Nonparametric Estimation and Testing of Fixed Effects Panel Data Models

Daniel J. Henderson
Department of Economics
State University of New York at Binghamton
Binghamton, NY 13902-6000
djhender@binghamton.edu

Raymond J. Carroll
Department of Statistics, Texas A&M University
College Station, TX 77843-3134,
carroll@stat.tamu.edu

Qi Li
Department of Economics, Texas A&M University
College Station, TX 77843-4228,
qi@econmail.tamu.edu

Abstract

In this paper we consider the problem of estimating nonparametric panel data models with so-called fixed effects, i.e., random effects that are correlated with the predictors in an unspecified manner. We derive the rate of convergence and asymptotic distribution of an iterative nonparametric kernel estimator. We further propose a test statistic for testing the null hypothesis of random effects against fixed effects in a nonparametric panel data regression model. We also extend the estimation method to the case of a semiparametric partially linear fixed effects model. Simulations are used to support the asymptotic development. We apply the methods to estimate the relationship between caloric intake and income using data obtained from the China Health and Nutrition Survey. With this data, a standard analysis ignoring the possibility of fixed effects suggests the implausible finding that at low levels of income, increases in income are not related to increases in caloric intake: the fixed effect approach gives estimates that are in accord with economic theory.

Some key words: Panel data; Nonparametric Kernel method; Marginal models; Partially linear model; Profile method; Semiparametric efficient bound.

Short title: Fixed Effects Panel Data Models

1 Introduction

Nonparametric and semiparametric kernel methods are becoming increasingly popular tools for statisticians/econometricians. Researchers have begun to gravitate toward nonparametric and semiparametric methods when there is little prior knowledge on specific (regression) functional forms or some known parametric specifications are deemed inadequate for the problem at hand. This often occurs when formal rejection of a parametric model yields no clues as to the direction in which to search for an improved parametric model. This growing popularity of nonparametric methods stems from their ability to relax functional form assumptions of an unknown model and let the data determine a function tailored to the data. This capacity to potentially reveal structure in the data that may be missed by common parametric specifications has encouraged growth in a variety of areas of statistics and econometrics.

The estimation of panel data models is no exception. Here the focus has been on both semiparametric (e.g. see Ke and Wang, 2001; Li and Stengos, 1996; Ullah and Roy, 1998) and nonparametric (e.g. see Henderson and Ullah, 2005; Lin and Carroll, 2000, 2001, 2006; Lin, Wang, Welsh and Carroll, 2004; Lin and Ying, 2001; Ruckstuhl, Welsh and Carroll, 1999; Wu and Zhang, 2002; Wang, 2003) estimation of random effects models. Estimation of these types of models are appropriate when the individual effect is independent of the regressors. This is common in many applications, where researchers often treat any unobserved individual heterogeneity as being distributed independently of the regressors. However, random effects estimators are inconsistent if the true model is one with fixed effects, i.e., individual effects which are correlated with the regressors. Indeed, Economists often view the assumptions for the random effects model as being unsupported by the data. In light of this we seek to develop both nonparametric and semiparametric fixed effects estimation procedures. These procedures will be consistent under either the random or fixed effects assumptions.

We present both nonparametric and semiparametric models which either take or do not take the correlation structure into account when estimating a fixed effects nonparametric/semiparametric panel data model. Our results show that for the nonparametric model, incorporating or ignoring the within-subject correlation leads to consistent estimation results. However, the incorporation of the correlation leads to an improvement in the estimated variance when the number of time periods is greater than two. For the semiparametric partially linear model, we also find that taking into account the correlation structure leads to efficient estimation of the finite dimensional (parametric) parameter.

The question of whether to use random or fixed effects in practice naturally arises with panel data. We know that when the individual effect is correlated with any of the regressors, random effects estimator becomes biased and inconsistent. The fixed effect estimator wipes out these individual effects and leads to consistent estimates. On the other hand, if the individual effects are independent of the regressors both estimators are consistent, then the random effects estimator is more efficient. This trade-off is common in econometrics and is often solved using a testing procedure. Here we develop a Hausman style test for the presence of fixed versus random effects. We also suggest a bootstrap procedure for the implementation of the test in practice.

Finally, we apply the model to study the relationship between caloric intake and income. We find that the random effects model produces an unreasonable relationship while the fixed effects model gives a more plausible result. In this example, our test for the presence of fixed effects is statistically significant, indicating the need to use a fixed-effects model. Thus, our empirical example shows the importance of the fixed effects specification in nonparametric panel data estimation.

The remainder of the paper is organized as follows: Section 2 gives the nonparametric estimation procedures when we both account for and ignore the correlation structure. In Section 3 we propose a test statistic for testing random effects against fixed effects in nonparametric panel data regression models. Section 4 generalizes the results to the case of a semiparametric partially linear model. Section 5 examines finite sample properties with a small Monte Carlo study. Section 6 reports an empirical study result using the caloric intake and income data obtained from the China Health and Nutrition Survey. Finally, Section 7 gives concluding remarks.

2 Fixed Effects Nonparametric Panel Data Models

Consider the following nonparametric panel data regression model with fixed effects

$$Y_{it} = \theta(Z_{it}) + \mu_i + \nu_{it}, \qquad (i = 1, ..., n;\ \ t = 1, ..., m) \tag{1}$$

where the functional form of $\theta(\cdot)$ is not specified. The covariate $Z_{it} = (Z_{it,1}, ..., Z_{it,q})$ is of dimension q, and all other variables are scalars. The random errors ν_{it} are assumed to be i.i.d. with a zero mean, finite variance and independent of Z_{it} for all i and t. Further, μ_i has a zero mean and finite variance. We allow μ_i to be correlated with Z_{it} with an unknown correlation structure. Hence, (1) is a fixed effects model. Alternatively, when μ_i is uncorrelated with Z_{it}, model (1) is a random effects model.

We consider the usual case of n large and m fixed, and assume that the data is independent across the i index. We take a first difference to remove the fixed effects

$$\widetilde{Y}_{it} \equiv Y_{it} - Y_{i1} = \theta(Z_{it}) - \theta(Z_{i1}) + \nu_{it} - \nu_{i1}. \tag{2}$$

Note that the above difference is to subtract observation $t = 1$ from t. From (1) we know that $E\left\{\theta(Z_{it})\right\} = E(Y_{it})$. Under this condition, $\theta(\cdot)$ defined in (2) is identified. We discuss two nonparametric estimators for $\theta(\cdot)$, one which utilizes the variance structure Σ of the terms $\widetilde{Y}_{it}$, and the other which ignores Σ. We will show that the former estimator is asymptotically more efficient when $m > 2$.

2.1 Motivation for Estimating $\theta(\cdot)$

A somewhat simplistic explanation for consideration of fixed effects models and the need for estimation of the function $\theta(\cdot)$ arises from considerations such as the following. Suppose that Y_{it} is the income of individual i at time period t, and Z_{it} is education of individual i at time period t , e.g., number of years of schooling. The fixed effects term μ_i in (1) includes the individual's unobservable characteristics such as ability, e.g., IQ level, characteristics which are not observable for the data at hand. Basically, this is an omitted variable bias problem.

In this problem, economists are interested in the marginal effects of education on income, after controlling for the unobservable individual ability factors. Hence, they are interested in the marginal effects in the average income change for an additional year of education regardless of whether the person has high or low ability. In this simple example, it is reasonable to believe that ability and education are positively correlated. If one does not control for the unobserved individual effects, then one would over estimate the true marginal effects of education on income (i.e., with an upwards bias).

2.2 An Estimator Using the Variance Structure

Our goal in this section is to derive an estimator which exploits the variance structure. We start by defining $\widetilde{\nu}_{it} = \nu_{it} - \nu_{i1}$ and $\widetilde{\nu}_i = (\widetilde{\nu}_{i2}, ..., \widetilde{\nu}_{im})^{\mathrm{T}}$, where the superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose of a matrix $(\cdot)$. The variance-covariance matrix of $\widetilde{\nu}_i$, defined as $\Sigma = \mathrm{cov}(\widetilde{\nu}_i | Z_{i1}, ..., Z_{im}) = \mathrm{cov}(\widetilde{\nu}_i)$, is given by

$$\Sigma = \sigma_\nu^2 (I_{m-1} + e_{m-1} e_{m-1}^{\mathrm{T}}),$$

where I_{m-1} is an identity matrix of dimension $(m-1) \times (m-1)$, and e_{m-1} is a $(m-1) \times 1$ vector of ones. It is easy to check that $\Sigma^{-1} = \sigma_\nu^{-2}(I_{m-1} - e_{m-1} e_{m-1}^{\mathrm{T}}/m)$. Following Wang (2003) and Lin and Carroll (2006) we use a profile likelihood approach to estimate $\theta(\cdot)$. The criterion function for individual i is given by

$$\mathcal{L}_i(\cdot) = \mathcal{L}(Y_i, \theta_i) = -\frac{1}{2}(\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1})^{\mathrm{T}} \Sigma^{-1} (\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}), \qquad (3)$$

where $\widetilde{Y}_i = (\widetilde{Y}_{i2}, ..., \widetilde{Y}_{im})^{\mathrm{T}}$, $\theta_{it} = \theta(Z_{it})$ and $\theta_i = (\theta_{i2}, ..., \theta_{im})^{\mathrm{T}}$.

Defining $\mathcal{L}_{i,t\theta} = \partial \mathcal{L}_i(\cdot)/\partial \theta_{it}$, and $\mathcal{L}_{i,ts\theta} = \partial^2 \mathcal{L}_i(\cdot)/(\partial \theta_{it} \partial \theta_{is})$, from (3) we obtain

$$\begin{aligned} \mathcal{L}_{i,1\theta} &= -e_{m-1}^{\mathrm{T}} \Sigma^{-1} (\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}); \\ \mathcal{L}_{i,t\theta} &= c_{t-1}^{\mathrm{T}} \Sigma^{-1} (\widetilde{Y}_i - \theta_i + \theta_{i1} e_{m-1}) \qquad \text{for } t \geq 2, \end{aligned}$$

where c_{t-1} is a vector of dimension $(m-1) \times 1$ with the $(t-1)$ element being 1 and all other elements being 0. Because Σ^{-1} is known up to a constant of proportionality, it can be considered known for purposes of estimating $\theta(\cdot)$, although not for inference.

Here we will maximize a kernel-weighted objective function. We start by defining the product kernel $K_h(v) = \prod_{j=1}^q h_j^{-1} k(v_j/h_j)$, where $k(\cdot)$ is a univariate kernel function. Next, let $f_t(\cdot)$ denote the density function of Z_{it}. Further, define $(Z_{it} - z)/h = \{(Z_{it,1} - z_1)/h_1, ..., (Z_{it,q} - z_q)/h_q\}^{\mathrm{T}}$ and $G_{it}(z, h) = [1, \{(Z_{it} - z)/h\}^{\mathrm{T}}]^{\mathrm{T}}$, where G_{it} is of dimension $(q+1) \times 1$. Finally, define $\theta^{(1)}(z) = \partial \theta(z)/\partial z$ as the first order derivative of θ with respect to z. We estimate the unknown function $\theta(z)$ by solving the first order condition

$$0 = \sum_{i=1}^n \sum_{t=1}^m K_h(Z_{it}, z) G_{it}(z, h) \mathcal{L}_{i,t\theta}[Y_i, \widehat{\theta}(Z_{i1}), ..., \widehat{\theta}(z) + \{(Z_{it} - z)/h\} \widehat{\theta}^{(1)}(z), ..., \widehat{\theta}(Z_{im})], \qquad (4)$$

where the argument $\mathcal{L}_{i,t\theta}$ is $\widehat{\theta}(Z_{is})$ for $s \neq t$ and $\widehat{\theta}(z) + \{(Z_{it} - z)/h\}\widehat{\theta}^{(1)}(z)$ when $s = t$.

2.2.1 An Iterative Procedure for Nonparametric Estimation

Equation (4) suggests the following iterative procedure. Suppose the current estimate of $\theta(z)$ at the $[\ell - 1]^{th}$ step is $\widehat{\theta}_{[\ell-1]}(z)$. Then the next step estimate of $\theta(z)$ is $\widehat{\theta}_{[\ell]}(z) = \widehat{\alpha}_0(z)$, where $(\widehat{\alpha}_0, \widehat{\alpha}_1)$ solve the following equation:

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z) G_{it}(z, h) \qquad (5) \\ &\qquad \times \mathcal{L}_{i,t\theta}[Y_i, \widehat{\theta}_{[\ell-1]}(Z_{i1}), ..., \widehat{\alpha}_0 + \{(Z_{it} - z)/h\}^{\mathrm{T}}\widehat{\alpha}_1, ..., \widehat{\theta}_{[\ell-1]}(Z_{im})]. \end{aligned}$$

Below we give an algorithm for estimating $\theta(\cdot)$. We note here that we need to use the restriction that $\sum_{i=1}^{n}\sum_{t=1}^{m}\{Y_{it} - \widehat{\theta}(Z_{it})\} = 0$ in order for $\theta(\cdot)$ to be uniquely defined based on (2), since $E(Y_{it}) = E\{\theta(Z_{it})\}$. The algorithm is linear in the Y_{it}'s. By defining

$$H_{i,[\ell-1]} = \begin{pmatrix} Y_{i2} - \widehat{\theta}_{[\ell-1]}(Z_{i2}) \\ \cdots \\ Y_{im} - \widehat{\theta}_{[\ell-1]}(Z_{im}) \end{pmatrix} - \{Y_{i1} - \widehat{\theta}_{[\ell-1]}(Z_{i1})\}e_{m-1},$$

we get

$$\begin{aligned} 0 &= \sum_{i=1}^{n} K_h(Z_{i1}, z) G_{i1} \left[-e_{m-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]} + e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1}\left\{\widehat{\theta}_{[\ell-1]}(Z_{i1}) - G_{i1}^{\mathrm{T}}(\alpha_0, \alpha_1)^{\mathrm{T}}\right\}\right] \\ &+ \sum_{i=1}^{n}\sum_{t=2}^{m} K_h(Z_{it}, z) G_{it} \left[c_{t-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]} + c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1}\left\{\widehat{\theta}_{[\ell-1]}(Z_{it}) - G_{it}^{\mathrm{T}}(\alpha_0, \alpha_1)^{\mathrm{T}}\right\}\right], \end{aligned}$$

where $G_{it} = G_{it}(z, h)$. Further, by defining

$$\begin{aligned} D_1 &= n^{-1}\sum_{i=1}^{n}\left\{e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1}K_h(Z_{i1}, z)G_{i1}G_{i1}^{\mathrm{T}} + \sum_{t=2}^{m} c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1}K_h(Z_{it}, z)G_{it}G_{it}^{\mathrm{T}}\right\}; \\ D_2 &= n^{-1}\sum_{i=1}^{n}\left\{e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1}K_h(Z_{i1}, z)G_{i1}\widehat{\theta}_{[\ell-1]}(Z_{i1})\right. \end{aligned}$$

$$+\sum_{t=2}^{m} c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1}K_h(Z_{it},z)G_{it}\widehat{\theta}_{[\ell-1]}(Z_{it})\Bigg\};$$

$$D_3 = n^{-1}\sum_{i=1}^{n}\left\{\sum_{t=2}^{m}K_h(Z_{it},z)G_{it}c_{t-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]} - K_h(Z_{i1},z)G_{i1}e_{m-1}^{\mathrm{T}}\Sigma^{-1}H_{i,[\ell-1]}\right\},$$

and then solving for α_0 and α_1 leads to $\{\widehat{\alpha}_0(z),\widehat{\alpha}_1(z)\}^{\mathrm{T}} = D_1^{-1}(D_2+D_3)$. The next step estimate of $\theta(z)$ is given by $\widehat{\theta}_{[\ell]}(z) = \widehat{\alpha}_0(z)$, while $\widehat{\alpha}_1(z)$ gives the next step derivative estimator of $\theta(z)$.

Wang (2003) considered the random effects case. In her model a consistent initial estimator can be obtained by replacing Σ by an identity matrix. The simulations reported in Wang (2003) show that a one-step iteration is nearly as efficient as the result for full convergence, and that it usually only takes 3 to 4 iterations to achieve full convergence. In our case, even when one replaces Σ by an identity matrix, (2) is an additive model with the restriction that the two additive functions have the same functional form, and an initial consistent estimator of $\theta(\cdot)$ can be obtained by the standard backfitting method, see for example Opsomer and Ruppert (1997).

2.2.2 Asymptotic Theory

To derive the asymptotic distribution of $\widehat{\theta}(z)$, we first give some regularity conditions and definitions.

Assumption 1: The random variables (Y_{it}, Z_{it}) are independent and identically distributed (iid) across the i index and Y_{it} has finite fourth moments for all t. Let $f_t(\cdot)$ denote the density function of Z_{it}; then both $f_t(\cdot)$ and $\theta(\cdot)$ are twice continuously differentiable functions. Let $\mathcal{S}_t$ denote the support of Z_{it}; then $f_t(z)$ is bounded from both below and above by some positive constant for all $z \in \mathcal{S}_t$.

Assumption 2: $K(v) = \prod_{s=1}^{q} k(v_s)$ is a product kernel, the univariate kernel function $k(\cdot)$ is a bounded, symmetric probability density function with $\kappa_2 \overset{def}{=} \int k(v)v^2$ is finite. As $n \to \infty$, $h_r \to 0$ for all $r = 1, ..., q$ and $nh_1 \cdots h_q \to \infty$.

We first make the following general definitions, the calculations of which will follow after the statement of the main result. Define $\Omega(z) = -\sum_{t=1}^{m} f_t(z) E\{\mathcal{L}_{i,tt\theta} | Z_{it} = z\}$. Further define $\epsilon_{it} = \mathcal{L}_{t\theta}\{\widetilde{Y}_i, Z_i, \theta(Z_{i1}), ..., \theta(Z_{it}), ..., \theta(Z_{im})\}$, and $\eta_n = \sum_{s=1}^{q} h_s^2 + (nh_1 \cdots h_q)^{-1/2}$. Also, define $b_r(z)$ be a bounded and continuous function that is the solution to

$$b_r(z) = \frac{\kappa_2}{2}\theta_{rr}(z) - \sum_{t=1}^{m}\sum_{s\neq t}^{m} f_t(z) E\left\{\mathcal{L}_{i,ts\theta}(\cdot) b_r(Z_{is}) | Z_{it} = z\right\} / \Omega(z),$$

where $\kappa_2 = \int v^2 k(v) dv$ and $\theta_{rr}(z) = \partial^2\theta(z)/\partial z_r^2$. In general, $b_r(\cdot)$ does not have a closed form expression.

Our main result in this section, one that is sketched in the Appendix, is the following.

Main Result: The estimator $\widehat{\theta}(z)$ has the asymptotic expansion

$$\widehat{\theta}(z) - \theta(z) = (\kappa_2/2)\sum_{r=1}^{q} h_r^2 b_r(z) - n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)\epsilon_{is}/\Omega(z) + o_p(\eta_n).$$

Thus, the asymptotic bias and variance of $\widehat{\theta}(z)$ are

$$\begin{aligned} \text{bias} &= (\kappa_2/2)\sum_{r=1}^{q} h_r^2 b_j(z) + o(\sum_{r=1}^{q} h_r^2); \\ \text{variance} &= \frac{\kappa^q}{nh_1 \cdots h_q}\frac{1}{\Omega(z)} + o\{(nh_1 \cdots h_q)^{-1}\}. \end{aligned}$$

Remark 1: The particular form of our problem means that many of the terms have simple expressions. In particular,

$$\begin{aligned} \mathcal{L}_{i,11\theta} &= -e_{m-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1} = -(m-1)/(m\sigma_\nu^2); \\ \mathcal{L}_{i,tt\theta} &= -c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{t-1} = -(m-1)/(m\sigma_\nu^2) \text{ for } t \geq 2; \\ \mathcal{L}_{i,1t\theta} &= -c_{t-1}^{\mathrm{T}}\Sigma^{-1}e_{m-1} = -\frac{1}{m\sigma_\nu^2} \qquad \text{for } t \geq 2; \end{aligned}$$

$$
\begin{aligned}
\mathcal{L}_{i,ts\theta} &= -c_{t-1}^{\mathrm{T}}\Sigma^{-1}c_{s-1} = \frac{1}{m\sigma_\nu^2} \qquad \text{for } t, s \geq 2 \text{ and } t \neq s; \\
\Omega(z) &= -\sum_{t=1}^{m} f_t(z)E\{\mathcal{L}_{tt\theta}(\cdot)|Z_t = z\} = \frac{m-1}{m\sigma_\nu^2}\sum_{t=1}^{m} f_t(z).
\end{aligned}
$$

Remark 2: If we further assume that $f_t(z) = f(z)$ for all t, then the asymptotic variance becomes

$$
\text{avar}\{\sqrt{nh_1\cdots h_q}\,\widehat{\theta}(z)\} = \frac{\sigma_\nu^2\kappa^q}{(m-1)f(z)}. \tag{6}
$$

Under the assumption that $h_r \sim n^{-1/(4+q)}$, and by defining $\kappa = \int k^2(v)dv$, we obtain the following asymptotic distribution for $\widehat{\theta}(z)$:

$$
(nh_1\cdots h_q)^{1/2}\left\{\widehat{\theta}(z) - \theta_0(z) - \sum_{r=1}^{q} h_r^2 b_r(z)\right\} \Rightarrow \text{Normal}\{0, \kappa^q/\Omega(z)\}.
$$

Remark 3: Obviously, $\Omega(z)$ can be consistently estimated by $\widehat{\Omega}(z) = (m-1)\sum_{t=1}^{m}\widehat{f}_t(z)/(m\widehat{\sigma}_\nu^2)$, where $\widehat{f}_t(z) = n^{-1}\sum_{i=1}^{n}K_h(Z_{it}, z)$, $\widehat{\sigma}_\nu^2 = \frac{1}{2n(m-1)}\sum_{i=1}^{n}\sum_{t=2}^{m}\widehat{u}_{it}^2$, and $\widehat{u}_{it} = Y_{it} - Y_{i1} - \{\widehat{\theta}(Z_{it}) - \widehat{\theta}(Z_{i1})\}$. If Z_{it} is strictly stationary in t, $\Omega(z) = (m-1)f(z)/\sigma_\nu^2$, and one can estimate $\Omega(z)$ by $(m-1)\widehat{f}(z)/\widehat{\sigma}_\nu^2$, where $\widehat{f}(z) = (nm)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}K_h(Z_{it}, z)$.

2.3 An Estimator Ignoring the Correlation Structure

In this section we derive the asymptotic distribution of a fixed effect estimator that ignores the variance structure Σ. In this case the objective function (3) is modified by replacing Σ^{-1} by I_{m-1}, and thus becomes

$$
\mathcal{L}_i(\cdot) = \mathcal{L}(Y_i, \theta_i) = -\frac{1}{2}\left(\widetilde{Y}_i - \theta_i + \theta_{i1}e_{m-1}\right)^{\mathrm{T}}\left(\widetilde{Y}_i - \theta_i + \theta_{i1}e_{m-1}\right).
$$

Then from (3) we obtain $\mathcal{L}_{i,t\theta}$ and $\mathcal{L}_{i,ts\theta}$ as

$$
\begin{aligned}
\mathcal{L}_{i,1\theta} &= -e_{m-1}^{\mathrm{T}}\left(\widetilde{Y}_i - \theta_i + \theta_{i1}e_{m-1}\right); \\
\mathcal{L}_{i,t\theta} &= c_{t-1}^{\mathrm{T}}\left(\widetilde{Y}_i - \theta_i + \theta_{i1}e_{m-1}\right), \qquad \text{for } t \geq 2.
\end{aligned}
$$

The iterative procedure is similar to before. Equation (5) remains of the same form, and we solve

$$\begin{aligned} 0 &= \sum_{i=1}^{n} K_h(Z_{i1}, z) G_{i1} \left[-e_{m-1}^{\mathrm{T}} H_{i,[\ell-1]} + (m-1) \left\{ \widehat{\theta}_{[\ell-1]}(Z_{i1}) - G_{i1}^{\mathrm{T}} (\alpha_0, \alpha_1)^{\mathrm{T}} \right\} \right] \\ &\quad + \sum_{i=1}^{n} \sum_{t=2}^{m} K_h(Z_{it}, z) G_{it} \left[c_{t-1}^{\mathrm{T}} H_{i,[\ell-1]} + \left\{ \widehat{\theta}_{[\ell-1]}(Z_{it}) - G_{it}^{\mathrm{T}} (\alpha_0, \alpha_1)^{\mathrm{T}} \right\} \right]. \end{aligned}$$

By replacing Σ^{-1} with I_{m-1}, and noting that $e_{m-1}^{\mathrm{T}} e_{m-1} = m-1$ and $c_{t-1}^{\mathrm{T}} c_{t-1} = 1$, analogous to the definitions of D_1, D_2 and D_3 we obtain

$$\begin{aligned} J_1 &= n^{-1} \sum_{i=1}^{n} \left\{ (m-1) K_h(Z_{i1}, z) G_{i1} G_{i1}^{\mathrm{T}} + \sum_{t=2}^{m} K_h(Z_{it}, z) G_{it} G_{it}^{\mathrm{T}} \right\}; \\ J_2 &= n^{-1} \sum_{i=1}^{n} \left\{ (m-1) K_h(Z_{i1}, z) G_{i1} \widehat{\theta}_{[\ell-1]}(Z_{i1}) + \sum_{t=2}^{m} K_h(Z_{it}, z) G_{it} \widehat{\theta}_{[\ell-1]}(Z_{it}) \right\}; \\ J_3 &= n^{-1} \sum_{i=1}^{n} \left\{ \sum_{t=2}^{m} K_h(Z_{it}, z) G_{it} c_{t-1}^{\mathrm{T}} H_{i,[\ell-1]} - K_h(Z_{i1}, z) G_{i1} e_{m-1}^{\mathrm{T}} H_{i,[\ell-1]} \right\}. \end{aligned}$$

Then solving for α_0 and α_1 leads to $\{\widetilde{\alpha}_0(z), \widetilde{\alpha}_1(z)\} = J_1^{-1}(J_2 + J_3)$. One can use the results of Section 2.2 to derive the asymptotic distribution of $\widetilde{\alpha}_0(z)$ by replacing Σ^{-1} by I_{m-1}. However, direct calculation of the asymptotic variance is quite simple. Under the assumption that $f_t(z) = f(z)$ for all $t = 1, ..., m$, it is easy to see that $J_1 = 2(m-1) f(z) \mathrm{diag}(1, \kappa_2 I_{m-1}) + o_p(1)$, where $\kappa_2 = \int v^2 k(v) dv$. It can be shown that the asymptotic variance of $J_2 + J_3$ comes from J_3 by replacing $H_{i,[\ell-1]}$ with ϵ_i, where $\epsilon_i = (\nu_{i2} - \nu_{i1}, ..., \nu_{im} - \nu_{i1})^{\mathrm{T}}$.

We decompose J_3 into $J_3 = J_{3,1} - J_{3,2}$, $J_{3,1} = n^{-1} \sum_{i=1}^{n} \sum_{t=2}^{m} K_h(Z_{it}, z) G_{it} c_{t-1}^{\mathrm{T}} H_{i,[\ell-1]}$ and $J_{3,2} = n^{-1} \sum_{i=1}^{n} K_h(Z_{i1}, z) G_{i1} e_{m-1}^{\mathrm{T}} H_{i,[\ell-1]}$. It is also easy to show that

$$\mathrm{avar}(J_{3,1}) = \frac{1}{n h_1 \cdots h_q} \{2(m-1) \sigma_\nu^2 \kappa^{q-1} f(z)\} \mathrm{diag}(\kappa, \kappa_{22} I_{m-1}),$$

where $\kappa_{22} = \int v^2 k(v)^2 dv$ and $\kappa = \int k(v)^2 dv$, and

$$\mathrm{avar}(J_{3,2}) = \frac{1}{n h_1 \cdots h_q} \{(m-1) m \sigma_\nu^2 \kappa^{q-1} f(z)\} \mathrm{diag}(\kappa, \kappa_{22} I_{m-1}),$$

and that $\mathrm{cov}(J_{3,1}, J_{3,2})$ has an order smaller than $O\{(nh_1 \cdots h_q)^{-1}\}$. Hence, we have that

$$\mathrm{avar}(J_3) = \frac{1}{nh_1 \cdots h_q}\{4(m-1)\sigma_\nu^2 \kappa^{q-1} f(z)\}\{\frac{(2+m)}{4}\}\mathrm{diag}(\kappa, \kappa_{22} I_{m-1}).$$

Thus, we immediately obtain the asymptotic variance of $\{\widetilde{\alpha}_0(z), \widetilde{\alpha}_1(z)\}^{\mathrm{T}}$ which is given by

$$\begin{aligned} & \frac{1}{4(m-1)^2 f(z)^2}\frac{1}{nh_1 \cdots h_q}\left\{4(m-1)\sigma_\nu^2 \kappa^{q-1} f(z)\right\}\left\{\frac{(2+m)}{4}\right\}\mathrm{diag}\{\kappa, (\kappa_{22}/\kappa_2^2) I_{m-1}\} \\ = & \frac{1}{nh_1 \cdots h_q}\frac{\kappa^{q-1}\sigma_\nu^2}{(m-1)f(z)}\frac{(2+m)}{4}\mathrm{diag}\{\kappa, (\kappa_{22}/\kappa_2^2) I_{m-1}\}. \end{aligned} \tag{7}$$

Comparing (7) with (6), we see that the relative asymptotic variance of $\widetilde{\alpha}_0(z)$ and $\widehat{\alpha}_0(z)$ is

$$\frac{\mathrm{avar}(\widetilde{\alpha}_0(z))}{\mathrm{avar}(\widehat{\alpha}_0(z))} = \frac{2+m}{4},$$

which equals one if $m = 2$ (as expected) and is greater than one when $m > 2$.

Here we note that when we replace Σ^{-1} with I_{m-1}, $\mathcal{L}_{i,11\theta} = -(m-1)$ and $\mathcal{L}_{i,tt\theta} = -1$ for $t \geq 2$, and $\mathcal{L}_{i,ts\theta} = -1$ for $t, s = 1, ..., m-1$ when t is different from s. Thus, $\Omega(z) = 2(m-1)f(z)$ and the leading bias term becomes

$$\begin{aligned} b_r(z) &= \frac{\kappa_2}{2}\theta_{rr}(z) - \sum_{t=1}^{m}\sum_{s \neq t}^{m} f_t(z) E\{\mathcal{L}_{i,ts\theta}(\cdot) b_r(Z_{is}) | Z_{it} = z\}/\Omega(z) \\ &= \frac{\kappa_2}{2}\theta_{rr}(z) + \frac{1}{2(m-1)}\sum_{t=1}^{m}\sum_{s=1, s\neq t}^{q} E\{b_r(Z_{is}) | Z_{it} = z\}. \end{aligned}$$

3 A Nonparametric Hausman Test

In this section we discuss how to test for the presence of random effects versus fixed effects in a nonparametric panel data model. The model remains as (1). Define $u_{it} = \mu_i + \nu_{it}$. The random effects specification assumes that μ_i is uncorrelated with the regressor Z_{it}, while for the fixed effects case, μ_i is allowed to be correlated with Z_{it} in an unknown way.

We are interested in testing the null hypothesis (H_0) that μ_i is a random effect versus the alternative hypothesis (H_1) that μ_i is a fixed effect. The null hypothesis can be written as

$$H_0 : E(\mu_i | Z_{i1}, ..., Z_{im}) \equiv 0.$$

The alternative hypothesis is the negation the null, i.e., H_1: $E(\mu_i | Z_{i1}, ..., Z_{im}) \neq 0$ on a set with positive measure. We maintain the assumption that $E(\nu_{it} | Z_{i1}, ..., Z_{im}) = 0$ under either H_0 or H_1. The null and the alternative hypotheses can then be equivalently written as

$$H_0 : E(u_{it} | Z_{i1}, ..., Z_{im}) \equiv 0,$$

where $u_{it} = \mu_i + \nu_{it}$, and

$$H_1 : E(u_{it} | Z_{i1}, ..., Z_{im}) \neq 0 \text{ on a set with positive measure.}$$

Our proposed test is based on the sample analogue of $J = E\left\{u_{it} E(u_{it}|Z_{it}) f(Z_{it})\right\}$. Note that $J = 0$ under H_0 and $J = E\left[\left\{E(u_{it}|Z_{it})\right\}^2 f(Z_{it})\right] > 0$ if the null is false. Hence, J serves as a proper candidate for testing H_0.

For notational simplicity, we impose an additional assumption that $f_t(\cdot) = f(\cdot)$ for all $t = 1, ..., m$. Let $\widehat{\theta}(z)$ denote a consistent estimator of $\theta(z)$ under the fixed effects assumption. Then a consistent estimator of u_{it} is given by $\widehat{u}_{it} = y_{it} - \widehat{\theta}(z_{it})$. Our feasible test statistic is given by

$$\begin{aligned} \widehat{J} &= (nm)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{m} \widehat{u}_{it} \widehat{E}_{-it}(\widehat{u}_{it}|Z_{it}) \widehat{f}_{-it}(Z_{it}) \\ &= \{nm(nm-1)\}^{-1} \sum_{i=1}^{n} \sum_{t=1}^{m} \sum_{j=1}^{n} \sum_{s=1, \{j,s\} \neq \{i,t\}}^{m} \widehat{u}_{it} \widehat{u}_{js} K_{h,it,js} \end{aligned}$$

where $K_{h,it,js} = K_h(Z_{it} - Z_{js})$, $K_h(v) = \prod_{\ell=1}^{q} h_l^{-1} k(v_\ell / h_\ell)$, $k(\cdot)$ is a univariate kernel function, and $E_{-it}(\widehat{u}_{it}|Z_{it}) = \{n(m-1)\}^{-1} \sum_{j=1}^{n} \sum_{s=1,\{js\}\neq\{it\}}^{m} \widehat{u}_{js} K_{h,it,js} / \widehat{f}_{-it}(Z_{it})$ and $\widehat{f}_{-it}(Z_{it}) = \{n(m-1)\}^{-1} \sum_{j=1}^{n} \sum_{s=1,js,\{js\}\neq\{it\}}^{m} K_{h.it,js}$ are the leave-one-out estimators of $E(u_{it}|Z_{it})$ and $f(Z_{it})$, respectively. It can be shown that $\widehat{J}$ is a consistent estimator of J. Hence, $\widehat{J} \to 0$ in probability under the null, and $\widehat{J} \to C$ if H_0 is false, where $C > 0$ is a positive constant. Therefore, one rejects H_0 when $\widehat{J}$ takes large positive values.

We conjecture that $\widehat{J}$, after proper normalization and centering, is asymptotically normally distributed. However, the derivation of such a result is quite complicated due to the iterative procedure involved in computing $\widehat{\theta}(\cdot)$. We leave the study of its asymptotic distribution to future research. Even if one derives the asymptotic distributions of $\widehat{J}$, it is well known that asymptotic theory does not provide good approximations for nonparametric kernel based tests in finite sample applications (e.g., see Härdle and Mammen, 1993; Lee and Ullah, 2000; Li and Wang, 1998). Therefore, we propose the following bootstrap procedure to approximate the finite sample null distribution of $\widehat{J}$.

3.1 A Bootstrap Procedure

Let $\widehat{u}_i = (\widehat{u}_{i1}, ..., \widehat{u}_{im})^{\mathrm{T}}$, where $\widehat{u}_{it} = Y_{it} - \widetilde{\theta}(Z_{it})$, and $\widetilde{\theta}(z)$ is the random effects estimator of $\theta(z)$. Compute the two-point wild bootstrap errors by $u_i^* = \{(1-\sqrt{5})/2\}\widehat{u}_i$ with probability $r = (1+\sqrt{5})/(2\sqrt{5})$ and $u_i^* = \{(1+\sqrt{5})/2\}\widehat{u}_i$ with probability $1-r$. Then generate Y_{it}^* via $Y_{it}^* = \widetilde{\theta}(Z_{it}) + u_{it}^*$. Call $\{Z_{it}, Y_{it}^*\}_{i=1,t=1}^{n,m}$ the bootstrap sample. Using the bootstrap sample to estimate $\theta(z)$ via the random effects method, denote the estimate by $\widetilde{\theta}^*(z)$, and then obtain the bootstrap residual by $\widehat{u}_{it}^* = Y_{it}^* - \widetilde{\theta}^*(Z_{it})$. The bootstrap test statistic $\widehat{J}^*$ is obtained as in $\widehat{J}$ except that $\widehat{u}_{it}$ $(\widehat{u}_{js})$ is replaced by

$\widehat{u}^*_{it}$ ($\widehat{u}^*_{js}$) wherever it occurs. This process is repeated a large number (B) of times. The empirical distribution of the B bootstrap statistics is then used to approximate the null distribution of the test statistic $\widehat{J}$.

The finite sample performance of the above bootstrap procedure is examined via simulations in Section 5.

4 A Partially Linear Model with Fixed Effects

Nonparametric regression suffers the curse of dimensionality problem when the dimension of the regressors is high. In this section we consider a semiparametric partially linear model where only a subset of the regressors enter the regression model nonparametrically. A partially linear panel data regression model with fixed effects is given by

$$Y_{it} = X_{it}^{\mathrm{T}}\beta + \theta(Z_{it}) + \mu_i + \nu_{it}, \qquad (i = 1, ..., n;\ \ t = 1, ..., m)$$

where X_{it} is of dimension $d \times 1$, and the other variables are as defined in Section 2.

Again we take the first difference to eliminate the fixed effects:

$$\widetilde{Y}_{it} = \widetilde{X}_{it}^{\mathrm{T}}\beta + \theta(Z_{it}) - \theta(Z_{i1}) + \widetilde{\nu}_{it}, \qquad (i = 1, ..., n;\ \ t = 2, ..., m) \tag{8}$$

where $\widetilde{Y}_{it} \equiv Y_{it} - Y_{i1}$, $\widetilde{X}_{it} \equiv X_{it} - X_{i1}$, and $\widetilde{\nu}_{it} = \nu_{it} - \nu_{i1}$. The criterion function for individual i is modified to

$$\begin{aligned} \mathcal{L}_i(\cdot) &= \mathcal{L}(Y_i, X_i, \beta, \theta_i) \\ &= -\frac{1}{2}\left(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta - \theta_i + \theta_{i1}e_{m-1}\right)^{\mathrm{T}} \Sigma^{-1} \left(\widetilde{Y}_i - \widetilde{X}_i^{\mathrm{T}}\beta - \theta_i + \theta_{i1}e_{m-1}\right), \end{aligned}$$

where $\widetilde{X}_i = (\widetilde{X}_{i2}, ..., \widetilde{X}_{im})^{\mathrm{T}}$. The derivative functions become

$$\begin{aligned} \mathcal{L}_{i,1\theta} &= -e_{m-1}\Sigma^{-1}\left(\widetilde{Y}_{i,-1} - \widetilde{X}_i\beta - \theta_i + \theta_{i1}e_{m-1}\right); \\ \mathcal{L}_{i,t\theta} &= c_{t-1}\Sigma^{-1}\left(\widetilde{Y}_i - \widetilde{X}_i\beta - \theta_i - \theta_{i1}e_{m-1}\right), \qquad \text{for } t \geq 2. \end{aligned}$$

and the second derivatives of $\mathcal{L}_{i,ts\theta}(\cdot)$ are analogous to those given in Section 2.

4.1 Computational Procedure

Following Wang et al (2005), we estimate $\theta(\cdot)$ and β by a profile-kernel approach. For given values of β and a current stage estimator $\widehat{\theta}(\cdot)$, we estimate $\theta(z)$ by $\widehat{\alpha}_0$, where $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ satisfy the following first order condition:

$$\begin{aligned} 0 &= \sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it},z)G_{it}(z,h) \\ &\qquad \times\mathcal{L}_{t\theta}\left[Y_i, X_i, \beta, \widehat{\theta}(Z_{i1},\beta), ..., \widehat{\alpha}_0 + \{(Z_{it}-z)/h\}^{\mathrm{T}}\widehat{\alpha}_1, ..., \widehat{\theta}(Z_{im,\beta})\right]. \end{aligned}$$

Compare (8) with (2). Let $\widehat{\theta}_y(\cdot)$ be the nonparametric estimator in model (2) and let $\widehat{\theta}_{x,r}(\cdot)$ be the nonparametric estimator in model (2) if Y_{it} is replaced by the r^{th} component of X_{it}. Further, let $\widehat{\theta}_x(z) = \{\widehat{\theta}_{x,1}(z), ..., \widehat{\theta}_{x,d}(z)\}^{\mathrm{T}}$. It is obvious by the linearity of the smoother and from (8) that

$$\widehat{\theta}(z,\beta) = \widehat{\theta}_y(z) - \widehat{\theta}_x(z)^{\mathrm{T}}\beta. \tag{9}$$

This means that $(\partial/\partial\beta)\widehat{\theta}(z,\beta) = -\widehat{\theta}_x(z)$. Therefore, we estimate β by the minimization of

$$\sum_{i=1}^{n}\begin{bmatrix} \widetilde{Y}_{i2} - \widetilde{X}_{i2}^{\mathrm{T}}\beta - \{\widehat{\theta}(Z_{i2},\beta) - \widehat{\theta}(Z_{i1},\beta)\} \\ \vdots \\ \widetilde{Y}_{im} - \widetilde{X}_{im}^{\mathrm{T}}\beta - \{\widehat{\theta}(Z_{im},\beta) - \widehat{\theta}(Z_{i1},\beta)\} \end{bmatrix}^{\mathrm{T}} \Sigma^{-1} \begin{bmatrix} \widetilde{Y}_{i2} - \widetilde{X}_{i2}^{\mathrm{T}}\beta - \{\widehat{\theta}(Z_{i2},\beta) - \widehat{\theta}(Z_{i1},\beta)\} \\ \vdots \\ \widetilde{Y}_{im} - \widetilde{X}_{im}^{\mathrm{T}}\beta - \{\widehat{\theta}(Z_{im},\beta) - \widehat{\theta}(Z_{i1},\beta)\} \end{bmatrix}.$$

We can now invoke (9) to get an explicit solution for $\widehat{\beta}$ and an explicit covariance matrix for the asymptotic distribution of $n^{1/2}(\widehat{\beta}-\beta)$. By defining $\widetilde{Y}_{it*} = \widetilde{Y}_{it} - \{\widehat{\theta}_y(Z_{it}) - \widehat{\theta}_y(Z_{i1})\}$, $\widetilde{Y}_{i*} = (\widetilde{Y}_{i2*}, ..., \widetilde{Y}_{im*})^{\mathrm{T}}$, $\widetilde{X}_{it*} = \widetilde{X}_{it} - \{\widehat{\theta}_x(Z_{it}) - \widehat{\theta}_x(Z_{i1})\}$ and $\widetilde{X}_{i*} = (\widetilde{X}_{i2*}, ..., \widetilde{X}_{im*})$, the estimate of β is given by

$$\widehat{\beta} = (\sum_{i=1}^{n}\widetilde{X}_{i*}\Sigma^{-1}\widetilde{X}_{i*}^{\mathrm{T}})^{-1}(\sum_{i=1}^{n}\widetilde{X}_{i*}\Sigma^{-1}\widetilde{Y}_{i*}).$$

Note that since we have a closed form solution for $\widehat{\beta}$, no iteration is needed. We estimate $\theta(\cdot)$ by the same method as discussed in Section 2 except now that $\widetilde{Y}_{it}$ is replaced by $\widetilde{Y}_{it} - \widetilde{X}_{it}^{\mathrm{T}}\widehat{\beta}$ whenever it occurs. At convergence, the resulting $\widehat{\theta}(z)$ has the same asymptotic distribution as described in Section 2. Then of course, we have a nonparametric regression model as covered in Section 2. Next, notice that $\widehat{\beta} - \beta = O_p(n^{-1/2})$ converges to zero faster than the nonparametric estimator $\widehat{\theta}(z) - \theta(z)$. Therefore, replacing β by $\widehat{\beta}$ will not affect the asymptotic distribution of $\widehat{\theta}(z)$.

To derive the asymptotic distribution of $\widehat{\beta}$, we first give some definitions. Let $\mathcal{G}$ denote the space of bounded, twice continuously differentiable functions. Define $g(\cdot) = (g_1(\cdot), g_2(\cdot), ..., g_d(\cdot))$ to be a $1 \times d$ vector function, with $g_j \in \mathcal{G}$. Define a $d \times 1$ vector function $\phi(\cdot)$ as the function that minimizes the following objective function:

$$\phi(\cdot) = \text{argmin}_{\phi(\cdot)=g(\cdot)\in\mathcal{C}^2} E\left[\begin{pmatrix} \widetilde{X}_{i2} - \{g(Z_{i2}) - g(Z_{i1})\} \\ \vdots \\ \widetilde{X}_{im} - \{g(Z_{im}) - g(Z_{i1})\} \end{pmatrix}^{\mathrm{T}} \Sigma^{-1} \begin{pmatrix} \widetilde{X}_{i2} - \{g(Z_{i2}) - g(Z_{i1})\} \\ \vdots \\ \widetilde{X}_{im} - \{g(Z_{im}) - g(Z_{i1})\} \end{pmatrix}\right].$$

The asymptotic distribution for the convergent $\widehat{\beta}$ is given by

$$n^{1/2}(\widehat{\beta} - \beta) \Rightarrow \text{Normal}(0, V^{-1}),$$

where

$$V = E\left[\begin{pmatrix} \widetilde{X}_{i2} - \{\phi(Z_{i2}) - \phi(Z_{i1})\} \\ \vdots \\ \widetilde{X}_{im} - \{\phi(Z_{im}) - \phi(Z_{i1})\} \end{pmatrix}^{\mathrm{T}} \Sigma^{-1} \begin{pmatrix} \widetilde{X}_{i2} - \{\phi(Z_{i2}) - \phi(Z_{i1})\} \\ \vdots \\ \widetilde{X}_{im} - \{\phi(Z_{im}) - \phi(Z_{i1})\} \end{pmatrix}\right].$$

Chamberlain (1992), Bickel et al. (1993), and Bickel and Kwon (2002) provide general treatment on inferences and efficient bounds analysis for semiparametric models. By following the same arguments as in Lin and Carroll (2006), one can show that V^{-1} is the semiparametric efficient lower bound for the asymptotic variance, among all estimators of β based upon the differences $Y_{it} - Y_{i1}$, when the regression errors ν_{it} in (1) have a Gaussian distribution.

5 Monte Carlo Simulations

This section uses Monte Carlo simulations to examine the finite sample performance of the panel data estimators. Following a methodology similar to Wang (2003), the following data generating process is used: $Y_{it} = \sin(2Z_{it}) + \mu_i + \nu_{it}$, where Z_{it} is *i.i.d.* uniform$[-1, 1]$, and ν_{it} is *i.i.d.* Normal$(0, 1)$. Let v_i denote an *i.i.d.* uniform$(-1, 1)$ sequence of random variables. We generate $\mu_i = v_i$ for the random effects model, and $\mu_i = v_i + Z_{i\cdot}$ for the fixed effects model, where $Z_{i\cdot} = m^{-1}\sum_{t=1}^{m} Z_{it}$. Note that Z_{it} and μ_i are correlated for the fixed effects model. The variances of ν_{it} and v_i are both fixed at unity. We use the Gaussian kernel function and the bandwidth is selected as $h = \widehat{\sigma}_z (nm)^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{it}\}_{i=1,t=1}^{n,m}$.

In the simulations reported below, we ignore the variance structure Σ in computing both the random effects and fixed effects estimators.[1] In this case, the random effects estimator is a simple local constant estimator (no iteration is needed). However, as noted previously, the fixed effects estimation procedure is iterative (even when one ignores the correlation structure of Σ^{-1}) and thus requires information on the previous iteration of θ. Thus we set the initial value of θ as follows: first, we use OLS to estimate the model using a fourth order polynomial. Next, we calculate the expected value of Y given Z for each observation using the OLS estimates and use these as our starting values for θ. Finally, we use the iterative method discussed in Section 2 to obtain estimates of the fixed effects estimator using the

---

[1] We also computed the MSE of the fixed effects estimator that takes care of the variance correlation structure by using the true Σ^{-1}, as expected, it gives smaller MSE compared with the one that ignores the variance correlation structure. These results are not reported here.

initial estimate of θ as described above. The convergence criterion is set to be $\sum_{i=1}^{n}\sum_{t=1}^{m}\{\widehat{\theta}_{[\ell]}(Z_{it}) - \widehat{\theta}_{[\ell-1]}(Z_{it})\}^2/\sum_{i=1}^{n}\sum_{t=1}^{m}\widehat{\theta}_{[\ell-1]}(Z_{it})^2 < 0.001$. Unlike the random effects iterative procedure in Wang (2003) which performs well with a one-step iteration, the fixed effects estimation generally needs an average of five to six iterations to obtain convergence.

We use both fixed effects and random effects methods to estimate $\theta(\cdot)$, and compute the average mean squared error (AMSE) by AMSE $= M^{-1}\sum_{p=1}^{M}(nm)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\{\widehat{\theta}(Z_{it,p}) - \theta_0(Z_{it,p})\}^2$, where the subscript p denotes the p^{th} replication. In each experiment we use $M = 5,000$ replications. The number of time periods (m) is fixed at 3, while the number of cross-sections (n) is varied to be 50, 100 and 200. The estimation results are given in Table 1. When the data generating process is that of a random effects model, we see that the random effects estimator has a smaller AMSE than the fixed effects estimator. This result is expected because the fixed effects estimator is not efficient. Also as expected, for both estimators, the AMSE decreases quickly as n gets larger. Next, when the data are generated via a fixed effects model, the regressor Z_{it} and the fixed effects μ_i are correlated. In this case the random effects estimator is inconsistent. Indeed, Table 1 shows that the random effects AMSE does not decrease as the sample size increases. In contrast, the fixed effects estimator that removes the fixed effects leads to consistent estimation results. Its AMSE decreases rapidly as n increases.

Next, we examine the finite sample performance of the nonparametric test for detecting a fixed effects model against a random effects model. The data generating process is the same as above except that we add another parameter (c_0) in generating the individual effects: $\mu_i = v_i + c_0 Z_{i\cdot}$. $c_0 = 0$ gives the random effects model, and $c_0 \neq 0$ leads to the fixed effects model. We consider $c_0 = 0$, 0.25, 0.5. The number of replications (M) here is set equal to 1000 in each setting, and the number of bootstraps within each replication is set at 400. From Table 2 we observe that the estimated sizes of the $\widehat{J}$ test (the case of $c_0 = 0$) are close to the nominal sizes. The last two rows ($c_0 = 0.25$, 0.5) give the estimated power of the $\widehat{J}$ test. We observe that the power increases rapidly as either the sample size increases, or as the correlation between the individual effects and the regressor increases (i.e., as c_0 increases). The limited simulation results seem to suggest that the bootstrap-based $\widehat{J}$ test performs well for the typical panel data situation of large n and small m.

6 An Empirical Application

In this section we apply the nonparametric estimation methods proposed in this paper to estimate the relationship between caloric intake and income using data obtained from the China Health and Nutrition Survey (CHNS). The CHNS project was designed to examine the effects of health, nutrition, and family planning policies and programs implemented by the national and local governments. It also looks to see how the social and economic transformation of the Chinese society is affecting the health and nutritional status of its population. The survey was conducted by an international team of researchers whose backgrounds include nutrition, public health, economics, sociology, Chinese studies, and demography. The panel survey data we use is over $m = 3$ time periods: 1989, 1991 and 1993.

The relationship between nutritional intake and income underlines the link between population health status and economic development in developing countries. Food can be regarded as the most important "necessity" when income is low. The income elasticity of nutrition is of particular interest to many researchers (e.g., see Deaton, 1997; Subramanian and Deaton, 1996). While most empirical studies along this line rely on a particular parametric functional form, typically a log-log regression, one advantage of our method is that there are no functional form assumptions besides some regularity conditions on smoothness.

The variables used in this study include: daily caloric intake (daycal, the dependent variable), per capita income adjusted by inflation (income), age, urban (a dummy variable indicating whether the individual is an urban or a rural resident) and gender (dummy). We are mainly interested in the relationship between caloric intake and income, and illustrate this relationship for urban men. We initially considered a partially linear specification $Y = X^{\mathrm{T}}\beta + \theta(Z) + u$, where $Y = \log(\text{daycal})$, $X =(\text{age}, \text{age}^2)^{\mathrm{T}}$, and $Z = \text{income}$. However, none of our results disclosed a statistically significant age effect, and thus in what follows we only display the nonparametric regression fit.

In the estimation of both the nonparametric fixed effects and random effects models we use a Gaussian kernel function. With the random effects approach we use the so-called working independence method which ignores any correlation structure: accounting for the correlation in the random effects estimator using the method of Wang (2003) shows little difference.

The estimation results are given in Figure 1. The figure plots the estimated curves for urban males (465 individuals). In the random effects model, the caloric intake is basically flat or even slightly decreasing as income increases through approximately 1,000 yuan, approximately \$120 U.S. dollars. This result contradicts almost all previous studies on caloric intake and income, which find that the relationship between the two is typically positive at very low income levels (e.g., see Deaton, 1997). The results displayed are for a bandwidth equal to 400, but the basic trend is relatively insensitive to the bandwidth choice. In contrast, the fixed effects estimation with a bandwidth equal to 400 shows that caloric intake first rises as income increases, and then decreases, a result more in keeping with economic theory: the basic shape is also not much affected by the bandwidth. Penalized regression splines using a Bspline basis gave essentially the same conclusions.

In Figure 2, we compare the kernel fits to fits from a cubic regression model. The power of flexible regression modeling is evident here, because the cubic fits either exaggerate the decline in caloric intake at higher income levels (random effects) or the increase in caloric intake at lower income levels (fixed effects).

We performed the bootstrap testing procedure based on $\widehat{J}$. For distribution of (inflation-adjusted) income over the periods of 1989, 1991, 1993 have quite close shapes and this justifies the use of the $\widehat{J}$ test by assuming that $f_t(\cdot) = f(\cdot)$ for all $t = 1, .., m$. We computed the p-value of the test statistic based on $1,000$ bootstrap statistics where the p-value is defined as the percentage of bootstrap statistics that are greater than $\widehat{J}$. The $1,000$ bootstrap statistic yields a p-value of 0.031 using the urban male data, suggesting that the unexpected shape of the random effects fit is due to violations of that model. That is, the correlation between the unobserved individual effects and the regressor (income) give a biased and inconsistent estimation result using the random effects specification. In contrast, the fixed effects estimation result is more plausible and consistent with the economic theory that for low income individuals, caloric intake increases as income increases, while for high income individuals, caloric intake and income are likely to be negatively correlated.

7 Discussion

In this paper we proposed using an kernel-based methodology to estimate a nonparametric panel data model with fixed effects. We suggested using a bootstrap procedure to test for the presence of random effects versus fixed effects in a nonparametric panel data set. We also extended the estimation method to the case of a partially linear fixed effects model. The empirical results on studying the relationship between income and caloric intake show that the fixed effects approach reveals a reasonable nonlinear relationship between daily nutritional intake and income, while the random effects specification leads to an implausible global negative correlation between the caloric intake and income.

Acknowledgments

Carroll's research was supported by a grant from the National Cancer Institute (CA-57030), and by the Texas A&M Center for Environmental and Rural Health via a grant from the National Institute of Environmental Health Sciences (P30-ES09106). Li's research was partially supported by the Private Enterprise Research Center, Texas A&M University.

Appendix: Sketch of Technical Arguments

Lemma 1 Let $\widehat{\theta}_{[\ell]}(z)$ and $\widehat{\theta}^{(1)}_{[\ell]}(z)$ be the l^{th} step iteration estimators of $\theta(z)$ and $\theta^{(1)}(z)$, respectively, then it satisfies the following equation

$$\widehat{\theta}_{[\ell]}(z) - \theta(z) = \frac{\kappa_2}{2}\sum_{r=1}^{q} h_r^2 b_r(z) - \frac{1}{n\Omega(z)}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)\epsilon_{it} \tag{A.1}$$

$$-\frac{1}{n\Omega(z)}\sum_{i=1}^{n}\sum_{t=1}^{m}\sum_{s\neq t}^{m} K_h(Z_{it}, z)\mathcal{L}_{i,ts\theta}(\cdot)\{\widehat{\theta}_{[\ell-1]}(z) - \theta(z)\} + o_p\{\sum_{r=1}^{q} h_r^2 + (nh_1\cdots h_q)^{-1/2}\},$$

where $\epsilon_{it} = \mathcal{L}_{i,t\theta}\{\widetilde{Y}_i, \theta(Z_{i1}), ..., \theta(Z_{it}), ..., \theta(Z_{im})\}$.

Proof:

By Taylor expansion of the first order equation (5) with respect to $\{\widehat{\theta}_{[\ell]}(z), \widehat{\theta}^{(1)}_{[\ell]}(z)\}$ at $\{\theta(z), \theta^{(1)}(z)\}$, we have that

$$\begin{aligned} 0 &= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)G_{it}(z,h)\mathcal{L}_{i,t\theta}(\cdot) \\ &\quad + n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)G_{it}(z,h)G_{it}(z,h)^{\mathrm{T}}\mathcal{L}_{i,tt\theta}(\cdot)\begin{pmatrix} \widehat{\theta}_{[\ell]}(z) - \theta(z) \\ \widehat{\theta}^{(1)}_{[\ell]}(z) - \theta^{(1)}(z) \end{pmatrix} \\ &\quad + o_p\{\sum_{r=1}^{q} h_r^2 + (nh_1\cdots h_q)^{-1/2}\}, \qquad \text{(A.2)} \end{aligned}$$

where the argument $(\cdot)$ is at $[\widetilde{Y}_i, \widehat{\theta}_{[\ell-1]}(Z_{i1}), ..., \theta(z) + \{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z), ..., \widehat{\theta}_{[\ell-1]}(Z_{im})]$. Note that $n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)G_{it}G_{it}^{\mathrm{T}}\mathcal{L}_{i,t\theta}(\cdot) \sim \sum_{t=1}^{m} E[K_h(Z_{it}, z)G_{it}G_{it}^{\mathrm{T}}\mathcal{L}_{i,tt\theta}]$ which converges to $-\Omega(z)\mathrm{diag}(1, \kappa_2 I_q)$, where $\kappa_2 = \int k(v)v^2 dv$. Hence, (A.2) leads to

$$\Omega(z)\{\widehat{\theta}(z) - \theta(z)\} = n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)\mathcal{L}_{i,t\theta}(\cdot) + o_p(\eta_n) \equiv A_n + o_p(\eta_n), \qquad \text{(A.3)}$$

where $\eta_n = \sum_{r=1}^{q} h_r^2 + (nh_1\cdots h_q)^{-1/2}$.

We decompose A_n into $A_n = A_{1n} + A_{2n}$, where A_{1n} is obtained from A_n with $\widehat{\theta}(\cdot)$ replaced by $\theta(\cdot)$, and $A_{2n} = A_n - A_{1n}$. Thus,

$$\begin{aligned} A_{1n} &= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z) \\ &\quad \times \mathcal{L}_{i,t\theta}\left[\widetilde{Y}_i, \theta(Z_{i1}), ..., \theta(z) + \{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z), ..., \theta(Z_{im})\right]. \qquad \text{(A.4)} \end{aligned}$$

We further write $A_{1n} = A_{1n1} - A_{1n2}$, where

$$\begin{aligned} A_{1n1} &= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)G_{it}(z,h)\epsilon_{it}; \qquad \text{(A.5)} \\ A_{1n2} &= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)\Big(\mathcal{L}_{i,t\theta}\{\widetilde{Y}_i, \theta(Z_{i1}), ...\theta(Z_{it}), ..., \theta(Z_{im})\} \\ &\quad - \mathcal{L}_{i,t\theta}[\widetilde{Y}_i, \theta(Z_{i1}), ..., \theta(z) + \{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z), ..., \theta(Z_{im})]\Big) \\ &= \frac{1}{2n}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)(Z_{it}-z)\theta^{(2)}(z)(Z_{it}-z)^{\mathrm{T}}\mathcal{L}_{i,tt\theta}(\cdot) + o_p(\eta_n) \\ &= \frac{\kappa_2}{2}\sum_{r=1}^{q} h_r^2\theta_{rr}(z)\Omega(z) + o_p(\eta_n), \qquad \text{(A.6)} \end{aligned}$$

where $\Omega(z) = -\sum_{t=1}^{m} f_t(z)E(\mathcal{L}_{i,tt}|Z_{it} = z) = (1-m)/(\sigma_\nu^2)$, $\theta_{rr}(z) = \partial^2\theta(z)/\partial z_r^2$, $\theta^{(2)}(z)$ is the $q \times q$ second order derivative matrix of $\theta(z)$, and $\eta_n = \sum_{r=1}^{q} h_r^2 + (nh_1 \cdots h_q)^{-1/2}$. In addition,

$$\begin{aligned}
A_{2n} &= A_n - A_{1n} = n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z) \\
&\quad \times \mathcal{L}_{i,j\theta}\left[\widetilde{Y}_i, \widehat{\theta}_{[\ell-1]}(Z_{i1}), ..., \theta(z) + \{(Z_{it}-z)/h\}^{\mathrm{T}}\theta^{(1)}(z), ..., \widehat{\theta}_{[\ell-1]}(Z_{im})\right] - A_{1n} \\
&= n^{-1}\sum_{i=1}^{n}\sum_{t=1}^{m}\sum_{s\neq t}^{m} K_h(Z_{it}, z)G_{it}(z,h)\mathcal{L}_{i,ts\theta}(\cdot)\{\theta_{[\ell-1]}(Z_{is}) - \theta(Z_{is})\} \\
&\quad + o_p(\sum_{r=1}^{q} h_r^2 + (nh_1 \cdots h_q)^{-1/2}). \qquad (A.7)
\end{aligned}$$

Accumulating these results yields (A.1).

Under the assumption that $n(h_1 \cdots h_q)^2 \to \infty$ and $n\sum_{r=1}^{q} h_r^6 \to 0$ as $n \to \infty$, at convergence, Lemma 1 leads to

$$\widehat{\theta}(z) - \theta(z) = \frac{\kappa_2}{2}\sum_{r=1}^{q} h_r^2 b_r(z) + \frac{1}{n\Omega(z)}\sum_{i=1}^{n}\sum_{t=1}^{m} K_h(Z_{it}, z)\epsilon_{it} + o_p(\eta_n), \qquad (A.8)$$

which in turn leads to the asymptotic bias and variance of $\widehat{\theta}(z)$ given by $(\kappa_2/2)\sum_{r=1}^{q} h_r^2 b_r(z) + o(\sum_{r=1}^{q} h_r^2)$ and $\kappa^q/\{nh_1 \cdots h_q\Omega(z)\} + o\{(nh_1 \cdots h_q)^{-1}\}$, respectively, by using the same arguments as in Wang (2003), or as in the proof of Lemma A.1 in Lin and Carroll (2006).

References

Bickel, P. J. Klaassen, A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Inference in Semiparametric Models.* John Hopkins University Press, Baltimore.

Bickel, P. J. and Kwon, J. (2002). Inference for Semiparametric Models: Some Current Frontiers (with Discussions). *Statistica Sinica*, 11, 863-960.

Chamberlain, G. (1992). Efficiency Bounds for Semiparametric Regression. *Econometrica*, 60, 567-596.

Deaton, A. (1997). *The Analysis of Household Surveys.* Baltimore: The Johns Hopkins University Press.

Härdle, W. and Mammen, E. (1993). Comparing Nonparametric versus Parametric Regression Fits. *Annals of Statistics*, 21, 1926-1947.

Henderson, D. J. and Ullah, A. (2005). A Nonparametric Random Effects Estimator. *Economics Letters*, 88, 403-407.

Ke, C. and Y. Wang (2001). Semiparametric Nonlinear Mixed-Effects Models and their Applications. *Journal of the American Statistical Association*, 96, 1272-1281.

Lee, T.-H. and Ullah, A. (2000). Nonparametric Bootstrap Tests for Neglected Nonlinearity in Time Series Regression Models. *Journal of Nonparametric Statistics*, 13, 425-451.

Li, Q. and Stengos, T. (1996). Semiparametric Estimation of Partially Linear Regression Models. *Journal of Econometrics*, 71, 389-397.

Li, Q. and Wang, S. (1998). A Simple Consistent Bootstrap Test for a Parametric Regression Function. *Journal of Econometrics*, 87, 145-165.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and Nonparametric Regression Analysis of Longitudinal Data (with Discussion). *Journal of the American Statistical Association*, 96, 103-126.

Lin, X., and Carroll, R. J. (2000). Nonparametric Function Estimation for Clustered Data when the Predictor is Measured without/with Error. *Journal of the American Statistical Association*, 95, 520-534.

Lin, X. and Carroll, R. J. (2001). Semiparametric Regression for Clustered Data using Generalised Estimation Equations. *Journal of the American Statistical Association*, 96, 1045-1056.

Lin, X. and Carroll, R. J. (2006). Semiparametric Estimation in General Repeated Measures Problems. *Journal of the Royal Statistical Society, Series B*, 68, 68-88.

Lin, X., Wang, N., Welsh, A. H. and Carroll, R. J. (2004). Equivalent Kernels of Smoothing Splines in Nonparametric Regression for Longitudinal/Clustered Data. *Biometrika*, 91, 177-194.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a Bivariate Additive Model By Local Polynomial Regression. *Annals of Statistics*, 25, 186-212.

Ruckstuhl, A. F., Welsh, A. H. and Carroll, R. J. (2000). Nonparametric Function Estimation of the Relationship Between Two Repeatedly Measured Variables. *Statistica Sinica*, 10, 51-71.

Subramanian, S. and Deaton, A. (1996). The Demand for Food and Calories. *Journal of Political Economy* 104, 133-162.

Ullah, A. and Roy, N. (1998). Nonparametric and Semiparametric Econometrics of Panel Data, in *Handbook of Applied Economics Statistics*, A. Ullah and D. E. A. Giles, eds., Marcel Dekker, New York, 579-604.

Wang, N. (2003). Marginal Nonparametric Kernel Regression Accounting for Within-Subject Correlation. *Biometrika*, 90, 43-52.

Wang, N., Carroll, R. J. & Lin, X. (2005). Efficient Semiparametric Marginal Estimation For Longitudinal/Clustered data. *Journal of the American Statistical Association*, 100, 147-157.

Wu, H. and Zhang, J. Y. (2002). Local Polynomial Mixed-Effects Models for Longitudinal Data. *Journal of the American Statistical Association*, 97, 883-897.

	Random Effects			Fixed Effects		
Data Process	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
Random	.0346	.0192	.0107	.0537	.0301	.0161
Fixed	.2895	.2880	.2905	.0721	.0399	.0211

Table 1: Average mean squared errors (AMSE) of the fixed and random effects estimators when the data generation process is a random effects model and when it is a fixed effects model.

Size of the $\widehat{J}$ test (m=3)									
	$n = 50$			$n = 100$			$n = 200$		
c_0	1%	5%	10%	1%	5%	10%	1%	5%	10%
0	.012	.056	.112	.015	.057	.109	.011	.054	.108
0.25	.176	.404	.518	.376	.626	.734	.672	.872	.922
0.5	.578	.834	.910	.935	.990	1.00	1.00	1.00	1.00

Table 2: Estimated size and power for the $\widehat{J}$ test for the null hypothesis that the random effects model is true.

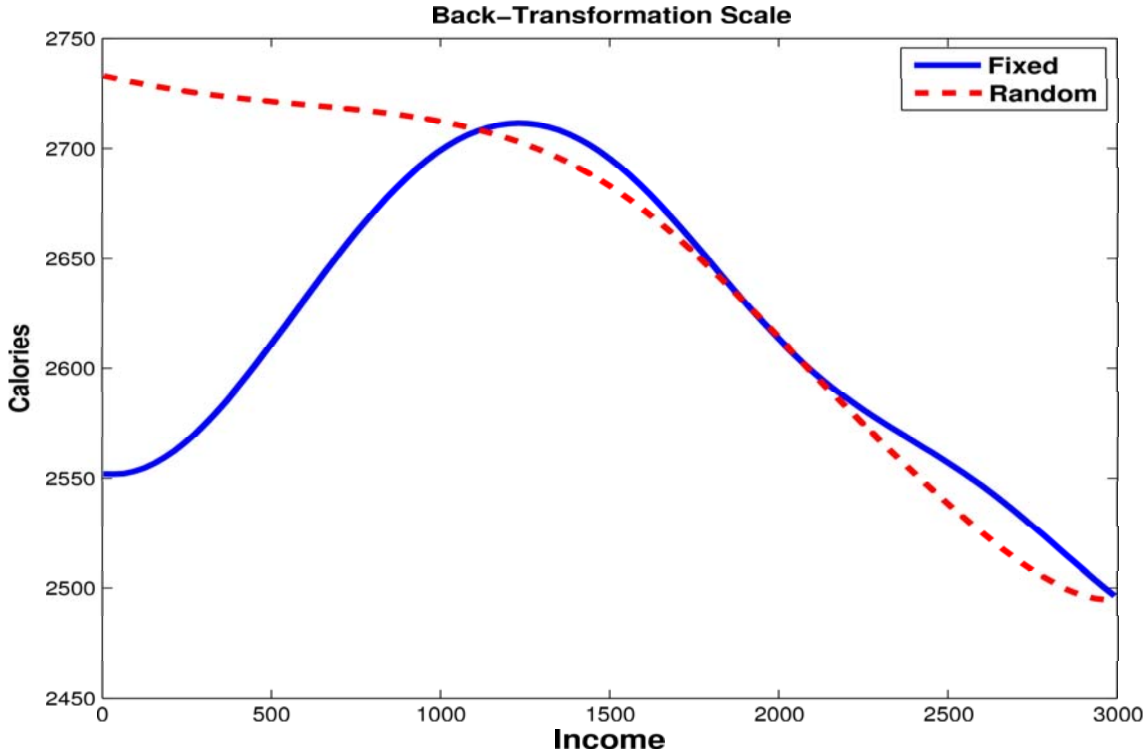

Figure 1: Fit to urban males. The functions were fit in the log scale and then back-transformed. The fixed effects fit is the solid line, while the random effects fit is the dashed line. The bandwidth for the fixed and random effects was 400.

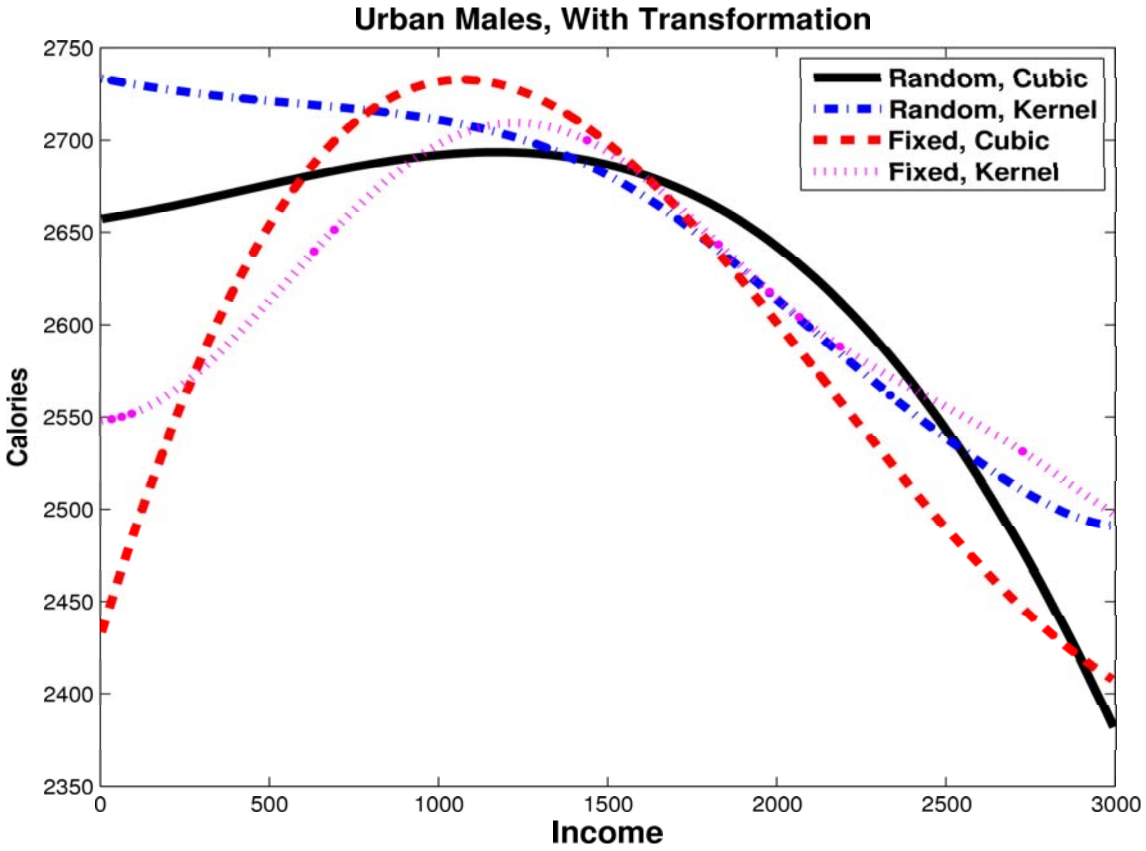

Figure 2: Fit to urban males. The functions were fit in the log scale and then back-transformed. Solid line: random effects cubic regression estimator. Dashed line: fixed effects cubic regression estimator. Dash-dotted line: random effects kernel regression estimator. Dotted line: fixed effects kernel regression estimator.