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

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

Myoung-jae Lee*
Dept. Economics, Korea University
Seoul 136-701, South Korea
myoungjae@korea.ac.kr

Body mass index (BMI), weight(kg)/ height(m)^2, is a widely used measure of obesity in medical science. In economics, there appeared studies 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 a priori reason to believe why this 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 (and white males); 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. This negative effect's magnitude is greater than that in Cawley (2004) who used a linear BMI model. The linear model gives the incorrect impression of the presence of a 'wage gain' by becoming slimmer than the average and of a 'wage loss' that is less than what it actually is when going above the average.

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

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

* Corresponding author: 82-2-3290-2229 (phone); 82-2-926-3601 (fax).

** The authors are grateful to Thierry Magnac and three anonymous reviewers for their detailed comments. The authors are also grateful to those who gave comments at seminars in various institutions, and also to the participants of the 18th European Workshop on Econometrics and Health Economics and the 2008 Korean Econometric Society Meeting.

1 Introduction

Obesity is a major and exacerbating public health problem. What is worrying is that, not only the level of obesity, but also its growth rate have been accelerating. With obesity defined as Body Mass Index (BMI) equal to or greater than 30 where BMI is weight in kg divided by height in meters squared, the age-adjusted prevalence of obesity was 22.9% in the U.S. during 1988–1994, which then escalated to 30.5% in 1999–2000 (Flegal et al. 2000). In other western developed countries, the prevalence of obesity has increased from around 10% during the mid 1980s to around 40% in the mid 1990s (Seidell and Flegal 1997). The epidemic of obesity is also an emerging problem in developing, transitional and newly industrialized countries (see, e.g., Dobson 2005).

There exist many early studies for obesity effects on wage/income, and the majority of them used cross-section data: Register and Williams (1990), Gortmaker et al. (1993), Loh (1993), Sargent and Blanchflower (1994), Hamermesh and Biddle (1994) and Pagan and Davila (1997). These early studies paid less attention to endogeneity problems caused by some omitted variables affecting both obesity and wage; e.g., individual time preference. A present-oriented individual who discounts future consumption heavily is more likely to be overweight due to their tendency to seek immediate gratification. These individuals tend to invest less in human capital and consequently will have lower wages. Another example is individual ability. Individuals with higher ability may do a better job of keeping themselves in shape and have higher wages. The failure to account for this kind of omitted variables casts doubt on the early studies' empirical findings.

Recent studies such as Averett and Korenman (1996), Cawley (2004), Baum and Ford (2004) and Morris (2006) dealt with the endogeneity problems using panel data or instrumental variable estimators (IVE), which made these studies more credible than the earlier ones. Although their findings differ much across data, models and econometric methods, a common finding seems to be that the obesity effects on wage are more salient and statistically significant for females than for males. For instance, Morris (2006) found that being obese or not, but not the degree of obesity, has a negative effect on male wage, but that both being obese or not and the degree of obesity have a negative effect on female wages. As will be discussed further below, insignificant (or weak even if significant) effects for males seem due to BMI (or weight) being a noisy measure of 'fatness' for men.

More recently, there appeared studies examining nonlinearity in the relationship between wage and BMI. Kline and Tobias (2008) set up a 'triangular model' to allow for endogenous BMI where ln(wage) equation is semi-linear with a nonparametric BMI component and the BMI reduced form is linear. They applied a Bayesian approach to their 1970 British Cohort Study data. Although their triangular model is semi-parametric, the Bayesian approach is heavily parametric. They found evidence of a nonlinear relationship between BMI and ln(wage). Gregory and Ruhm (2009) also used a semi-linear model where the model is nonparametric only for BMI; they applied the two-stage procedure in Robinson (1988) and used PSID-drawn data. The endogeneity problem was dealt with by a 'control function approach', where the residual from a linear regression of BMI on the instruments (and others) is used to control for the endogeneity source. As in other studies, they found different gender patterns. For example, there is a steep wage loss when BMI increases from around 23 to 27 for white females, but then the wage function becomes almost flat when BMI goes above 27. For white men, wage increases monotonically until BMI reaches 25, and then it decreases steadily afterwards.

Although BMI has been used as a primary measure of obesity, there are a couple of reasons why BMI is not an ideal measure of obesity. First, by using weight, not 'fatness', BMI cannot tell muscle weight from fat weight (Burkhauser and Cawley 2008 and Johansson et al. 2009). A muscular man with a high weight is not the same as an obese man with the same weight; their productivity, health care costs and physical attractiveness would differ much. Second, BMI is a tightly specified function of weight and height. Considering so many different ways how weight and height can be combined, it is surprising how the particular functional form has been so widely used. Even if we take the ratio form in BMI, why should the height exponent be 2, not something else, say 2.5 or 3?

The goal of this paper is to find weight effects on wage, for which we use a subset of the NLSY data originally used in Cawley (2004)—more details on the data will be provided later and we are grateful to Professor Cawley for generously providing us the data. Weight can be a good measure of obesity for women as muscular women are rare, whereas that may not be the case for men as there are many muscular men. Hence, as in most studies, we obtain statistically significant results for women, but not for men. Although our analysis will show weight effect on wage, it will not show how specifically weight influences wage. That is, our analysis will not show through which routes (productivity, health care costs, physical attractiveness,...) weight influences wage. These routes are causal paths linking weight to wage.

Suppose weight has a significant effect on wage and we have a productivity variable. If the productivity variable is significant when included in the model while weight loses its significance, then weight affects wage only through productivity. If both the productivity variable and weight are significant but the weight effect magnitude decreases, then roughly speaking, weight influences wage through productivity and some other routes as in the path diagram below.

	↗	productivity	⟶	⟶	↘	
weight	⟶	health care costs	⟶	⟶	⟶	wage
	↘	physical attractiveness	⟶	employer discrimination	↗	
			↘	client preference	↗	

The rest of this paper is organized as follows. Section 2 maps out our econometric methodologies. Section 3 shows our empirical analyses. Finally, Section 4 draws conclusions. Presenting our main conclusion in advance, we find no weight effect on wage up to the average weight, beyond which there is a large negative effect. That is, for a given height, the wage function is flat over the under-weight range relative to the mean weight 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 BMI effect's magnitude 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 large negative effect 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 for the over-weight.*

2 Methods for Related-Effect Semi-Linear Panel Model

2.1 Model and Estimation Strategy

Consider a semi-linear 'related-effect' panel data model:

$$Y_{it} = \rho(C_i, X_{it}) + W_{it}'\beta + \alpha_i + U_{it}, \qquad i = 1, \ldots, N, \quad t = 1, 2$$

where Y_{it} is a response variable (ln(wage)), C_i is a time-constant regressor (height), X_{it} is a time-variant regressor (weight), $\rho(C_i, X_{it})$ is an unknown function of C_i and X_{it}, W_{it} is other covariates with parameter β, 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

$$E(U_{it}|C_i, X_{i1}, X_{i2}, W_{i1}, W_{i2}) = 0, \quad t = 1, 2;$$

where i indexes individuals and t indexes time periods. Assume iid (independent and identically distributed) across i.

While BMI specifies $\rho(c, x)$ as proportional to x/c^2, this is a very tight specification. The 'related-effect' semi-linear panel model relaxing this restriction is relevant also for growth curve estimation as in Müller (1988) 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 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 the last display, regressors in all periods 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), Wooldridge (1997) and Arellano and Carrasco (2003).

To remove the possibly endogeneity-causing α_i, we difference the model to obtain

$$\Delta Y_i = \mu(C_i, X_{i1}, X_{i2}) + \Delta W_i'\beta + \Delta U_i \quad \text{where}$$
$$\mu(C_i, X_{i1}, X_{i2}) \equiv \rho(C_i, X_{i2}) - \rho(C_i, X_{i1})$$
$$\Delta Y_i \equiv Y_{i2} - Y_{i1}, \;\; \Delta W_i \equiv W_{i2} - W_{i1} \;\; \text{and} \;\; \Delta U_i \equiv U_{i2} - U_{i1}.$$

Our estimation of $\rho(\cdot)$ proceeds in multiple-stages as follows.

The first step is to obtain an estimator b_N for β using the Robinson's (1988) approach for semi-linear models. Take $E(\cdot|C_i, X_{i1}, X_{i2})$ on the differenced model to get

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

Subtract this from the differenced model to get

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

Replace the unknown conditional means by kernel nonparametric estimators, say $\hat{E}(\Delta Y|C_i, X_{i1}, X_{i2})$ and $\hat{E}(\Delta W_i|C_i, X_{i1}, X_{i2})$, and then obtain the Least Squares Estimator (LSE) b_N for β.

The second step is to estimate $\mu(C_i, X_{i1}, X_{i2})$ by a kernel regression of the 'residual'

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

on (C_i, X_{i1}, X_{i2}); let $\hat{\mu}$ be the kernel estimator for μ. Since b_N is $\sqrt{N}$-consistent while $\hat{\mu}$ is $\sqrt{Nh^2}$-consistent where h is the kernel bandwidth, β is as good as known as far as $\hat{\mu}$ is concerned. If desired, we can allow β to vary over time: replace $\Delta W_i'\beta$ with $W_{i2}'\beta_2 - W_{i1}'\beta_1$ to use both W_{i1} and W_{i2} as regressors.

The third step is to find ρ from μ (i.e., $\hat{\rho}$ from $\hat{\mu}$). For this, we will use the 'marginal integration estimator (MIE)' in Linton and Nielsen (1995) and Newey (1994). In MIE, the first component $\rho(C_i, X_{i2})$ in $\mu(C_i, X_{i1}, X_{i2})$ is "removed" by integrating $\mu(C_i, X_{i1}, X_{i2})$ over X_{i2} and the second component $\rho(C_i, X_{i1})$ is removed by integrating $\mu(C_i, X_{i1}, X_{i2})$ over X_{i1}. Kondo and Lee (2002) dealt with a similar problem but they removed each component by differentiation, not by integration. Integration tends to be a better operation than differentiation as integration smooths out things and speeds up the convergence rate. We thus adopt integration in this paper. Once the two MIE's for ρ are obtained, we combine them into a single estimator $\hat{\rho}$. Intuitively, this is done by a (weighted) average of the two MIE's. This averaging approach with MIE's can be seen in Azomahou et al. (2006), whereas an analogous approach based on differentiation appeared in Kondo and Lee (2002). The averaging halves the asymptotic standard deviation (SD), which is a substantial efficiency gain.

Further examining literature related to our model and estimation strategy, Li and Stengos (1996) extended the Robinson's (1988) approach for cross-section semi-linear models to panel data, where α_i is assumed to be unrelated to the regressors in the nonparametric component although they examined how to relax this restriction in a paragraph. Lin and Carroll (2006) studied semi-linear panel models where ρ is a function of time-varying regressors. A differenced model was considered briefly for matched-pair samples, but 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, however, the model with a time-constant C_i non-separable from X_{it} has not been specifically examined in the literature. Also MIE to be explained in the next section is much simpler than the profile-based kernel methods. The theoretical literature for MIE is large, but the applied econometric literature as in Rodriguez-Poo et al. (2005) and Bontemps et al. (2008)

seems scarce.

2.2 Marginal Integration Estimators (MIE) and Their Average

As β is as good as known for $\hat{\mu}$, we focus on the model without W_{it} to ease exposition:

$$Y_{it} = \rho(C_i, X_{it}) + \alpha_i + U_{it} \implies \Delta Y_i = \mu(C_i, X_{i1}, X_{i2}) + \Delta U_i.$$

Define

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

Let $f_z(z)$ be the density function for $Z = z$; denote the components of Z also as Z_1 $(= C)$, Z_2 $(= X_1)$, and Z_3 $(= X_2)$.

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 (ϕ's) for K: $K(z) = \phi(z_1)\phi(z_2)\phi(z_3)$. When W_{it} is in use, we just have to replace ΔY_{it} with $Q_{it} = \Delta Y_{it} - \Delta W_{it}' b_N$ in $\hat{\mu}(c, x_1, x_2)$.

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 $p = 0.5, 1, 2, 3$ where '7' comes from 4 plus the number of regressors; in general, $pN^{-1/(k+4)}$ is used with k being the number of regressors. When the dimension of z is small, the best practical way to choose p is plotting $\hat{f}_z(z)$; if the dimension of z is large, 'cross-validation' can be used. In our case, z is three-dimensional, and thus we can plot 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 $\hat{\mu}(z) = \hat{g}_z(z)/\hat{f}_z(z)$, we get two MIE's:

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

where f_t denotes the X-density for time t.

The averaging estimator of the two MIE's to take the advantage of the additive structure of μ in ρ is, with $x = x_1 = x_2$,

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

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 consistent for $\rho(c,x)$ up to a function of c. But, by combining the two estimators as in this display, due to the constant $1/2$, the resulting estimator has SD twice smaller than when only a single estimator is used.

If $\rho(c,x) = \alpha x/c^2$ as in BMI for some parameter α, then

$$\int \rho(c,x)\frac{f_1(x)+f_2(x)}{2}dx = \alpha\frac{1}{c^2}\int x\frac{f_1(x)+f_2(x)}{2}dx$$
$$\Longrightarrow \quad m(c,x) = \alpha\frac{x-\mu_{12}}{c^2} \quad \text{where} \quad \mu_{12} \equiv \int x\frac{f_1(x)+f_2(x)}{2}dx.$$

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

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, then fixing c means that the resulting graphs for $m(c,x)$ should be linear in x because $\rho(c,x) = \alpha x/c^2$. In our empirical analysis later, we will fix c at the lower quartile (LQ), median (MED) and upper quartile (UQ).

Although we assumed that the same functional form ρ holds in two periods, we can in fact easily allow a time-varying intercept, say τ_t:

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

where $\Delta\tau \equiv \tau_2 - \tau_1$. But $\Delta\tau$ will get cancelled out in the difference $\hat{\mu}_{c2}(c,x) - \hat{\mu}_{c1}(c,x)$. This shows that a time-varying intercept is allowed in our approach.

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 $f_1 = f_2$,

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

then $\int m(c,x) f_0(x) dx = 0$ by construction. With $\rho(c,x) = \alpha x/c^2$ in BMI, $m(c,x) = \alpha\{x - E(x)\}/c^2$ where $E(x) = \int x f_0(x) dx$.

2.3 Asymptotic Distribution of Averaged Estimator

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

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

$$\text{where} \quad \tilde{m}_1(x_1) \equiv \int \hat{m}(x_1, x_2) q(x_2) dx_2 \text{ for a weighting function } q(x_2),$$

$\hat{m}(x_1, x_2)$ 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, i.e., $\tilde{m}_1(x_1)$ 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

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

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, p.610-613).

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

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

As the two estimators are asymptotically independent, it holds that

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

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,

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

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

$$\hat{m}(c,x)\pm1.96\int L(s)^2ds\cdot\{\frac{\hat{V}_{1cx}+\hat{V}_{2cx}}{4Nh^2}\}^{1/2}.$$

$\int L(s)^2ds$ is a known number; with $L(\cdot)$ being the $N(0,1)$ density, $\int L(s)^2ds\simeq0.283$. As noted already, when three bandwidths $\hat{\sigma}_c h_0$, $\hat{\sigma}_{x1}h_0$ and $\hat{\sigma}_{x2}h_0$ are used for C_i, X_{i1} and X_{i2}, the last display becomes

$$\hat{m}(c,x)\pm1.96\int L(s)^2ds\cdot\{\frac{\hat{V}_{1cx}}{4N\hat{\sigma}_c\hat{\sigma}_{x1}h_0^2}+\frac{\hat{V}_{2cx}}{4N\hat{\sigma}_c\hat{\sigma}_{x2}h_0^2}\}^{1/2}.$$

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 use 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 Empirical Analysis: Weight Effect on Wage

3.1 Two Far-Apart Waves for White Females

Our study uses the same data as used in Cawley (2004). The panel data used in Cawley (2004) has about 100,000 observations when pooled. Cawley found that only white females have statistically significant negative weight effects on wage that are stable across different models and estimators. His main finding based on linear models is that *two SD (roughly 65 pounds) weight difference is associated with 9 percent wage difference.*

BMI is an error-ridden measure of obesity, and the measurement error is likely to bias the estimated BMI effect toward zero. To avoid this problem, Burkhauser and Cawley (2008) used body fat percentage (and some other obesity indicators). But Burkhauser and Cawley (2008) still found that BMI is a reliable measure for white females' 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 white females in our empirical analysis (and white males in the appendix to some extent). 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 obtain 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 discussed later in detail, unfortunately, the asymptotic variance and its estimator presented above do not work well for our data; the main problem seems to be $\hat{f}(c, x_1, x_2)$ in the denominator, which becomes very small, and 'trimming' (removing observations with too small $\hat{f}(c, x_1, x_2)$'s) provided little help although it is supposed to work in theory. 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.

3.2 Descriptive Statistics and Densities

Whereas the detailed information on the original data can be found in Cawley (2004), Table 'Descriptive Statistics for White Females' presents the mean, SD, minimum and maximum of the used variables. The variable 'wage' is bottom-coded 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.

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		

Dummy variables indicating whether job experience, job tenure, local unemployment rate, schooling-dad, schooling-mom and intelligence are missing or not were also used as regressors (the former three are time-varying while the latter three are not) in the linear model estimators and in estimating 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.

As height and weight are of prime interest to our study, we drew their densities in Figure 1. 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 heavier.

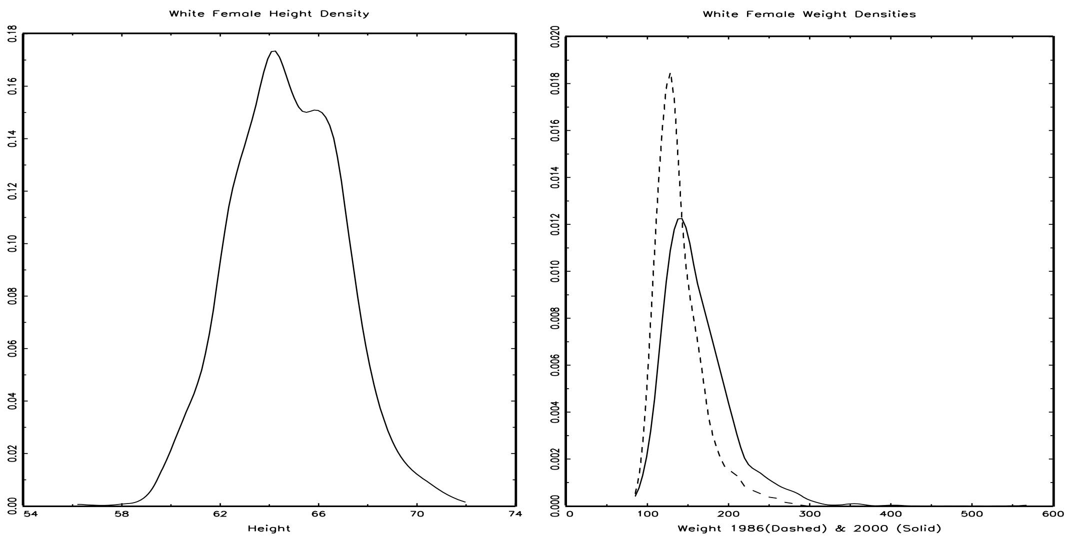

Figure 1: Density Functions for Height and Weights (1986,1990)

3.3 Panel LSE

We applied panel LSE as explained in Lee (2002); the LSE allows an arbitrary correlation between $\alpha_i + U_{i1}$ 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 Wald test statistic 51.7 and p-value 0.02. But the parameter constancy for BMI and height only was not rejected with Wald test statistic 0.416 and p-value 0.812. 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 MDE implementation details if interested.

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)

Although the model specification differs somewhat from that in Cawley (2004), the estimate for BMI is not much different, hovering around -0.01. Fixing height at the average height 1.65m, as BMI increases by 1 units (i.e., weight increases by 1.65^2=2.72kg), the wage decreases by 1%, which translates to 3.7% reduction when weight increases by 10kg—somewhat larger than about 3% in Cawley (2004).

3.4 Main Empirical Results

Before presenting the nonparametric results, we summarize its implementation:

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); the multiplicative factor 2 was chosen by 'eye-balling' on the final $\rho(c, x)$ figures. Then the LSE of $\Delta Y_i - \hat{E}(\Delta Y|C_i, X_{i1}, X_{i2})$ on $W_{it} - \hat{E}(W_{it}|C_i, X_{i1}, X_{i2})$, $t = 1, 2$, is done for b_N.

2. 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.

3. Then MIE's $\hat{\mu}_{c1}(c, x) = N^{-1}\sum_i \hat{\mu}(c, x, X_{i2})$ and $\hat{\mu}_{c2}(c, x) = N^{-1}\sum_i \hat{\mu}(c, X_{i1}, x)$ were computed to get the final averaging estimator $\hat{m}(c, x) = 0.5\{\hat{\mu}_{c2}(c, x) - \hat{\mu}_{c1}(c, x)\}$.

In Step 1, high-order kernels are needed but they performed poorly, which is why the $N(0, 1)$ kernels were employed. For W_{it}, we used only the time-varying regressors. In view of the test that rejected 'H_0 : parameter constancy for the regressors other than BMI and height' with p-value 0.2, we used (W_{i1}, W_{i2}) instead of ΔW_i for a slightly more general specification.

Figure 2 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. The weight effect seems slightly stronger in 1986 (when the women were young) than in 2000 (when the women were relatively old).

In drawing a CB around $\hat{m}(c, x)$, as once mentioned, the asymptotic variance estimator did not work well: the CB was too wide without trimming and too narrow with trimming. Instead, we applied nonparametric bootstrap, resampling from the original data with replacement while using the same bandwidth. One way of getting a CB is using the lower and

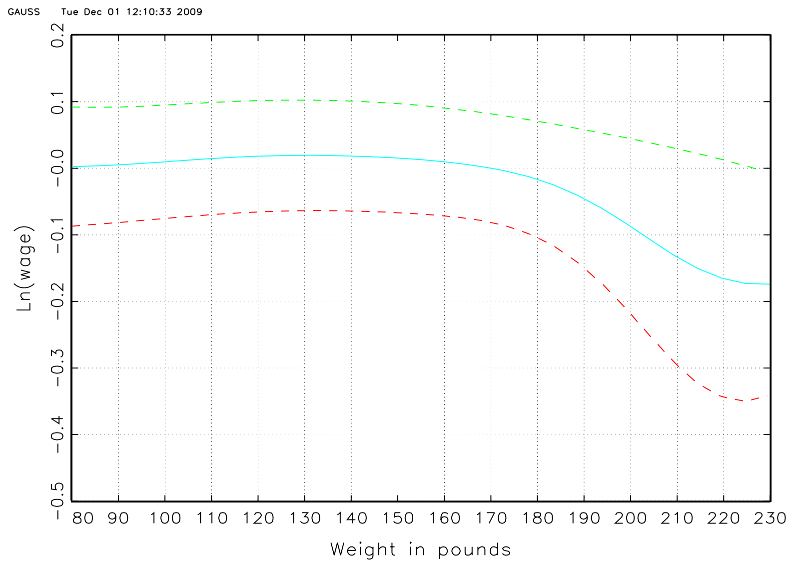

Figure 2: Ln(Wage) vs. Weight 2000 (top), 1986 (bottom) and Combined at MED Height

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 100 times to obtain the bootstrap $SD_{boot}(c, x)$, and the CB used is

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

i.e., in Figures 3-5, the ln(wage) function is shown with height fixed respectively at the LQ (63.2 inches$\simeq$160cm), MED (64.9 inches$\simeq$165cm) and UQ (66.5 inches$\simeq$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 $m(c, x)$. To see if other studies experienced the same problem or not with MIE, we looked into those listed earlier. Rodriguez-Poo et al. (2005, p.714) noted that their asymptotic variance estimator gave too narrow CB's (as in our experience with trimming); instead they used a bootstrap in Härdle et al (2004) which requires two bandwidths and gives a conservative CB. Azomahou (2006) used 'wild bootstrap' to obtain CB's which requires again two bandwidths; the bootstrap is described in their unpublished longer version of the paper. Bontemps et al. (2008) did not show CB's for their additive nonparametric model. Although Gregory and Ruhm (2009) did not use MIE, they also used a wild bootstrap with two bandwidths for their semi-linear model. The main attraction of bootstrap is in its data-dependent automatic nature of inference. When bootstrap depends on two bandwidths, it seems to lose much of its attractions, which is why we use the simple data-resampling ('nonparametric') bootstrap in our application. Kondo and Lee (2003, p.42) showed an example of "peril" in using a two-bandwidth test—two might be one too many.

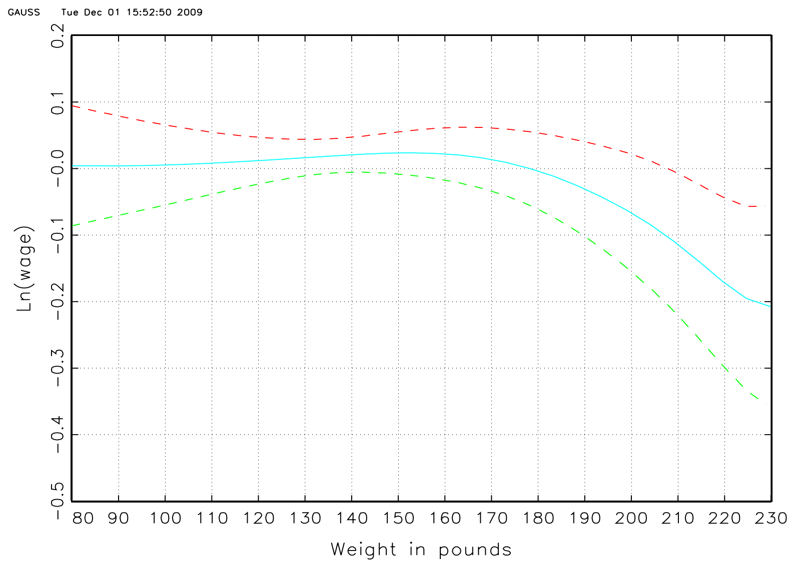

Figure 3: Ln(Wage) Function at LQ Height (about 160cm)

Looking at Figures 3-5, the ln(wage) function appears flat (or increasing slightly) up to the average weight 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 3, it is virtually impossible to fit a linear line in the CB, implying a significant nonlinear effect; in Figure 4, 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 5, 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 in becoming slimmer than the normal (average) weight. But our nonparametric estimator paints a different picture: there is a wage gain only for those with an above-average weight when they lose 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 a 9% wage gain.

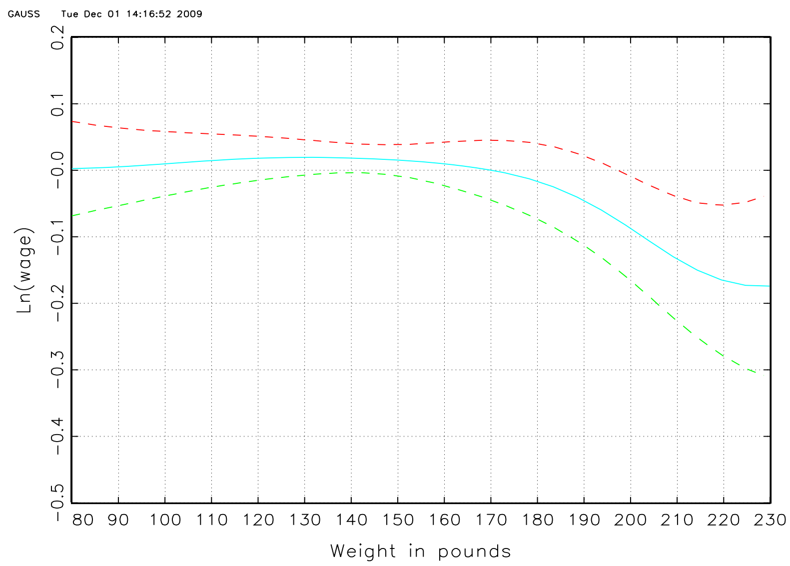

Figure 4: Ln(Wage) Function at MED Height (about 165cm)

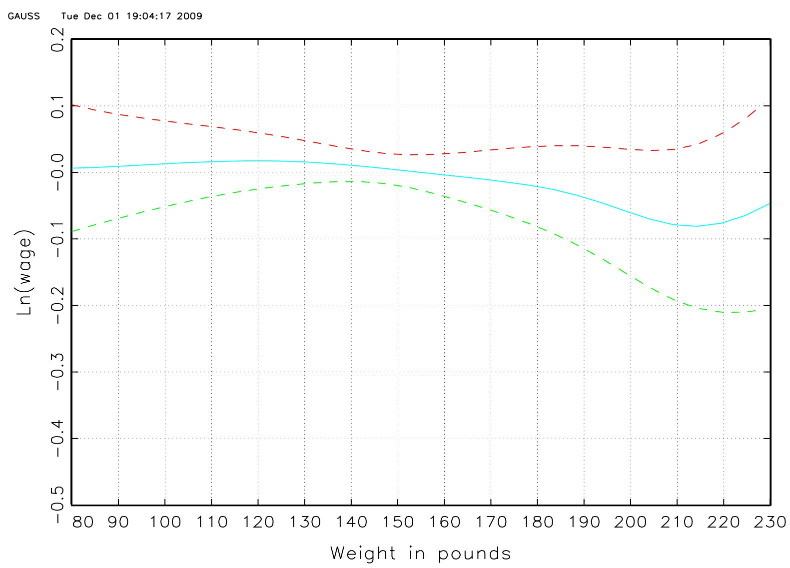

Figure 5: Ln(Wage) Function at UQ Height (about 170cm)

Figures 3-5 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 with above-average 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 (relative to the average weight) and the negative effect over the over-weight range.* This is one of the most interesting findings from the nonparametric method—something not foreseen using only the linear model.

As an obesity measure, BMI is not bad at all for white 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 procedure to the white males in the NLSY data, and indeed obtain no statistically significant findings. In a sense, this provides support to our methodology. Despite no significant findings, some interesting features do exist for white males as noted in the appendix.

3.5 Model Checks: Nonlinearity and Endogeneity

One may wonder if our nonparametric approach is really necessary. As a "compromise" between our approach and the panel linear model, we can think of a semi-linear model with a nonparametric component only for BMI as considered in Kline and Tobias (2008) and Gregory and Ruhm (2009): $Y_{it} = \tilde{\rho}(BMI_{it}) + W_{it}'\beta + \alpha_i + U_{it}$ for some nonparametric function $\tilde{\rho}$. We can difference this equation to remove α_i and obtain two MIE's and then the averaging estimator. Figure 6 shows the result where the bandwidth is $1.25N^{-1/6}SD$ and W_{it} is the same as in the above figures. Although there is some evidence of wage gain over BMI 25-28 and then wage penalty over BMI 28 to 32, the flat line at 0 can fit the CB. Since Gregory and Ruhm (2009) used wage not ln(wage), we also tried $Y_{it} = wage_{it}$ to get Figure 7, which is similar to Figure 6 in shape (going up over BMI 22-28 and then down over 28-32). The CB in Figure 7 also includes the flat zero line—no significant finding, differently from the approach with $\rho(C_i, X_{it})$. Hence it seems worth the effort going for $\rho(C_i, X_{it})$ instead of the simpler $\tilde{\rho}(BMI_{it})$.

To go beyond an informal graphical comparison, we implemented an artificial regressor-type test analogous to tests in Wooldridge (1992) as follows. Set up

$$(i) : \Delta Y_i = \tilde{\mu}(BMI_{i1}, BMI_{i2}) + W_{i1}'\beta_1 + W_{i2}'\beta_2 + \xi_{\rho 1}\widehat{\Delta\rho_i} + \xi_{\rho 2}\widehat{\Delta\rho_i}^2 + error_i$$

where $\widehat{\Delta\rho_i} \equiv \rho(\widehat{C_i, X_{i2}}) - \rho(\widehat{C_i, X_{i1}})$ and $\rho(\widehat{C_i, X_{it}})$ is the fitted value from the MIE-based averaging estimator. The adequacy of the simpler model can be tested with $H_0 : \xi_{\rho 1} = \xi_{\rho 2} = 0$, which is also reminiscent of the RESET test. When LSE was applied to this model, we obtained $0.649\widehat{\Delta\rho_i} - 0.557\widehat{\Delta\rho_i}^2$ with t-values 1.88 and -0.81 respectively. As the p-value for the t-value 1.88 is 0.06, we may say that $\widehat{\Delta\rho_i}$ is "borderline significant".

For a higher power test, we also set up

$$(ii) : \Delta Y_i = \xi_{\tilde{\rho}}\{\tilde{\rho}(\widehat{BMI_{i2}}) - \tilde{\rho}(\widehat{BMI_{i1}})\} + W_{i1}'\beta_1 + W_{i2}'\beta_2 + \xi_{\rho 1}\widehat{\Delta\rho_i} + \xi_{\rho 2}\widehat{\Delta\rho_i}^2 + error_i$$

where $\tilde{\rho}(\widehat{BMI_{it}})$ is the fitted value from the MIE-based averaging estimator. The LSE result for $\xi_{\rho 1}$ and $\xi_{\rho 2}$ is $0.691\widehat{\Delta\rho_i} - 1.853\widehat{\Delta\rho_i}^2$ with t-values 1.89 and -2.38 respectively. Hence the H_0 is rejected easily in this test, which adds a support to the above conclusion for '$\rho(C_i, X_{it})$ worthiness' based only on graphs.

The two tests trade off (dis-)advantages. The first test is theoretically sound because the fact that $\widehat{\Delta\rho_i}$ is a generated regressor in (i) does not alter the asymptotic distribution of the first test statistic under the H_0, whereas the fact that $\tilde{\rho}(\widehat{BMI_{i2}}) - \tilde{\rho}(\widehat{BMI_{it}})$ is a generated regressor in (ii) may alter the asymptotic distribution of the second test statistic even under the H_0. But the power of the first test is likely to be weaker because it does not use the information $\tilde{\mu}(BMI_{i1}, BMI_{i2}) = \tilde{\rho}(BMI_{i2}) - \tilde{\rho}(BMI_{i1})$ thus inappropriately allowing for interaction between BMI_{i1} and BMI_{i2}.

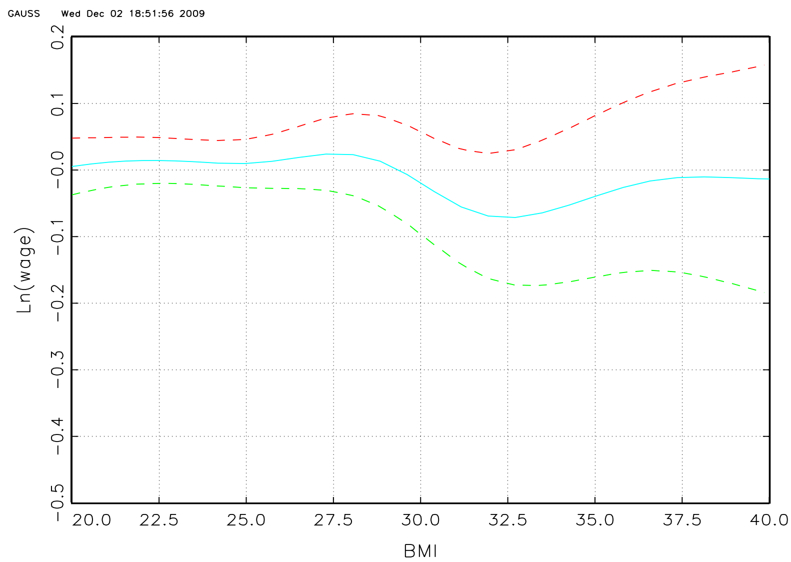

Figure 6: Ln(Wage) v. BMI Function: Semi-Linear Panel

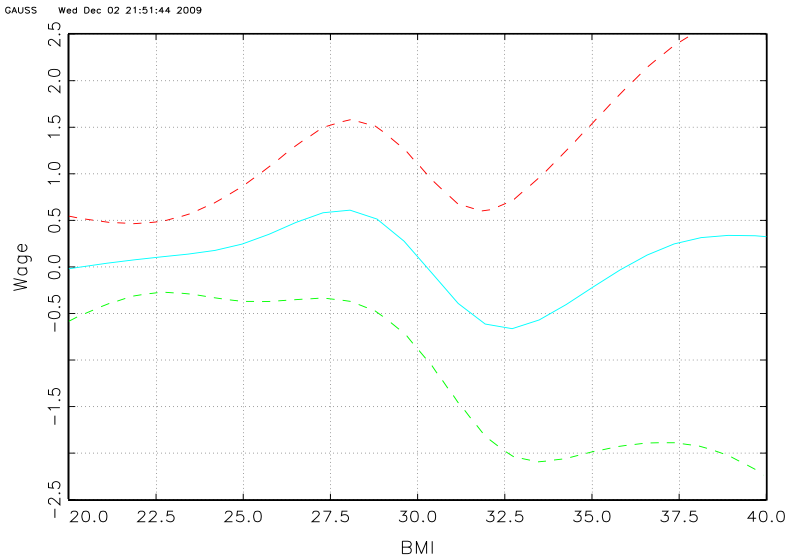

Figure 7: Wage v. BMI Function: Semi-Linear Panel

Weight X_{it} can be endogenous through its relation to U_{it}, not just through α_i. If so, removing α_i is not enough to make X_{it} exogenous. For this, we tried the following control function approach. Set up a triangular system

$$Y_{it} = \rho(C_i, X_{it}) + W_{it}'\beta + \alpha_i + U_{it}, \quad U_{it} = \zeta V_{it} + \varepsilon_{it}$$
$$X_{it} = C_i'\gamma_c + W_{it}'\gamma_w + Z_{it}'\gamma_z + \delta_i + V_{it}$$

where ζ, γ_c, γ_w and ζ_z are parameters, Z_{it} is an instrument for X_{it}, δ_i is a time-constant error in the X_{it} equation, V_{it} is a time-varying error in the X_{it} equation and ε_{it} is an error term independent of $(C_i, X_{i1}, X_{i2}, W_{i1}, W_{i2})$. In this model, the endogeneity of X_{it} with respect to U_{it} is captured by ζV_{it} in U_{it}. Difference the model to get ($\Delta X_i \equiv X_{i2} - X_{i1}$, $\Delta Z_i \equiv Z_{i2} - Z_{i1}$ and $\Delta V_i \equiv V_{i2} - V_{i1}$)

$$\Delta Y_i = \mu(C_i, X_{i1}, X_{i2}) + \Delta W_i'\beta + \zeta \Delta V_i + \Delta U_i \quad \text{and} \quad \Delta X_i = \Delta W_i'\gamma_w + \Delta Z_i'\gamma_z + \Delta V_i.$$

LSE can be applied to the ΔX_i equation to obtain the residual $\widehat{\Delta V}_i$, which can be then used as an extra regressor (i.e., 'control function') in the ΔY equation (as before, for a slightly higher generality, we use W_{i1} and W_{i2} as regressors instead of ΔW_i in the ΔY_i equation).

In our data, the only available variable for Z_{it} is a sibling BMI. But there are many women without any sibling, and even if they do, they may not be aware of their siblings'

BMI. This resulted in 59% data loss: $N = 537$ from $N = 1302$. The first-stage LSE result is $\Delta X_i = ..., +1.22\Delta Z_i$ with $R^2 = 0.048$ and the t-value for ΔZ_i is 4.12. Although the instrument does explain ΔX_i with ΔW_i controlled, the explanatory power is rather weak in view of $R^2 = 0.048$. The ensuing nonparametric MIE-based procedure returned insignificant findings with the flat zero curve falling in the CB; the figure is omitted to save space.

4 Conclusions

Does obesity matter for wage? The answer is yes at least for white females in the U.S. when BMI is used as an obesity measure; 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 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, 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 a nonparametric marginal integration approach. Differently from the linear model finding, we found no evidence of wage gain from a weight loss for average- (or under-) weight women. Also, for over-weight women relative to the average weight, the gain from a 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 medical 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 nonparametric procedure to the white males in the NLSY data for waves 1986 and 2000. To save space, we omit descriptive statistics and the linear model results to present only the three lines (1986, 2000 and the average line). As the white females did, the white males also gained weight over the 14 year span (from average 175 lb to 197). The linear model estimation returned no significant relationship between ln(wage) and BMI. The omitted figures with CB's also gave no significant findings as the flat zero lines are included in the CB's when height is fixed at LQ, MED and UQ. Note that the scales of Figure 8 are greater than those for the female figures. Although no statistically significant finding is available for white men, still the following is notable.

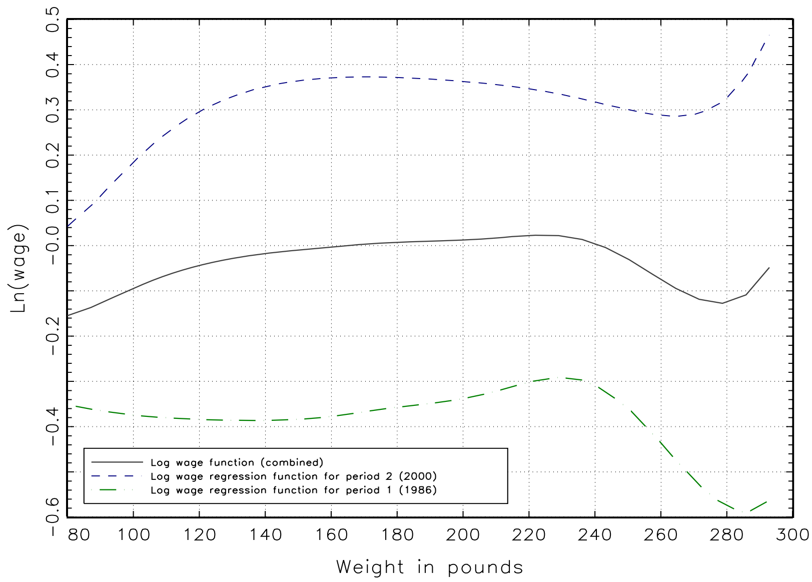

Figure 8: Ln(Wage) vs. Weight 2000 (top), 1986 (bottom) and Combined at MED Height

First, the under-weight are penalized more severely in year 2000 than in 1986, whereas the over-weight are penalized less in 2000 than in 1986. This may be due to the growing trend of valuing fitness and muscle: the under-weight is penalized more in 2000 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. Second, there seems to be a wage gain of about 12% from 50 lb loss from 280 to 230 lb. Third, the starting weight for a wage loss is about 230 lb—well above the year 2000 average weight of 197 lb; this is in sharp contrast to the females, for whom the starting weight for a wage loss is about 160 lb, which is close to the year 2000 average weight. Fourth, the wage loss from being under-weight is more visible than for females.

REFERENCES

Arellano, M. and R. Carrasco, 2003, Binary choice panel data models with predetermined variables, Journal of Econometrics 115, 125–157.

Averett, S. and S. Korenman, 1996, The economic reality of the beauty myth, Journal of Human Resources 31, 304-330.

Azomahou, T., F. Laisney and P. N. Van, 2006, Economic development and CO2 emissions: a nonparametric panel approach, Journal of Public Economics 90, 1347-1363.

Baum, C.L. and W.F. Ford, 2004, The wage effects of obesity: a longitudinal study, Health Economic 13, 885–899.

Bontemps, C., M. Simioni and Y. Surry, 2008, Semiparametric hedonic price models: assessing the effects of agricultural nonpoint source pollution, Journal of Applied Econometrics 23, 825–842.

Burkhauser, R.V. and J. 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.

Dobson, R., 2005, Number of overweight women in developing countries is increasing, British Medical Journal 330, 620.

Flegal, K.M., M.D. Carroll, C.L. Ogden and C.L. Johnson, 2002, Prevalence and trends in obesity among U.S. adults, 1999–2000, Journal of American Medical Association 288, 1723–1727.

Gortmaker, S.L., A. Must, J.M. Perrin, A.M. Sobol and W.H. Dietz, 1993, Social and economic consequence of overweight in adolescence and young adulthood, New England Journal of Medicine 329, 1008–1012.

Gregory, C.A. and C.J. Ruhm, 2009, Where does the wage penalty bite?, NBER Working Paper 14984, National Bureau of Economic Research, Cambridge, MA.

Hamermesh, D.S. and J.E. Biddle, 1994, Beauty and the labor market, American Economic Review 84, 1174–1194.

Harper, B., 2000, Beauty, stature and the labor market: a British cohort study, Oxford Bulletin of Economics and Statistics 62, 771-800.

Härdle W., S. Huet, E. Mammen and S. Sperlich. 2004, Bootstrap inference in semiparametric generalized additive models, Econometric Theory 20, 265–300.

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.

Johansson, E, P. Böckerman, U. Kiiskinen and M. Heliövaara, 2009, Obesity and labour market success in Finland: the difference between having a high BMI and being fat, Economics and Human Biology 7, 36–45.

Kline, B. and J.L. Tobias, 2008, The wages of BMI: Bayesian analysis of a skewed treatment-response model with nonparametric endogeneity, Journal of Applied Econometrics 23, 767–793.

Kondo, Y. and M.J. Lee, 2002, Nonparametric derivative estimation for related-effect panel data, presented at the 10th International Conference on Panel Data at Berlin.

Kondo, Y. and M.J. Lee, 2003, Hedonic price index estimation under mean independence of time dummies from quality-characteristics, Econometrics Journal 6, 28-45.

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.

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.

Loh, E.S., 1993, The economic effects of physical appearance, Social Science Quarterly 74, 420–438.

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

Morris, M., 2006, Body mass index and occupational attainment, Journal of Health Economics 25, 347–364.

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.

Pagan, J.A. and A. Davila, 1997, Obesity, occupational attainment, and earnings: consequences of obesity, Social Science Quarterly 8, 756–770.

Register, C.A. and D.R. Williams, 1990, Wage effects of obesity among young workers, Social Science Quarterly 71, 130–141.

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

Rodriguez-Poo, J.M., S. Sperlich and I. Fernandez, 2005, Semiparametric three-step estimation methods for simultaneous equation systems, Journal of Applied Econometrics 20, 699–721.

Sargent, J.D. and D.G. Blanchflower, 1994, Obesity and stature in adolescent and earning in young adulthood. Analysis of a British birth cohort, Archives of Pediatrics and Adolescent Medicine 148, 681–687.

Seidell, J. and K.M. Flegal, 1997, Assessing obesity: classification and epidemiology, British Medical Bulletin 53, 238-252.

Wooldridge, J. M., 1992, A test for functional form against nonparametric alternatives, Econometric Theory 8, 452-475.

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