma}_{x1} h_0}) \cdot \phi(\frac{X_{i2} - x_2}{\hat{\sigma}_{x2} h_0})$$

where $\hat{\sigma}_c$, $\hat{\sigma}_{x1}$ and $\hat{\sigma}_{x2}$ are the sample SD's for C_i, X_{i1} and X_{i2}, respectively, and h^3 in the preceding display is replaced by the product of the three different bandwidths $\hat{\sigma}_c h_0$, $\hat{\sigma}_{x1} h_0$ and $\hat{\sigma}_{x2} h_0$. Then, set $h_0 = pN^{-1/7}$ to try, say $p = 0.5, 1, 2, 3$. The best way to choose p is drawing $\hat{f}_z(z)$ when the dimension of z is small. In our case, z is three-dimensional, and thus we can draw only its two-dimensional 'cross-sections'. Then choose a value of p that gives "not too rough nor too smooth" cross-sectional figures. A practical rule-of-thumb value for p is 1, which can be used at the first attempt.

With $\hat{\mu}(z) = \hat{g}_z(z)/\hat{f}_z(z)$, obtain the averaged (i.e., integrated) versions and their linear combination:

$$\hat{\mu}_{c1}(c, x_1) \equiv \frac{1}{N} \sum_i \hat{\mu}(c, x_1, X_{i2}), \quad \hat{\mu}_{c2}(c, x_2) \equiv \frac{1}{N} \sum_i \hat{\mu}(c, X_{i1}, x_2)$$

$$\hat{m}(c,x) \equiv \frac{\hat{\mu}_{c2}(c,x) - \hat{\mu}_{c1}(c,x)}{2}.$$

A three-dimensional graph is needed to plot $m(c, x)$, but in practice, it will be simpler to plot a number of two-dimensional graphs with c fixed at some points. If the BMI functional form is correct, fixing c means that the resulting graphs should be all linear because $\rho(c, x) = x/c^2$. In our empirical analysis later, we will fix c at the lower quartile (LQ), median (MED) and upper quartile (UQ).

2.3 Asymptotic Variance

For a two-dimensional regression function, say $m(x_1, x_2)$, in a cross-section linear model $Y_i = m(X_{i1}, X_{i2}) + U_i$, Linton and Nielsen (1995) stated that, under the homoskedasticity assumption $E(U|X_1 = x_1, X_2 = x_2) = \sigma^2$,

$$\sqrt{Nh}\{\tilde{m}_1(x_1) - \int m(x_1, x_2)q(x_2)dx\} \rightsquigarrow N\{0, \int L(s)^2 ds \cdot \sigma^2 \int \frac{q(x_2)^2}{f(x_1, x_2)} dx_2\}$$
$$\text{where} \quad \tilde{m}_1(x_1) \equiv \int \hat{m}(x_1, x_2)q(x_2)dx_2$$

for a weighting function $q(x_2)$, $\hat{m}(x_1, x_2)$ is a kernel estimator with the product kernel $L((X_{i1} - x_1)/h)L((X_{i2} - x_1)/h)$ and $f(x_1, x_2)$ is the joint density function for $(X_1 = x_1, X_2 = x_2)$.

A couple of extensions for this finding are notable. First, for heteroskedastic errors, $\sigma^2 \int \{q(x_2)^2/f(x_1, x_2)\}dx_2$ should be replaced with

$$\int \sigma^2(x_1, x_2)\frac{q(x_2)^2}{f(x_1, x_2)} dx_2.$$

Second, for $\tilde{m}_1(x_1) + \tilde{m}_2(x_2)$ where $\tilde{m}_2(x_2)$ is defined analogously to $\tilde{m}_1(x_1)$, its asymptotic variance is just the sum of the two individuals variances, that is, $\tilde{m}_1(x_1)$ and $\tilde{m}_2(x_2)$ are asymptotically independent. The sum is considered here because $\tilde{m}_1(x_1)$ was designed originally for additive nonparametric models, say $m(x_1, x_2) = m_1(x_1) + m_2(x_2)$. Third, if $q(x_2) = f_2(x_2)$, then the asymptotic variance allowing for heteroskedasticity becomes

$$\int L(s)^2 ds \int \sigma^2(x_1, x_2)\frac{f_2(x_2)^2}{f(x_1, x_2)} dx_2 = \int \sigma^2(x_1, x_2)\frac{f_2(x_2)}{f_{1|2}(x_1|x_2)} dx_2.$$

where $f_{1|2}(x_1|x_2) \equiv f(x_1, x_2)/f_2(x_2)$. This asymptotic variance also holds when the empirical distribution is used for $q(x_2)dx_2$ as in Linton (1997) to result in

$$\hat{m}_1(x_1) = \frac{1}{N}\sum_i \hat{m}(x_1, X_{i2}).$$

In the following, we present the asymptotic distribution for $\hat{m}(c,x)$ for the three-regressor case. The main steps in deriving the asymptotic variance for the above two-regressor case are presented in Lee (2010).

Generalizing the two-regressor case to three-regressor case, we get

$$\begin{aligned}
\sqrt{Nh^2}\{\hat{\mu}_{c1}(c,x)-\mu_{c1}(c,x)\} &\rightsquigarrow N[0,\ \{\int L(s)^2 ds\}^2 V_{1cx}], \quad V_{1cx} \equiv \int \sigma^2(c,x,x_2)\frac{f_2(x_2)^2}{f(c,x,x_2)}dx_2 \\
\sqrt{Nh^2}\{\hat{\mu}_{c2}(c,x)-\mu_{c2}(c,x)\} &\rightsquigarrow N[0,\ \{\int L(s)^2 ds\}^2 V_{2cx}], \quad V_{2cx} \equiv \int \sigma^2(c,x_1,x)\frac{f_1(x_1)^2}{f(c,x_1,x)}dx_1 \\
\text{where } \sigma^2(c,x_1,x_2) &\equiv V(\Delta Y|C=c, X_1=x_1, X_2=x_2).
\end{aligned}$$

As the two estimators are asymptotically independent, we get

$$\sqrt{Nh^2}\{\hat{m}(c,x)-m(c,x)\} \rightsquigarrow N[0,\ \{\int L(s)^2 ds\}^2 \frac{V_{1cx}+V_{2cx}}{4}].$$

Define

$$\tilde{\mu}(c,x_1,x_2) \text{ as } \hat{\mu}(c,x_1,x_2) \text{ with its } \Delta Y \text{ replaced with } (\Delta Y)^2.$$

Then $\tilde{\mu}(c,x_1,x_2) \to^p E\{(\Delta Y)^2|c,x_1,x_2\}$, and thus

$$\tilde{\mu}(c,x_1,x_2) - \{\hat{\mu}(c,x_1,x_2)\}^2 \to^p \sigma^2(c,x_1,x_2).$$

Observe now

$$\begin{aligned}
V_{1cx} &= \int \sigma^2(c,x,x_2)\frac{f_2(x_2)}{f(c,x,x_2)}f_2(x_2)dx_2 \simeq \frac{1}{N}\sum_i \sigma^2(c,x,X_{i2})\frac{f_2(X_{i2})}{f(c,x,X_{i2})} \\
&\simeq \frac{1}{N}\sum_i [\{\tilde{\mu}(c,x,X_{i2}) - (\hat{\mu}(c,x,X_{i2}))^2\}\frac{\hat{f}_2(X_{i2})}{\hat{f}(c,x,X_{i2})}] \equiv \hat{V}_{1cx}.
\end{aligned}$$

Doing analogously,

$$V_{2cx} \simeq \frac{1}{N}\sum_i [\{\tilde{\mu}(c,X_{i1},x) - (\hat{\mu}(c,X_{i1},x))^2\}\frac{\hat{f}_1(X_{i1})}{\hat{f}(c,X_{i1},x)}] \equiv \hat{V}_{2cx}.$$

Therefore, a 95% asymptotic point-wise confidence interval (CI) for $m(c,x)$ is

$$\hat{m}(c,x) \pm 1.96 \int L(s)^2 ds \cdot \{\frac{\hat{V}_{1cx}+\hat{V}_{2cx}}{4Nh^2}\}^{1/2}.$$

$\int L(s)^2 ds$ is a known number; with $L(\cdot)$ being the $N(0,1)$ density, $\int L(s)^2 ds \simeq 0.283$. As noted already, we use a different bandwidth for each regressor in practice; for our problem, we use $\hat{\sigma}_c h_0$, $\hat{\sigma}_{x1} h_0$ and $\hat{\sigma}_{x2} h_0$. In this case, the last display becomes

$$\hat{m}(c,x) \pm 1.96 \int L(s)^2 ds \cdot \{\frac{\hat{V}_{1cx}}{4N\hat{\sigma}_c\hat{\sigma}_{x1}h_0^2} + \frac{\hat{V}_{2cx}}{4N\hat{\sigma}_c\hat{\sigma}_{x2}h_0^2}\}^{1/2}.$$

Although the above CI can be used for different points of x (with c fixed at one value) to get a 'confidence band (CB)' connecting those CI's, this lowers the coverage probability of the CB. For instance, suppose we obtain $\hat{m}(c,x)$ for 31 different evaluation points $x^{(j)}$, $j = 1, ..., 31$, of x. Then, with the asymptotic independence across the evaluation points holding, the coverage probability of the CB is only $0.95^{31} = 0.204$. If we use the critical value 2.93 instead of 1.96, then the coverage probability of the CB becomes $0.9966^{31} = 0.900$ as the coverage probability of one CI is 0.9966. We will be using 2.93 later for our CB's, but it should be noted that the CB's obtained this way are likely to be too conservative, because the asymptotic independence is indeed "asymptotic". In reality, adjacent CI's are likely to be positively related. That is, if one CI at $x^{(1)}$ contains $\rho(c, x^{(1)})$, then another CI at $x^{(2)}$ is likely to contain $\rho(c, x^{(2)})$ as well when $x^{(1)}$ and $x^{(2)}$ are close to each other.

3 Semi-Linear Model

Recall the semi-linear model with an extra regressor vector W_{it} in (1.5). First-differencing yields, with $\Delta W_i \equiv W_{i2} - W_{i1}$,

$$\Delta Y_i = \mu(C_i, X_{i1}, X_{i2}) + \Delta W_i'\beta + \Delta U_i \qquad (3.1)$$

which is a semi-linear cross-section model. As the time-constant elements of W_{it} drop out in this differenced model, there is no need to include time-constant elements in W_{it} even if they are relevant for Y_{it}.

Following the Robinson's (1988) idea of removing the nonparametric component, take $E(\cdot|C_i, X_{i1}, X_{i2})$ on (3.1) to get

$$E(\Delta Y|C_i, X_{i1}, X_{i2}) = \mu(C_i, X_{i1}, X_{i2}) + E(\Delta W_i'|C_i, X_{i1}, X_{i2})\beta. \qquad (3.2)$$

Subtract (3.2) from (3.1) to get

$$\Delta Y_i - E(\Delta Y|C_i, X_{i1}, X_{i2}) = \{\Delta W_i - E(\Delta W_i|C_i, X_{i1}, X_{i2})\}'\beta + \Delta U_i. \qquad (3.3)$$

Using this, the following three-stage estimator can be done.

First, obtain kernel estimators, say $\hat{E}(\Delta Y|C_i, X_{i1}, X_{i2})$ and $\hat{E}(\Delta W_i|C_i, X_{i1}, X_{i2})$, for the conditional means in (3.3). Second, get the Least Squares Estimator (LSE) b_N for β of

$$\Delta Y_i - \hat{E}(\Delta Y|C_i, X_{i1}, X_{i2}) \quad \text{on} \quad \Delta W_i - \hat{E}(\Delta W_i|C_i, X_{i1}, X_{i2});$$

the estimator is $\sqrt{N}$-consistent under some regularity conditions. Third, construct the '$\Delta Y - \Delta W' b_N$ residual'

$$Q_i \equiv \Delta Y_i - \Delta W_i' b_N$$

and proceed as in the preceding section treating Q_i as ΔY_i.

Some remarks are in order. First, since b_N is $\sqrt{N}$-consistent while the ensuing nonparametric procedure is $\sqrt{Nh^2}$-consistent, using b_N is as good as knowing β asymptotically. That is, we can ignore the presence of W_i as far as estimating $\rho(C_i, X_{it})$ goes. Second, we can easily allow β to vary over time: replace $\Delta W_i'\beta$ with $W_{i2}'\beta_2 - W_{i1}'\beta_1$ to estimate β_2 and β_1 using both W_{i2} and $-W_{i1}$ as the regressors. Third, for our purpose of estimating ρ, β does not have to be estimated "perfectly", because we just need the residual Q_i. That is, estimating each of β is not of concern; rather, the function $W_{it}'\beta$ as a whole should be estimated accurately even if each component of β may not be. Hence, if W_{it} has a multicollinearity problem, it would be better to use part of W_{it} instead of using all of W_{it} so that β for the used part can be estimated accurately rather than all components of β get estimated inaccurately.

4 Empirical Analysis: Weight Effect on Wage

4.1 Two Far-Apart Waves for White Females

Our study uses the same data as used in Cawley (2004) who drew the original data from NLSY (the National Longitudinal Survey of Youth)—we are grateful to Professor Cawley for providing us the data. The panel data used in Cawley (2004) has about 100,000 observations when pooled. Cawley found that only white females have significant negative weight effects on wage that are stable across different models and estimators. His main finding summarized in the abstract of the paper is that *two SD (roughly 65 pounds) weight difference is associated with 9 percent wage difference*. This came from LSE using BMI as a regressor along with some other regressors.

Burkhauser and Cawley (2008) found that BMI is not a good measure of obesity, as BMI does not distinguish muscle weight and fat weight; they obtained the correct obesity classification using body fat percentage (and some others). This resulted in misclassifying many non-obese persons to obese, and the problem is severe for men but weak for women as there are far more muscular men than women. Phrases cited in Burkhauser and Cawley (2008) condemning BMI are "Farewell to Body-Mass Index" and "the final nail in the casket for body-mass index as an independent cardiovascular risk factor". These demonstrate that BMI is an error-ridden measure of obesity, and the measurement error is likely to bias the estimated BMI effect toward zero. But Burkhauser and Cawley (2008) still found that BMI is a reliable measure for white female obesity, almost as good as body fat percentage in explaining employment status. Hence, among the various gender and ethnic groups examined in Cawley (2004), we use only while females in our empirical analysis. The original unbalanced panel data in Cawley (2004) has 13 unequally spaced waves for the period 1981-2000. Our analysis is based on a balanced panel of $N = 1302$ only for two waves 1986 and 2000 for the following reasons.

First, an unbalanced panel data set is cumbersome to use for difference-based methods because we should make sure of differencing two waves only for those individuals observed in the two waves. This problem can be avoided if we trim the unbalanced panel to make it balanced, but there is a trade-off: more waves means the smaller N. In our case, using the two waves 1986 and 2000 gives $N = 1302$, but if we try to use eight equally spaced (every other year) waves, then N becomes about 600.

Second, if we use more than two waves, then we can use each pair for the waves. This brings up the question on how to combine multiple estimators (one from each possible pair) to find an optimal estimator; recall that, even for two waves, we have a linear combination of two estimators. This does not seem to be an easy task theoretically, to say the least. Also even if this is done, the resulting optimal estimator is likely to depend on the estimators' variances and covariances. As will be noted later, unfortunately, the asymptotic variance and its estimator presented above do not work well for our data; the main problem is the appearance of $\hat{f}(c, x_1, x_2)$ in the denominator becoming very small and thus blowing the variance estimator "off the chart". Hence we use a bootstrap for our data, which is admittedly ad-hoc. Going further with the ad-hoc bootstrap variance-covariance estimates to obtain the optimal estimator might be too far-fetching. In addition, deriving a CB with the optimal estimator would call yet another round of bootstrap, and the whole procedure would be extremely time-consuming.

Third, for our three-stage procedure, the main explanatory power for $\rho(c, x)$ comes from the variation in weight. As weight tends to change little year to year, even if we use all the waves, the extra gain brought in by using all waves instead of two far-apart waves is likely to be small. Hence we chose the two waves 1986 and 2000 for our analysis, taking into account this "far-apartness" and sample size.

4.2 Descriptive Statistics and Densities

Table 1: Descriptive Statistics for White Females

	Wave 1986		Wave 2000	
	Mean (SD)	Min, Max	Mean (SD)	Min, Max
wage ($)	7.65 (3.99)	1, 47.7	12.7 (10.9)	1, 146
weight (lb)	139 (29.5)	84.6, 279	161 (41.8)	82.3, 572
in school	0.151 (0.358)	0, 1	0.048 (0.215)	0, 1
age youngest	0.985 (1.99)	0, 13	6.45 (6.08)	0, 27
# kids	0.651 (0.932)	0, 5	1.63 (1.20)	0, 6
married	0.474 (0.500)	0, 1	0.672 (0.470)	0, 1
married but	0.114 (0.318)	0, 1	0.233 (0.423)	0, 1
job experience	5.03 (2.54)	0, 11.4	14.0 (5.90)	0, 23.7
job tenure	2.09 (2.14)	0, 10.1	6.17 (5.80)	0, 23.1
local unemp<6	0.257 (0.437)	0, 1	0.878 (0.328)	0, 1
local unemp>9	0.296 (0.457)	0, 1	0.026 (0.160)	0, 1
white collar	0.630 (0.483)	0, 1	0.373 (0.484)	0, 1
part-time	0.788 (0.409)	0, 1	0.895 (0.307)	0, 1
north east	0.190 (0.393)	0, 1	0.174 (0.379)	0, 1
north central	0.312 (0.463)	0, 1	0.318 (0.466)	0, 1
south	0.325 (0.469)	0, 1	0.339 (0.474)	0, 1
height (inch)	64.7 (2.20)	56.2, 72.1		
age	24.6 (2.21)	21, 29		
schooling	13.1 (2.26)	0, 20		
schooling-dad	11.4 (4.25)	0, 20		
schooling-mom	11.5 (3.22)	0, 20		
intelligence	0.163 (0.908)	-3.21, 2.40		

Whereas the detailed information on the original data can be found in Cawley (2004), Table 'Descriptive Statistics for White Females' presents the mean, standard deviation (SD), minimum and maximum of the variables used in our study. The variable 'wage' is bottomcoded at \$1, 'age youngest' is the age of the youngest child, 'married' is married with the spouse present, 'married but' is married with spouse absent, 'job experience' is the years of the actual work experience, 'job tenure' is the years at the current job, 'local unemp<6' is the county unemployment rate less than 6%, 'local unemp>9' is the county unemployment rate greater than 9%, "part-time' is working less than 20 hours per week, 'schooling' is the highest grade completed, and 'intelligence' is a measure of cognitive ability from the ten Armed Services Aptitude Battery tests administered in 1980.

In addition to these regressors, dummy variables indicating whether some variables are missing or not were also used as regressors in the LSE below and in the three-stage procedure to estimate ρ, but their descriptive statistics are omitted in Table 1; those dummies will not be further mentioned. According to Cawley (2004), the sample selection problem of using only those who work does not seem to matter much, and we proceed along with this statement without further mentioning the sample selection aspect in the remainder of this paper.

As height and weight are of prime interest to our study, we drew their densities in Figure 1. The height density suggests a possibility of bimodality, which may be interpreted as two groups of heights. The weight densities show that, as the white females get older, their weight distribution gets more dispersed toward the right tail, i.e., on average the women became more obese.

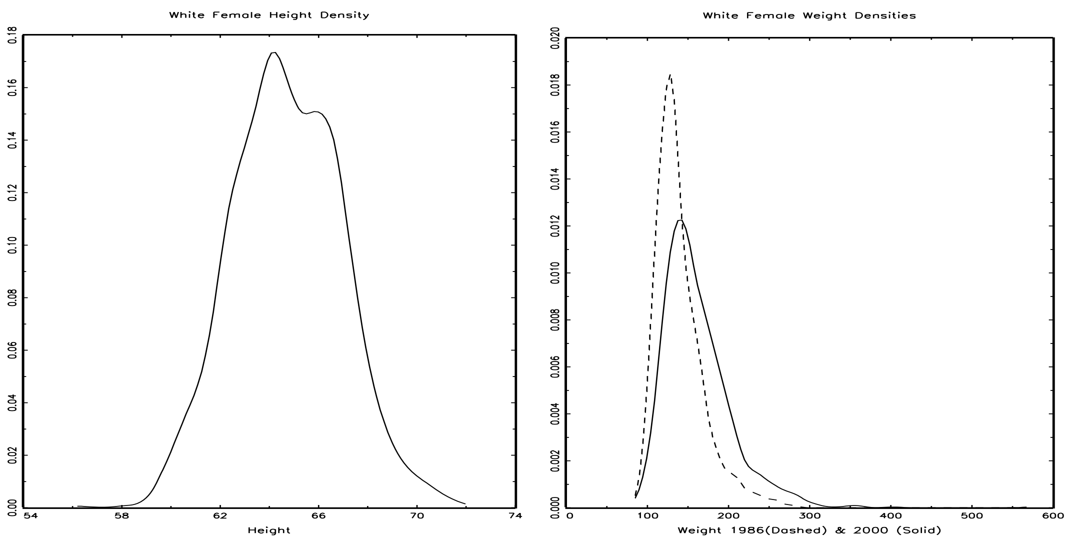

Figure 1: Density Functions for Height and Weights (1986,1990)

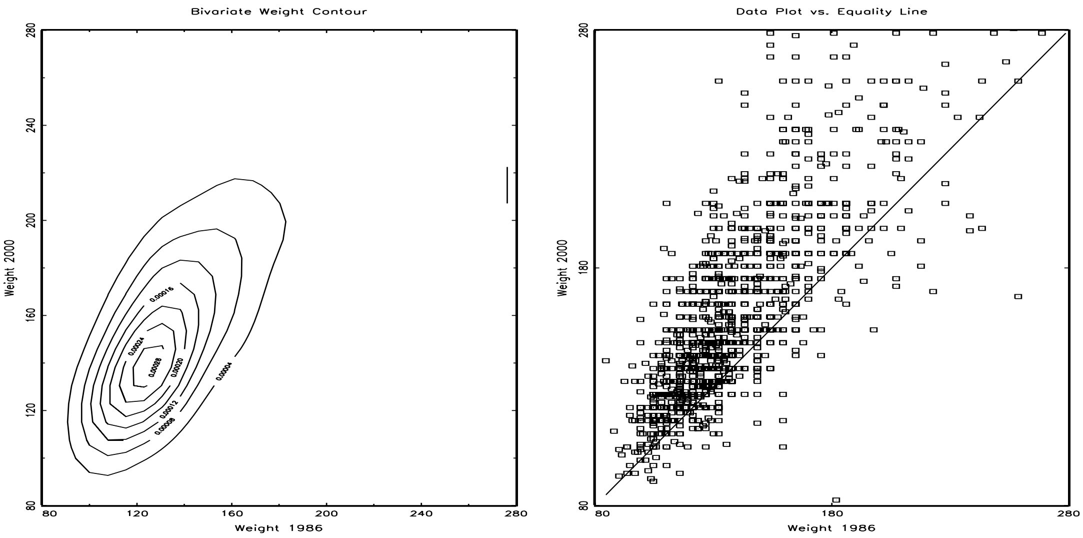

Figure 2: Weight (1986,1990) Density Contours and Data Plots

To get a more detailed understanding on weight change, Figure 2 presents the contours and data plot along with the 45% equality line for 1986 and 2000 weights. Clearly, most women became heavier over the 14 year span, and there is no other visible interesting pattern such as bimodality or only some weight-subgroup getting heavier. Heteroskedasticity depending on the earlier weight is also noticeable on the right panel. When LSE was run for 2000 weight on 1986 weight, the result is

$$\underset{SD\ (t-value):}{Weight\ 2000 =} \underset{5.15\ (1.33)}{6.857} + \underset{0.04\ (28.0)}{1.111} \times (Weight\ 1986), \quad R^2 = 0.613.$$

Adding other regressors to this equation raises R^2 somewhat but R^2 seems to fall short of 0.65 no matter what.

4.3 Panel LSE

We applied panel LSE as explained in Lee (2002); the LSE allows an arbitrary correlation for $\alpha_i + U_{i1}$ and $\alpha_i + U_{i2}$. The LSE results are in Table 2 'Panel LSE and MDE for Ln(Wage): White Females'. There are three columns with b_N (tv), among which the first two are for 1986 and 2000. The columns indicate that the estimates tend to differ across the two years. Testing for the parameter constancy, we reject the H_0 with the Wald test statistic 51.7 and p-value 0.02. But the parameter constancy for BMI and height only was not rejected with the Wald test statistic 0.416 and p-value 0.812.

Table 2: Panel LSE and MDE for Ln(Wage): White Females

Variables	b_N (tv) 1986	b_N (tv) 2000	b_N (tv) MDE
one	0.636 (0.36)	-2.735 (-0.49)	1.433 (3.24)
BMI	-0.010 (-3.64)	-0.008 (-4.52)	-0.010 (-6.23)
height	0.004 (0.72)	0.004 (0.66)	0.006 (1.24)
age	0.111 (0.87)	0.220 (0.85)	0.019 (1.03)
$\text{age}^2/100$	-0.276 (-1.11)	-0.308 (-0.98)	-0.064 (-2.33)
age×schooling	0.003 (1.13)	0.000 (0.09)	0.002 (2.95)
schooling	-0.115 (-1.69)	-0.065 (-0.44)	-0.091 (-4.55)
$\text{schooling}^2/10$	0.027 (2.83)	0.042 (1.95)	0.031 (3.88)
schooling-dad	0.005 (1.12)	0.010 (1.61)	0.005 (1.30)
schooling-mom	-0.007 (-0.97)	0.005 (0.59)	-0.002 (-0.32)
in school	-0.032 (-0.83)	-0.031 (-0.57)	-0.049 (-1.64)
intelligence	0.093 (5.02)	0.056 (2.82)	0.075 (5.34)
$\text{intelligence}^2 \times 10$	-0.001 (-0.52)	0.003 (1.83)	0.001 (0.90)
age youngest	-0.001 (-0.08)	-0.004 (-1.68)	-0.003 (-1.21)
# kids	-0.032 (-1.65)	-0.011 (-0.76)	-0.019 (-1.82)
married	-0.028 (-0.94)	0.102 (1.89)	0.007 (0.29)
married but	0.020 (0.46)	0.159 (2.78)	0.064 (2.06)
job experience	0.032 (3.59)	0.030 (5.80)	0.024 (6.24)
job tenure	0.066 (3.78)	0.050 (5.58)	0.057 (10.3)
$(\text{job tenure})^2$	-0.004 (-1.55)	-0.002 (-3.40)	-0.002 (-6.08)
local unemp<6	0.091 (2.83)	0.226 (4.67)	0.130 (5.13)
local unemp>9	-0.033 (-1.05)	0.075 (0.71)	0.001 (0.02)
white collar	0.183 (5.93)	0.091 (2.73)	0.151 (7.13)
part-time	0.088 (2.36)	-0.013 (-0.18)	0.081 (2.63)
north east	-0.066 (-1.52)	-0.072 (-1.35)	-0.081 (-2.36)
north central	-0.132 (-3.41)	-0.185 (-3.74)	-0.166 (-5.52)
south	-0.102 (-2.69)	-0.141 (-2.83)	-0.120 (-4.04)

The last column with MDE (minimum distance estimator) is obtained imposing the all-parameter constancy restriction; the numbers in this MDE column may be regarded as an weighted average of the 1986 and 2000 columns, and they are shown as a 'reference'; see Lee (2002) for the implementation details of the MDE if interested. Although the model specification differs somewhat from that in Cawley (2004), the estimate for BMI is almost the same, hovering around -0.01.

4.4 Marginal Integration for Height-Fixed Log Wage Function

In this section, we apply the three-stage procedure to estimate $\rho(c, x)$:

1. $\hat{E}(\Delta Y|C_i, X_{i1}, X_{i2})$ and $\hat{E}(W_{it}|C_i, X_{i1}, X_{i2})$, $t = 1, 2$, are estimated with the kernel method where the product of $N(0,1)$ kernels is used and the bandwidth for each regressor is $2 * SD \times N^{-1/6}$ ($= 1.7 * SD \times N^{-1/7}$ for 3-dimensional smoothing) where the multiplicative factor 2 was chosen by 'eye-balling' on the final $\rho(c, x)$ figures. In theory, high-order kernels are needed, but they performed poorly, which is why the usual normal kernels were employed.

2. LSE of $\Delta Y_i - \hat{E}(\Delta Y|C_i, X_{i1}, X_{i2})$ on $W_{it} - \hat{E}(W_{it}|C_i, X_{i1}, X_{i2})$, $t = 1, 2$, was done to get b_N. Recall that we rejected the H_0 : parameter constancy for the regressors other than BMI and height.

3. With $Q_i \equiv \Delta Y_i - (W'_{i1}, W'_{i2})b_N$, $\hat{\mu}(c, x_1, x_2)$ for $\mu(c, x_1, x_2) = E(Q|C = c, X_1 = x_1, X_2 = x_2)$ was obtained. Then $\hat{\mu}_{c1}(c, x) = N^{-1}\sum_i \hat{\mu}(c, x, X_{i2})$ and $\hat{\mu}_{c2}(c, x) = N^{-1}\sum_i \hat{\mu}(c, X_{i1}, x)$ were obtained to get the final estimator

$$\hat{m}(c, x) \equiv \frac{\hat{\mu}_{c2}(c, x) - \hat{\mu}_{c1}(c, x)}{2} \to^p \rho(c, x) - \int \rho(c, x)\frac{f_1(x) + f_2(x)}{2}dx.$$

Figure 3 presents the three functions $\hat{\mu}_{c2}(c, x)$ (top), $-\hat{\mu}_{c1}(c, x)$ (bottom) and $\hat{m}(c, x)$. Clearly $\hat{m}(c, x)$ falls halfway between the other two curves. Because of the women getting heavier over time, we have $f_2 \neq f_1$, which makes $\int \rho(c, x)f_2(x)dx \neq \int \rho(c, x)f_1(x)dx$. As the result, the levels of the top and bottom curves differ. The averaged in-between curve $\hat{m}(c, x)$ picks up the same curvature information/restriction while re-leveling the curve with the averaged density of f_1 and f_2.

Whereas the 2000 curve is monotonically declining, the 1986 curve has a trough at weight $\simeq 220$ pounds. This seems to have occurred because the 1986 curve has $-\int \rho(c, x)f_2(x)dx$, and there are many obese women with high wages in 2000. The women were young in 1986 and thus mostly employed by others getting paid relatively low, but this has changed by 2000. The trough might be an outcome of the reverse causality of high wage causing obesity. To deal with the reverse causality, Cawley (2004) instrumented BMI with a sibling BMI, but he could not reject the H_0 that the estimates do not change with or without the instrument. In addition to the trough presence, the weight effect seems slightly stronger in 1986 (when the women were young) than in 2000 (when the women were relatively old). It might be better to restrict the same functional form restriction only up to, say 190 lb. Here we also note some outliers in weight: the maximum weight is only 279 lb 1986, but there were 10 women above 300 in 2000, 2 above 400 and 1 above 500.

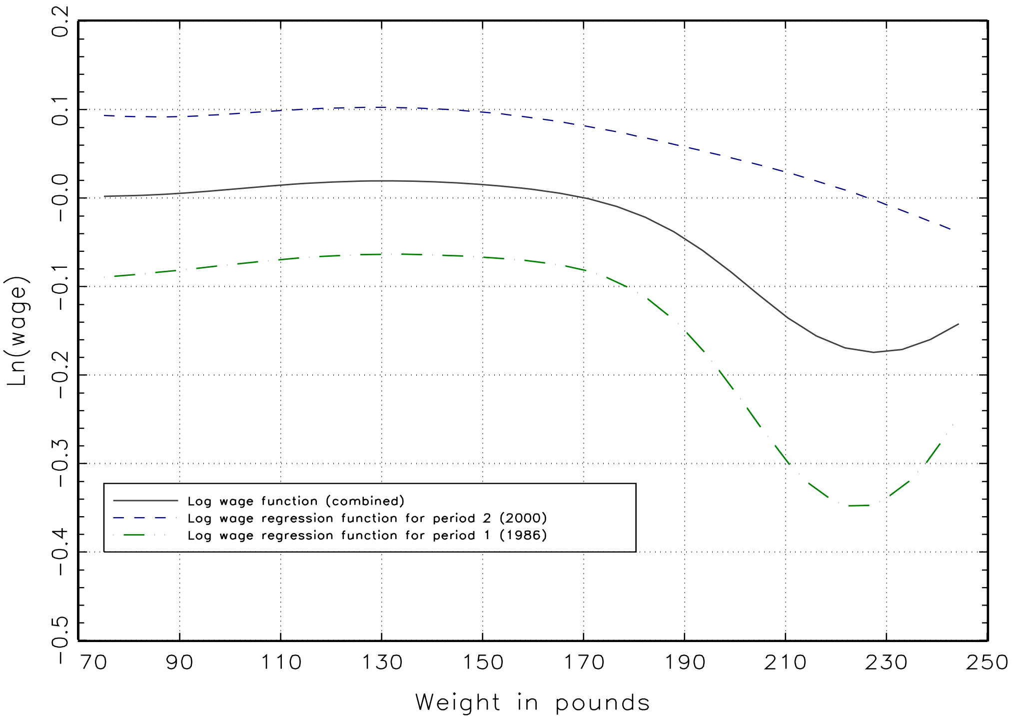

Figure 3: Ln(Wage) vs. Weight 2000 (top), 1986 (bottom) and Combined at MED Height

In drawing a CB (confidence band) around $\hat{m}(c,x)$, as already noted, the asymptotic variance estimator did not work well due to the random denominator $\hat{f}(c,x_1,x_2)$. Instead, we applied nonparametric bootstrap, resampling from the original data with replacement. One way of getting a CB is using the lower and upper 5% bootstrap quantiles at each evaluation point (c,x). But this requires a rather high bootstrap repetition number. To save time, we did the bootstrap 31 times to obtain the bootstrap $SD_{boot}(c, x)$, and the CB used is

$$\hat{m}(c, x) \pm 2.93 \times SD_{boot}(c, x), \quad \text{where } c \text{ is fixed at LQ, MED and UQ heights;}$$

i.e., in Figures 4-6, the log wage function is shown with height fixed respectively at the LQ (63.2 inches; about 160cm), MED (64.9 inches; about 165cm) and UQ (66.5 inches; about 170cm). Admittedly, there is no proof that this bootstrap is consistent in the sense of giving the correct asymptotic coverage error. But with the asymptotic variance estimator failing, there seems to be no other way to gauge the precision in estimating $m(c, x)$ with $\hat{m}(c, x)$.

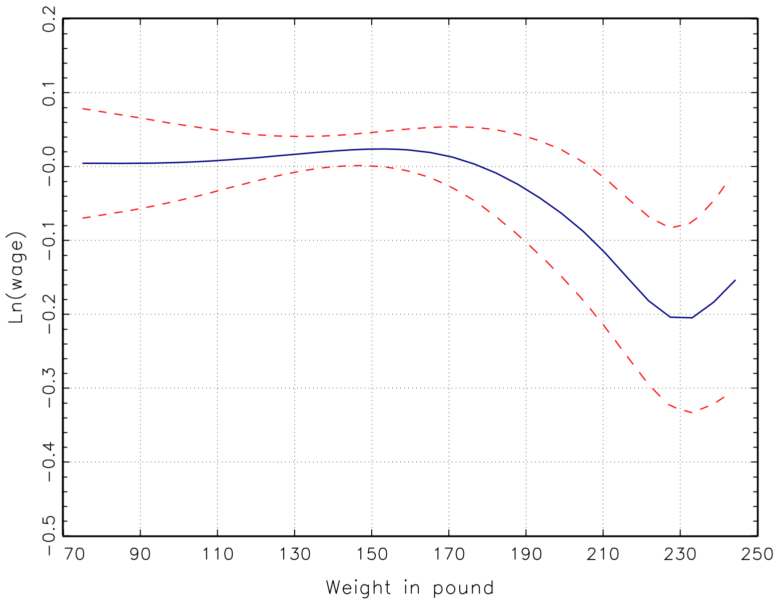

Figure 4: Ln(Wage) Function at LQ Height (about 160cm)

Looking at Figures 4-6, the log wage function appears flat (or increasing slightly) up to the average weight of 140-160 pounds, and then decreasing rapidly afterward. The curve drops more rapidly for relatively shorter women (with LQ or MED height) than taller women (with MED or UQ height), which is natural because the same amount of weight gain is more visible for shorter women. In Figure 4, it is virtually impossible to fit a linear line in the CB, implying a significant nonlinear effect; in Figure 5, it is possible to fit a linear line but not the zero line, implying a significant effect but not necessarily a non-linear effect; in Figure 6, the zero line can be fit, implying no significant effect.

Based on the linear model results as in Cawley (2004), one might conclude that there is a wage gain in reducing weight regardless of the current weight, implying a wage gain

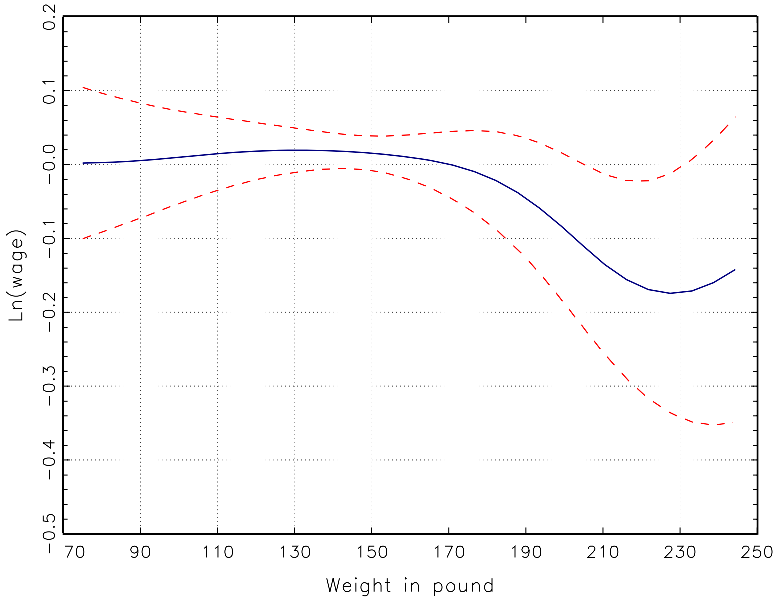

Figure 5: Ln(Wage) Function at MED Height (about 165cm)

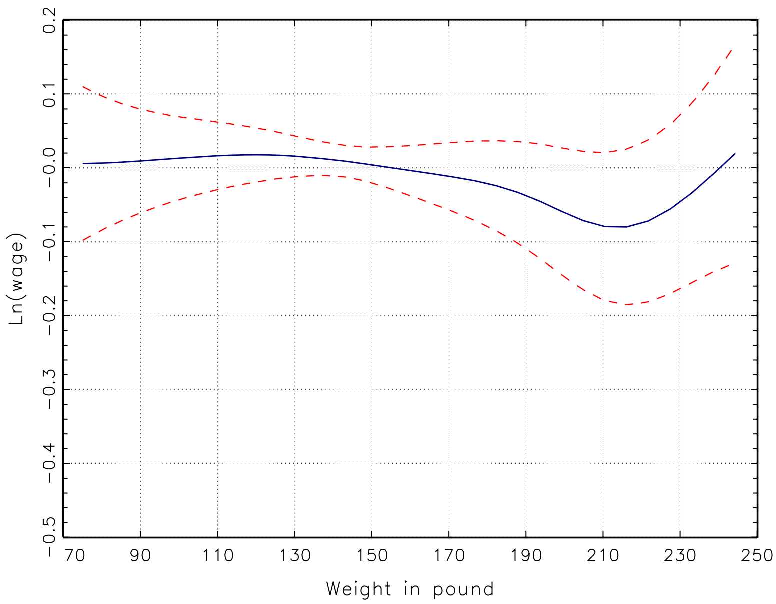

Figure 6: Ln(Wage) Function at UQ Height (about 170cm)

in becoming slimmer than the normal (average) weight. But our nonparametric estimator paints a different picture: there is a wage gain only for those over-weight, but there is no wage gain (or even a loss) for those with the average or lower weight. Recall the main finding of Cawley (2004): 65 pound loss for 9% wage gain. Figures 4-6 show the gain of about 20%, 17% and 8% over the 60 pound loss from about 220 to 160 lb. Hence the weight-loss gain for those over-weight is greater than what the linear model suggests. This is because essentially *the linear model estimate is a mixture of the zero effect over the under-weight range and the negative effect over the over-weight range*. This is one of the most interesting findings from the nonparametric method—something one could not have foreseen using only the linear model.

As an obesity measure, BMI is not bad at all for while females but poor for males, as observed in Burkhauser and Cawley (2008). Cawley (2004) could not find any BMI effect for white males using linear models. In the appendix, we apply the same nonparametric three-stage procedure to the while males in the NLSY data to see indeed no significant findings. In a sense, this provides a support to our methodology. Despite no significant findings, some interesting features do exist for white males; see the appendix if interested.

5 Conclusions

Does obesity matter for wage? The answer is yes at least for white females in the U.S. when body mass index (BMI) is used as a measure of obesity; this was the finding from the conventional linear models. But BMI is a rather special—in fact, too tightly specified—function of weight and height. As popular as BMI may be, it is highly unlikely that the single functional form is suitable for different response variables for which BMI have been used, and this means that there is likely to be a better functional form than BMI that relates weight and height to wage. In this paper, we posited a semi-linear panel data model where the model has a nonparametric function $\rho(C_i, X_{it})$ of height C_i and weight X_{it} and a linear function of the other regressors. After removing the unit-specific effect by first-differencing the model, in essence, we ended up with $\mu(C_i, X_{i1}, X_{i2}) = \rho(C_i, X_{i2}) - \rho(C_i, X_{i1})$. The main task was then to recover ρ from a nonparametric estimator of μ, imposing the restriction that μ consists of two functions of the same form. We did the task using the nonparametric marginal integration approach. Differently from the linear model finding, we found no evidence of wage gain from weight loss for normal- (or under-) weight women. Also, for over-weight women, the gain from weight loss is greater than what the linear model suggested.

Our study warns against using BMI in a blindfolded fashion. Rather, for each response variable of interest, there should be a more suitable function of weight and height than BMI, which should be sought after. When the response variable is related to some illness/disease, although the difference between the "advices" from the linear and the semi-linear models may not be much for a single individual, giving the wrong advice using the linear function may amount to many lives lost unnecessarily for the entire population.

APPENDIX

This appendix applies the same three-stage procedure to the white males in the NLSY data for waves 1986 and 2000. Descriptive statistics of the white males are in Table 3. As in the white females, the white males also gained weight over the 14 year span.

Table 4 displays the panel LSE and MDE results, which indicate that there is no significant linear relationship between ln($wage$) and BMI. Figure 7 corresponding to Figure 3 shows the graphs for 2000 (top), 1986 (bottom) and the combination (middle). It is interesting to see that the under-weight are penalized more severely in 2000 than in 1986, whereas the over-weight are penalized less severely in 2000 than in 1996. This may be explained by the growing trend of valuing fitness and muscle: the under-weight is penalized more these days due to lack of muscles, but the over-weight is not because muscle weight can take a higher proportion of weight in 2000 than in 1986.

Figures 8-10 corresponding to Figures 4-6 show negative effects for being under-weight as well as being over-weight. However, differently from the white female case, the zero line can fit in all three figures, implying no significant effects in all figures. Nevertheless, if we look at the estimated lines in Figures 8-10, then the following features are notable.

First, a wage gain of about 12% from 50 lb loss from about 280 to 230 can be seen. Differently from the white females, this feature is similar across the LQ, MED and UQ heights, possibly because the interquartile range is small (about 6 cm) for males compared with that for females (about 10cm). Second, the starting weight for wage loss is about 230 that is well above the 2000 average weight 197. This is in sharp contrast to the females, for whom the starting weight for wage loss is about 160 that is almost the same as the 2000 average weight 161. Third, the wage loss from being under-weight is more visible than for females; for females, hardly any wage loss was seen from being under-weight.

Table 3: Descriptive Statistics for White Males

	Wave 1986		Wave 2000	
	Mean (SD)	Min, Max	Mean (SD)	Min, Max
wage ($)	2.11 (0.497)	1, 4.78	2.58 (0.650)	1, 6.21
weight (lb)	175 (30.2)	108, 348	197 (38.7)	73.2, 471
in school	0.137 (0.344)	0, 1	0.018 (0.134)	0, 1
age youngest	0.533 (1.577)	0, 21	4.55 (5.36)	0, 29
# kids	0.415 (0.759)	0, 1	1.58 (1.31)	0, 9
married	0.403 (0.491)	0, 1	0.679 (0.467)	0, 1
married but	0.060 (0.238)	0, 1	0.181 (0.385)	0, 1
job experience	5.24 (2.68)	0, 11.4	14.4 (6.71)	0, 23.9
job tenure	2.32 (2.40)	0, 12.4	7.22 (6.30)	0, 24.9
local unemp<6	0.273 (0.446)	0, 1	0.876 (0.329)	0, 1
local unemp>9	0.280 (0.449)	0, 1	0.023 (0.151)	0, 1
white collar	0.357 (0.479)	0, 1	0.221 (0.415)	0, 1
part-time	0.841 (0.366)	0, 1	0.982 (0.134)	0, 1
north east	0.191 (0.393)	0, 1	0.176 (0.381)	0, 1
north central	0.340 (0.474)	0, 1	0.345 (0.476)	0, 1
south	0.289 (0.454)	0, 1	0.307 (0.461)	0, 1
height (inch)	69.9 (2.38)	62.6, 78.7		
age	24.6 (2.27)	21, 29		
schooling	13.0 (2.30)	0, 20		
schooling-dad	11.6 (4.35)	0, 20		
schooling-mom	11.5 (3.28)	0, 20		
intelligence	0.119 (0.947)	-3.53, 2.10		

Table 4: Panel LSE and MDE for Ln(Wage): White Males

Variables	b_N (tv) 1986	b_N (tv) 2000	b_N (tv) MDE
one	-0.997 (-0.70)	5.130 (1.15)	1.071 (2.71)
BMI	0.000 (0.02)	-0.004 (-1.55)	-0.003 (-1.40)
height	0.013 (2.82)	0.001 (0.27)	0.010 (2.78)
age	0.149 (1.42)	-0.125 (-0.57)	0.046 (2.79)
age^2/100	-0.266 (-1.24)	0.111 (0.39)	-0.138 (-5.56)
age×schooling	0.001 (0.25)	0.001 (0.56)	0.004 (6.74)
schooling	-0.038 (-0.60)	-0.148 (-1.31)	-0.131 (-5.80)
schooling^2/10	0.018 (1.69)	0.056 (3.23)	0.023 (2.58)
schooling-dad	-0.001 (-0.33)	0.012 (2.16)	0.004 (1.18)
schooling-mom	-0.006 (-1.06)	-0.002 (-0.21)	-0.002 (-0.38)
in school	-0.185 (-4.53)	-0.108 (-0.94)	-0.177 (-4.82)
intelligence	0.085 (4.97)	0.122 (5.93)	0.096 (6.94)
intelligence^2×10	0.001 (0.51)	0.000 (-0.03)	0.000 (0.31)
age youngest	0.000 (-0.03)	-0.003 (-1.22)	-0.002 (-0.87)
# kids	0.020 (0.97)	0.016 (1.18)	0.029 (2.64)
married	0.055 (1.90)	0.190 (4.08)	0.091 (3.86)
married but	0.067 (1.42)	0.107 (2.10)	0.051 (1.57)
job experience	0.022 (2.74)	0.036 (5.32)	0.018 (4.74)
job tenure	0.069 (4.68)	0.029 (3.21)	0.035 (6.59)
(job tenure)^2	-0.005 (-3.10)	-0.001 (-2.25)	-0.001 (-4.60)
local unemp<6	0.017 (0.55)	0.199 (4.42)	0.047 (2.03)
local unemp>9	-0.109 (-3.99)	0.023 (0.25)	-0.078 (-3.18)
white collar	0.032 (1.13)	0.123 (3.09)	0.070 (3.23)
part-time	0.089 (2.28)	-0.319 (-1.47)	0.084 (2.29)
north east	-0.003 (-0.07)	-0.018 (-0.36)	0.009 (0.28)
north central	-0.080 (-2.36)	-0.143 (-3.07)	-0.082 (-2.94)
south	-0.090 (-2.57)	-0.091 (-1.98)	-0.066 (-2.33)

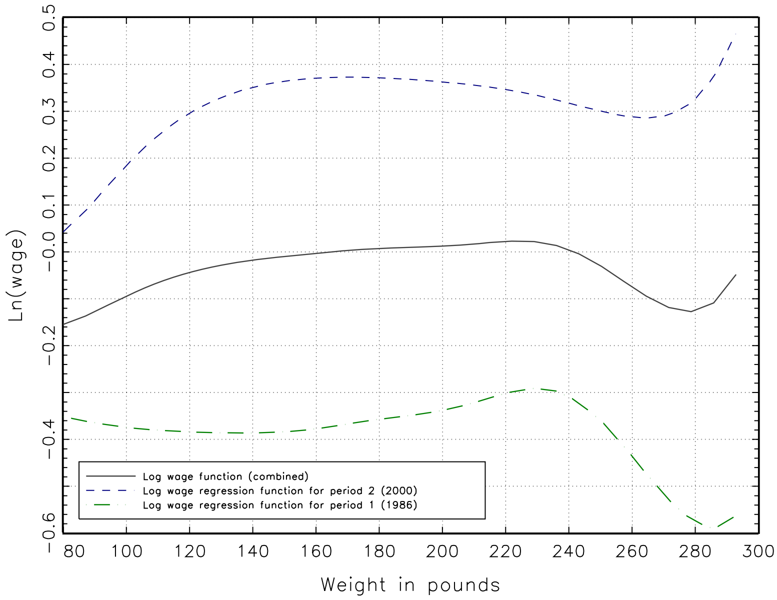

Figure 7: Ln(Wage) vs. Weight 2000 (top), 1986 (bottom) and Combined at MED Height

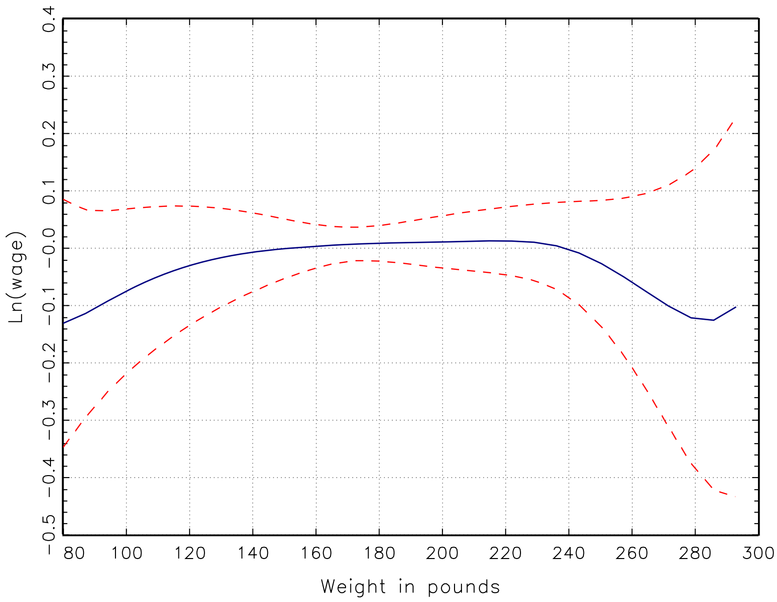

Figure 8: Ln(Wage) Function at LQ Height (about 175cm)

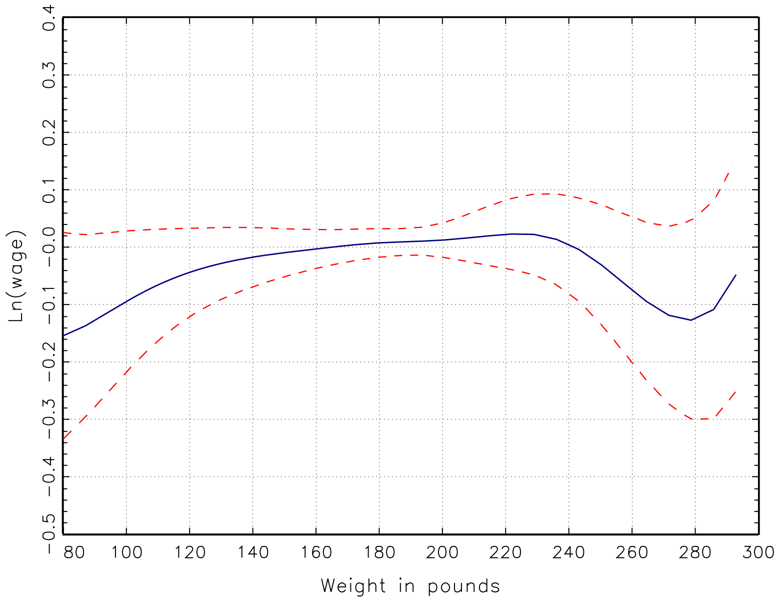

Figure 9: Ln(Wage) Function at MED Height (about 178cm)

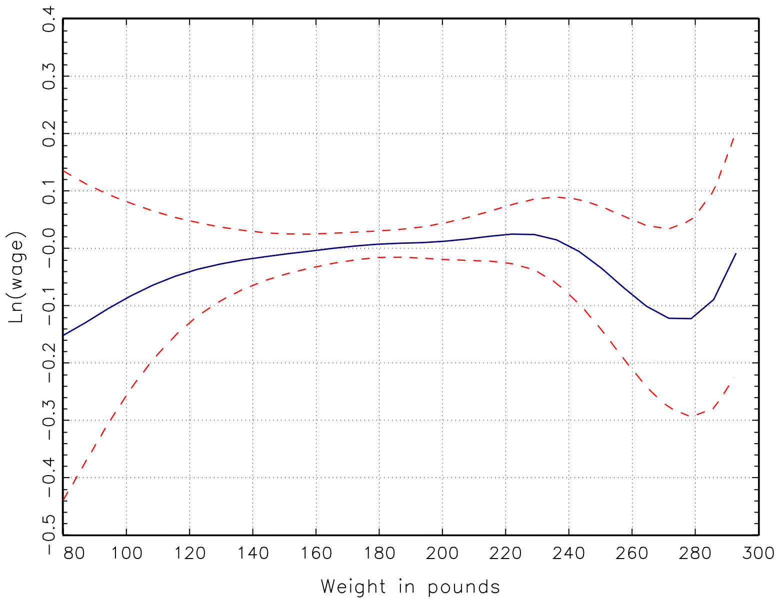

Figure 10: Ln(Wage) Function at UQ Height (about 181cm)

REFERENCES

Burkhauser, R.V. and Cawley, 2008, Beyond BMI: the value of more accurate measures of fatness and obesity in social science research, Journal of Health Economics 27, 519-529.

Cawley, J., 2004, The impact of obesity on wages, Journal of Human Resources 39, 451-474.

Chamberlain, G., 1992, Comment: sequential moment restrictions in panel data, Journal of Business and Economic Statistics 10, 20–26.

Henderson, D.J., R.J. Carroll and Q. Li, 2008, Nonparametric estimation and testing of fixed effects panel data models, Journal of Econometrics 144, 257–275

Holtz-Eakin, D., W. Newey and H. S. Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371–1395.

Honoré, B. E., 1992, Trimmed LAD and LSE of truncated and censored regression models with fixed effects, Econometrica 60, 533–565.

Kyriazidou, E., 1997, Estimation of a panel data sample selection model, Econometrica 65, 1335–1364.

Lee, M. J., 1999, A root-N consistent semiparametric estimator for related-effect binary response panel data, Econometrica 67, 427–434.

Lee, M.J., 2002, Panel data econometrics: methods-of-moments and limited dependent variables, Academic Press.

Lee, M.J., 2010, Micro-econometrics: methods of moments and limited dependent variables, Springer, forthcoming.

Li, Q., and T. Stengos, 1996, Semiparametric estimation of partially linear panel data models, Journal of Econometrics 71, 389-397.

Lin, X. and R.J. Carroll, 2006, Semiparametric estimation in general repeated measures problems, Journal of the Royal Statistical Society (Series B) 68, 69–88.

Linton, O.B., 1997, Efficient estimation of additive nonparametric regression models, Biometrika 84, 469-473.

Linton, O.B. and J. P. Nielsen, 1995, A kernel method of estimating structured nonparametric regression based on marginal integration, Biometrika 82, 93–100.

Manski, C. F., 1987, Semiparametric analysis of random effects linear models from binary response panel data, Econometrica 55, 357–362.

Müller, H. G., 1988, Nonparametric regression analysis of longitudinal data, Springer.

Newey, W., 1994, Kernel estimation of partial means and a general variance estimator, Econometric Theory 10, 233-253.

Robinson, P.M., 1988, Root-N consistent semiparametric regression, Econometrica 56, 931-954.

Wooldridge, J. M., 1997, Multiplicative panel data models without the strict exogeneity assumption, Econometric Theory 13, 667–678.