SEMIPARAMETRIC ESTIMATION OF REGRESSION MODELS FOR PANEL DATA

by

Joel L. Horowitz*
Department of Economics
University of Iowa
Iowa City, IA 52242

and

Marianthi Markatou
Department of Statistics
Columbia University
New York, New York 10027

June 1993

ABSTRACT

Linear models with error components are widely used to analyze panel data. Some applications of these models require knowledge of the probability densities of the error components. Existing methods handle this requirement by assuming that the densities belong to known parametric families of distributions (typically the normal distribution). This paper shows how to carry out nonparametric estimation of the densities of the error components, thereby avoiding the assumption that the densities belong to known parametric families. The nonparametric estimators are applied to an earnings model using data from the Current Population Survey. The model's transitory error component is not normally distributed. Use of the nonparametric density estimators yields estimates of the probability that individuals with low earnings will become high earners in the future that are much lower than the estimates obtained under the assumption of normally distributed error components.

JEL Classification: C13, C14, C23

____* Author to whom correspondence should be addressed. We thank Kevin Murphy, George Neumann, Whitney Newey and Gene Savin for helpful suggestions concerning this research. Susan Greene provided research assistance. The research was supported in part by NSF grant no. DMS-9208820.

SEMIPARAMETRIC ESTIMATION OF REGRESSION MODELS FOR PANEL DATA

1. Introduction

Linear models with error components are widely used in applied econometrics to analyze panel data, and there is a large literature on how to carry out estimation and inference with such models. Chamberlain (1984) and Hsiao (1986) review this literature. Some applications, such as estimation of transition probabilities and first passage times, require knowledge of the probability densities of the error components. Existing methods handle this requirement by assuming that the densities belong to known parametric families of distributions (typically the normal distribution). In this paper, we show how to carry out nonparametric estimation of the probability densities of the error components, thereby avoiding the assumption that the densities belong to known parametric families. To illustrate the usefulness of the nonparametric estimators, we use a well-known panel data set to estimate an earnings model. The model's transitory error component is not normally distributed. Use of the nonparametric density estimators yields estimates of the probability that individuals with low earnings will become high earners in future that are much lower than the estimates obtained under the assumption of normally distributed error components.

One drawback of our estimators is that they converge very slowly. In an important special case that we investigate in detail, the rate of convergence is $(\log n)^{-1}$, where n is the number of individuals in the panel. This is an excruciatingly slow rate of convergence by conventional standards. We show, however, that it is the fastest possible rate under our relatively weak assumptions. Thus, slow convergence is intrinsic to the problem we are dealing with, not a defect of the estimators.

Because of their slow rate of convergence, our estimators are likely to be useful only with fairly large data sets. In the empirical parts of this paper, we use a data set consisting of observations on 1779 individuals sampled randomly from the Current Population Survey (CPS). Other large panel data sets that are widely used in economics include the National Longitudinal Surveys of Labor Market Experience and the Panel Study of Income Dynamics. We present Monte Carlo evidence that our estimators work well with panels of 1000 individuals, which is below the sizes of the data sets just mentioned.

The main technical problem that we must solve is deconvolution of a probability density. In deconvolution, one wishes to estimate a certain density function but cannot sample the random variable that has this density. Instead, one samples a random variable whose density is the convolution of the density of interest and another density. The problem of deconvolution has been investigated by Carroll and Hall (1988), Stefanski and Carroll (1990, 1991), and Fan (1991) in the context of estimating errors-in-variables models. Although the models we are concerned with here are different from those considered by these authors, our techniques are similar to theirs, and we use several of their results in our analysis.

The remainder of this paper is organized as follows. Section 2 presents the model we analyze and gives an empirical example that illustrates why density estimators that do not require parametric assumptions may be useful for the analysis of panel data. Section 3 presents our estimators and describes their properties. Section 4 describes the results of a Monte Carlo investigation of the behavior of the estimators. Section 5 illustrates the use of the estimators in an application. Section 6 presents the conclusions of the paper. The proofs of theorems are in the appendix.

2. The Model and an Example

a. The Model

We consider the following model:

$$Y_{jt} = \beta' X_{jt} + U_j + \varepsilon_{jt}; \quad j = 1,...,n; \quad t = 1,...,T \qquad (2.1)$$

where Y_{jt} is the observed value of the dependent variable for individual j at time t, X_{jt} is a vector of observed explanatory variables for individual j and time t; β is a conformable vector of parameters to be estimated; U_j is an unobserved, random, individual effect; and ε_{jt} is an unobserved random variable that is independently and identically distributed across both individuals and time periods. We assume that U and ε are independent of one another and that their distributions satisfy regularity conditions that are given below. In what follows, we will refer to U as the permanent component and ε as the transitory component of the total error U + ε. We shall be concerned with the situation in which n is large but T may not be. Thus, asymptotic distributional results will be developed under the assumption that $n \to \infty$ while T stays constant.

The standard methods for estimating the parameters of (2.1) are based on least squares. See Hsiao (1986) for a description. The only information about the distributions of

U and ε provided by these methods consists of variance estimates. Thus, applications requiring knowledge of the distributions of U and ε are possible only if these distributions are known up to scale. This can lead to difficulties for reasons that are illustrated in Section 2b.

b. An Example

To illustrate why nonparametric density estimators may be useful in models such as (2.1), we consider a model for annual earnings estimated using a panel of length T = 2. The estimation data set consists of 1779 white, male, full-time workers, aged 18-65 years, sampled randomly from the matched March 1986 and 1987 CPS. Each individual is included in the sample for each year, so the data form a panel of length T = 2.

In this section, we are concerned with investigating the distribution of the transitory error component ε. We do this by examining the empirical distribution of the residuals from ordinary least-squares estimation of the differenced model

$$Y_{j2} - Y_{j1} = \beta'(X_{j2} - X_{j1}) + \varepsilon_{j2} - \varepsilon_{j1} \; ; \; j = 1,...,n \qquad (2.2)$$

The dependent variable Y in (2.2) is the natural logarithm of real annual earnings from wages. The explanatory variables X are listed in Table 1. There is no intercept.[1]

To illustrate why nonparametric estimators of the densities of the error components may be useful, we carry out a graphical test of normality of the distribution of ε. If ε is normally distributed, the residuals from (2.2) are also normally distributed up to random sampling error. Thus, we can test for normality of ε by testing for normality of the residuals in (2.2). Let F_n denote the empirical distribution function of these residuals, and let Φ denote the cumulative normal distribution function. If the residuals are normally distributed up to random sampling error, a plot of $\Phi^{-1}[F_n(v)]$ against v will consist of scatter around a straight line. Figure 1 shows this plot. It is clear that ε is not normally distributed; the tails of its distribution are thicker than those of the normal distribution. The implications of this finding for applications will be illustrated in Sections 4 and 5.

Since the distribution of ε appears to be thick-tailed, one might consider approximating it with a Cauchy distribution. If ε is Cauchy distributed, the least-squares estimator of β in (2.2) is not consistent, but the least-absolute-deviations (LAD) estimator of β is. The residuals from the model estimated by LAD will be Cauchy distributed up to random sampling error, and a plot of $\tan\{\pi[F(v) - 0.5]\}$ against v will consist of scatter around a straight line.[2] Figure 2 shows the resulting plot for our data. It is clear that ε is not Cauchy distributed; the tails of its distribution are too thin.

Of course, the fact that ε has neither the normal nor the Cauchy distribution does not rule out the possibility of finding a tractable parametric family of distributions that fits the data. However, a parametric model that is found through a specification search amounts to an informal nonparametric estimator whose statistical properties are unknown. We now present formal nonparametric estimators of the densities of the error components and discuss their statistical properties.

3. Nonparametric Estimators of the Densities of ε and U

This section presents the nonparametric estimators of the densities of ε and U in (2.1). We begin with an informal discussion that motivates the estimators.

a. Motivation

Let b_n be a $n^{1/2}$-consistent estimator of β in (2.1), possibly one of the least-squares estimators described by Hsiao (1986), among others. Let $\{_{jt}$: j = 1,...,n; t = 1,...,T} denote the residuals from the corresponding estimate of (2.1):

$$_{jt} = Y_{jt} - b_n'X_{jt} \qquad (3.1)$$

In addition, let $\{_{jt}$: j = 1,...,n; t = 2,...,T} denote the residuals from the estimate of the differenced model for $Y_{jt} - Y_{j1}$:

$$_{jt} = (Y_{jt} - Y_{j1}) - b_n'(X_{jt} - X_{j1}). \qquad (3.2)$$

Observe that as $n \to \infty$ while T remains fixed, $_{jt}$ converges in distribution to the random variable $\nu \equiv U + \varepsilon$, and $_{jt}$ converges in distribution to the random variable η that is distributed as the difference between two independent realizations of ε. Thus, the estimation data $\{Y_{jt}, X_{jt}\}$ provide estimates of random variables that are distributed as ν and η. However, the data do not provide estimates of U and ε, whose distributions are the objects of interest in this discussion. The problem that must be solved here is to obtain estimates of the distributions of U and ε from estimates of ν and η. This amounts to deconvoluting two densities because the probability density of ν is the convolution of the densities of U and ε, and the probability density of η is the convolution of the density of ε with itself.

To see how the densities of U and ε can be estimated, let h_ν and h_η denote the characteristic functions of ν and η, respectively. That is

$$h_\nu(\tau) = \int_{-\infty}^{\infty} e^{iz\tau} f_\nu(z)dz$$

and

$$h_\eta(\tau) = \int_{-\infty}^{\infty} e^{iz\tau} f_\eta(z)dz,$$

where $i = (-1)^{1/2}$, f_ν is the probability density function of ν, and f_η is the probability density function of η. Let h_U and h_ε denote the characteristic functions of U and ε, respectively. Then it is easily shown that

$$h_\nu(\tau) = h_U(\tau)h_\varepsilon(\tau)$$

and

$$h_\eta(\tau) = |h_\varepsilon^2(\tau)| ,$$

where $|\bullet|$ denotes the modulus of the complex variable between the bars. If the distribution of ε is such that $h_\varepsilon(\tau)$ is real and strictly positive for all finite τ, then

$$h_\varepsilon(\tau) = h_\eta^{1/2}(\tau)$$

and

$$h_U(\tau) = h_\nu(\tau)/h_\eta^{1/2}(\tau) .$$

It follows from the inversion formula for characteristic functions that the densities of ε and U (f_ε and f_U, respectively) are given by

$$f_\varepsilon(z) = (1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau} h_\eta^{1/2}(\tau)\, d\tau \tag{3.3}$$

and

$$f_U(z) = (1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau} [h_\nu(\tau)/h_\eta^{1/2}(\tau)]d\tau. \tag{3.4}$$

Equations (3.3) and (3.4) would solve the deconvolution problem if h_ν and h_η were known.

Of course, h_ν and h_η are not known in applications, but they can be estimated by the empirical characteristic functions of and . These are

$$ {}_{\nu}(\tau) = (nT)^{-1} \sum_{j=1}^{n} \sum_{t=1}^{T} \exp(i\tau {}_{jt}). \qquad (3.5) $$

and

$$ {}_{\eta}(\tau) = [n(T-1)]^{-1} \sum_{j=1}^{n} \sum_{t=2}^{T} \exp(i\tau {}_{jt}), \qquad (3.6) $$

It is shown in the appendix that under regularity conditions, $|{}_\eta|$ and ${}_\nu$ consistently estimate h_η and h_ν, respectively. Thus, one might consider estimating f_ε and f_U by replacing h_η and h_ν with $|{}_\eta|$ and ${}_\nu$ in (3.3) and (3.4). In general, however, the integrals in (3.3) and (3.4) do not exist when h_η and h_ν are replaced with empirical analogs. To overcome this problem, we convolute the empirical distributions of and with the distribution of a suitable continuously distributed random variable that becomes degenerate as $n \to \infty$. This amounts to kernel smoothing of the empirical distributions of and .

To carry out the smoothing, let g be a bounded, real characteristic function with support [-1,1], and let ζ be the random variable that has this characteristic function. Let $\{\lambda_{n\varepsilon}\}$ and $\{\lambda_{nU}\}$ be sequences of positive constants satisfying $\lambda_{n\varepsilon} \to 0$ and $\lambda_{nU} \to 0$ as $n \to \infty$. The idea behind the smoothing procedure is to use the inversion formula for characteristic functions to estimate the densities of the random variables $\varepsilon + \lambda_{n\varepsilon}\zeta$ and $U + \lambda_{nU}\zeta$. Since $\lambda_{n\varepsilon} \to 0$ and $\lambda_{nU} \to 0$ as $n \to \infty$, the resulting estimators converge to the densities of ε and U.

Specifically, observe that $h_\varepsilon(\tau)g(\lambda_{n\varepsilon}\tau)$ is the characteristic function of $\varepsilon + \lambda_n\zeta$ evaluated at the point τ, and $h_\nu(\tau)g(\lambda_{nU}\tau)$ is the characteristic function of $\nu + \lambda_n\zeta$ evaluated at τ. These quantities can be estimated by $|\hat{h}_\eta(\tau)|^{1/2}g(\lambda_{n\varepsilon}\tau)$ and $[\hat{h}_\nu(\tau)|\hat{h}_\eta(\tau)|^{-1/2}]g(\lambda_{nU}\tau)$. The corresponding estimators of f_ε and f_ν are:

$$\hat{f}_\varepsilon(z) = (1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}|\hat{h}_\eta(\tau)|^{1/2}g(\lambda_{n\varepsilon}\tau)d\tau \qquad (3.7)$$

and

$$\hat{f}_U(z) = (1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}\hat{h}_\nu(\tau)|\hat{h}_\eta(\tau)|^{-1/2}g(\lambda_{nU}\tau)d\tau. \qquad (3.8)$$

These are the estimators of f_ε and f_U that are proposed in this paper. We now give conditions under which they are consistent and discuss their rates of convergence.

b. Consistency and Rates of Convergence of the Estimators

We now show that under regularity conditions, $\hat{f}_\varepsilon(z)$ and $\hat{f}_U(z)$ converge in probability to $f_\varepsilon(z)$ and $f_U(z)$, respectively, uniformly over $z \in (-\infty,\infty)$. We make the following assumptions:

A1. The distributions of U and ε are absolutely continuous with respect to Lebesgue measure, and f_ε is symmetrical about 0. Moreover, f_U and f_ε are everywhere twice continuously differentiable with uniformly bounded derivatives, and h_ε is strictly positive everywhere.

A2. The distribution of X has bounded support.

A3. b_n is a $n^{1/2}$-consistent estimator of β. That is, $n^{1/2}(b_n - \beta) = O_p(1)$ as $n \to \infty$.

A4. Let $A_{n\varepsilon} = (\log n)/[n^{1/2}h_\varepsilon(1/\lambda_{n\varepsilon})^2]$ and $B_{n\varepsilon} = 1/[n^{1/2}\lambda_n h_\varepsilon(1/\lambda_{n\varepsilon})^2]$. Define A_{nU} and B_{nU} by replacing $\lambda_{n\varepsilon}$ with λ_{nU} in $A_{n\varepsilon}$ and $B_{n\varepsilon}$. As $n \to \infty$: $\lambda_{n\varepsilon} \to 0$, $\lambda_{nU} \to 0$, $B_{n\varepsilon}/\lambda_{n\varepsilon} \to 0$, $B_{nU}/\lambda_{nU} \to 0$, $A_{n\varepsilon}/\lambda_{n\varepsilon} = O(1)$, and $A_{nU}/\lambda_{nU} = O(1)$.

Assumption A1 insures, among other things, that f_ε and f_U are identified. Examples of distributions with strictly positive characteristic functions are the normal, the Cauchy and scale mixtures of these. A2 can always be satisfied by dropping observations with "large" values of X. At the expense of greater complexity in the proofs, A2 can be weakened to permit distributions of X with unbounded support. A3 insures that random sampling errors in the estimate of β are asymptotically negligible. A4 restricts the rate at which $\lambda_{n\varepsilon} \to 0$ and $\lambda_{nU} \to 0$ as $n \to \infty$. The λ_n's must converge more slowly if the tails of h_ε are thin than if they are thick. To illustrate, if ε has the normal (Cauchy) distribution, A4 is satisfied if $\lambda_n = c(\log n)^{-1/2}$ ($\lambda_n = c(\log n)^{-1}$) for any sufficiently small $c > 0$.

The following theorem establishes uniform consistency of $\hat{f}_\varepsilon$ and $\hat{f}_U$. It also gives their uniform rates of convergence in probability.

<u>Theorem 1</u>: *Let* g *be a bounded, real characteristic function with support* [-1,1]. *If* g *is twice differentiable in a neighborhood of* 0 *and A1-A4 hold, then as* $n \to \infty$

$$\sup_z |\hat{f}_\varepsilon(z) - f_\varepsilon(z)| = O_p(\lambda_{n\varepsilon}^2) + O_p(B_{n\varepsilon}/\lambda_{n\varepsilon}) + o_p[(A_{n\varepsilon}/\lambda_{n\varepsilon})^{1/2}] \quad (3.9)$$

and

$$\sup_u |\hat{f}_U(u) - f_U(u)| = O_p(\lambda_{nU}^2) + O_p(B_{nU}/\lambda_{nU}) + o_p[(A_{nU}/\lambda_{nU})^{1/2}]. \quad (3.10)$$

z

Theorem 1 implies that the rates of convergence in probability of $_{\varepsilon}$ and $_{U}$ are controlled by the rates at which $\lambda_{n\varepsilon}$ and λ_{nU} converge to 0. The latter rates are controlled by A4. In general, faster rates of convergence of $_{\varepsilon}$ and $_{U}$ are possible when h_{ε} is thick-tailed than when it is thin-tailed. To provide some insight into the resulting rates of convergence, we investigate them in detail for the special case of normally distributed ε.

If $\varepsilon \sim N(0,\sigma_{\varepsilon}^{2})$, then $h_{\varepsilon}(\tau) = \exp(-0.5\sigma_{\varepsilon}^{2}\tau^{2})$. Given this h_{ε}, it is not difficult to show that A4 cannot be satisfied if $\lambda_{n\varepsilon}$ and λ_{nU} converge to 0 faster than $(\log n)^{-1/2}$. Thus, if ε and U are normally distributed, the fastest possible uniform rate of convergence in probability of $_{\varepsilon}$ and $_{U}$ is $(\log n)^{-1}$. In addition, it may be shown that the fastest possible pointwise rate of convergence in probability is $(\log n)^{-1}$ and the fastest possible rate of convergence in probability of the integrated squared errors of $_{\varepsilon}$ and $_{U}$ is $(\log n)^{-2}$.

Although these rates are very slow, they cannot be increased under assumption A1 and certain additional mild regularity conditions. Slow convergence is intrinsic to the deconvolution problem and is the price that must be paid for lack of *a priori* knowledge of the densities of ε and U. To see why, first consider f_U. It follows from Theorem 1 of Carroll and Hall (1988) that if $\varepsilon \sim N(0,\sigma_{\varepsilon}^{2})$ and f_U is assumed to have k bounded derivatives, then the pointwise rate of convergence in probability of an estimator of f_U cannot exceed $(\log n)^{-k/2}$. Here, we assume the existence of two derivatives of f_U, so k = 2 and the upper bound on the pointwise rate of convergence is $(\log n)^{-1}$. This is the rate achieved by $_{U}$.

Now consider f_ε. We will prove that under regularity conditions, $(\log n)^{-2}$ is the fastest possible rate of convergence in probability of the integrated squared error of an estimator of f_ε. This is the rate achieved by $_\varepsilon$. The following notation will be used in addition to that previously defined:

$_\varepsilon$ = Any estimator of f_ε that is symmetrical about 0 and pointwise consistent.

$_\eta$ = The estimator of f_η whose characteristic function evaluated at the point τ is $_\varepsilon(\tau)^2 g(\lambda_{n\varepsilon}\tau)^2$, where $_\varepsilon$ is the characteristic function of $_\varepsilon$.

$_\eta'$ = The derivative of $_\eta$.

f_η^* = The density whose characteristic function evaluated at the point τ is $h_\varepsilon(\tau)^2 g(\lambda_{n\varepsilon}\tau)^2$.

The following theorem, which is a modified version of Theorem 3.1 of Stefanski and Carroll (1990), gives the required result.

Theorem 2: *Assume that A1 holds and that* $\varepsilon \sim N(0,\sigma_\varepsilon^2)$. *Suppose there are sequences of constants* $\{a_n\}$, $\{c_{1n}\}$, $\{c_{2n}\}$, *and* $\{c_{3n}\}$ *such that as* $n \to \infty$: $a_n \to 0$; c_{1n}, c_{2n}, *and* c_{3n} *have non-zero finite limits; and*

(a) $\int_{-\infty}^{\infty} \mathrm{Var}[\ _\eta(z)]dz = c_{1n}/(na_n)$

(b) $\int_{-\infty}^{\infty} \mathrm{Var}[f_\eta'(z)]dz = c_{2n}/(na_n^3)$

(c) $\int_{-\infty}^{\infty} [E\hat f_\eta(z) - f_\eta^*(z)]^2 dz \le c_n a_n^4$.

Then

$$ISE = \int_{-\infty}^{\infty} [\hat f_\varepsilon(z) - f_\varepsilon(z)]^2 dz$$

converges in probability to 0 *at a rate that does not exceed* $(\log n)^{-2}$.

To understand the significance of this theorem, note that as a consequence of (2.2), η is observable up to random sampling error in b_n, whereas ε in (2.1) is not observable. Thus, any estimator of f_ε must be derived explicitly or implicitly from an estimator of f_η. Theorem 2 shows that when the estimator of f_η satisfies certain conditions, the corresponding estimator of f_ε has an integrated squared error whose rate of convergence in probability does not exceed $(\log n)^{-2}$. The estimator $\hat f_\varepsilon$ in (3.7) satisfies the conditions of Theorem 2 with $a_n = \lambda_{n\varepsilon}$ and $\hat f_\eta$ the probability density whose characteristic function evaluated at the point τ is $|\hat\phi_\eta(\tau)|\, g(\lambda_{n\varepsilon}\tau)^2$. Up to asymptotically negligible terms introduced by the modulus operator, this $\hat f_\eta$ is a nonparametric kernel estimator. Specifically,

$$\hat f_\eta(z) = [n(T-1)\lambda_{n\varepsilon}]^{-1} \sum_{j=1}^{n} \sum_{t=2}^{T} K_g\left(\frac{z - \hat\eta_{jt}}{\lambda_{n\varepsilon}}\right) + o_p(\lambda_{n\varepsilon}^2),$$

where K_g is the probability density whose characteristic function is g^2. More generally, a_n is the bandwidth if (apart from asymptotically negligible terms) $\hat f_\eta$ is a kernel estimator with a non-negative kernel, and $a_n = k_n/n$ if $\hat f_\eta$ is a k_n-nearest neighbor estimator.

The assumption that $_{\varepsilon}$ is symmetrical about 0 does not restrict the generality of the conclusion of theorem 2. This is because the rate of convergence in probability of the integrated squared error of $_{\varepsilon}$ cannot exceed the rate of convergence of the integrated squared error of the symmetrized estimator $[_{\varepsilon}(z) + _{\varepsilon}(-z)]/2$.

The rates of convergence of $_{\varepsilon}$ and $_{U}$ can be increased from those discussed here by assuming that f_{ε} and f_U have more than two derivatives. If f_{ε} and f_U are assumed to have k > 2 derivatives, faster rates of convergence can be achieved by replacing g with a bounded, real function whose support is [-1,1] and that satisfies (0) = 1 and $d^r g(z)/dz^r\big|_{z=0} = 0$ for r = 1,...,k. If this is done for the case $\varepsilon \sim N(0,\sigma_{\varepsilon}^2)$, $U \sim N(0,\sigma_U^2)$, the pointwise and uniform rates of convergence in probability of $_{\varepsilon}$ and $_{U}$ in (3.7) and (3.8) are $(\log n)^{-k/2}$. The integrated squared errors of $_{\varepsilon}$ and $_{U}$ converge in probability at the rate $(\log n)^{-k}$. Arguments similar to those made above show that under mild regularity conditions, these rates cannot be improved without assuming the existence of more than k derivatives of f_{ε} and f_U.

c. First Passage Times

In Section 5, we shall be concerned with estimating the probability distributions of certain first passage times of individuals' earnings. In this section, we define the distributions of interest and explain how they are estimated.

The first passage time for individual i is the smallest t for which Y_{it} exceeds a specified threshold, say y*. In this paper, we will be concerned with the first passage time conditional on the initial value of Y for individual i, Y_{i1}, and the covariates X_{it}. Given an integer $\theta > 1$, let $P(\theta | y_1, y^*, x)$ denote the probability that the first passage time for threshold y^* and individual i exceeds θ conditional on $Y_{i1} = y_1$ and $X_{it} = x_t$ $(t = 1,...,\theta)$. Then

$$P(\theta | y_1, y^*, x) = \Pr(Y_{i2} \le y^*, ..., Y_{i\theta} \le y^* | Y_{i1} = y_1).$$

If U is independent of X, some algebra shows that

$$P(\theta | y_1, y^*, x) = \qquad (3.11)$$

$$f_\nu(y_1 - \beta' x_{i1})^{-1} \int_{-\infty}^{\infty} \left[f_\varepsilon(y_1 - \beta' x_1 - u) \prod_{j=2}^{\theta} F_\varepsilon(y^* - \beta' x_j - u) \right] f_U(u) du,$$

where f_ν is the probability density of the $\nu = U + \varepsilon$, and F_ε is the cumulative distribution function of ε. $P(\theta | y_1, y^*, x)$ can be estimated consistently by

$$\hat{P}(\theta | y_1, y^*, x) = \qquad (3.12)$$

$$\hat{f}_\nu(y_1 - b_n' x_{i1})^{-1} \int_{-\rho_n}^{\rho_n} \left[\hat{f}_\varepsilon(y_1 - b_n' x_1 - u) \prod_{j=2}^{\theta} \hat{F}_\varepsilon(y^* - b_n' x_j - u) \right] \hat{f}_U(u) du,$$

where b_n is a $n^{1/2}$-consistent estimator of β; $\hat{f}_\varepsilon$ and $\hat{f}_U$ are given by (3.7) and (3.8); $\hat{f}_\nu$ is a kernel estimator of the density of ν; $\hat{F}_\varepsilon$ is the estimator of the cumulative distribution function of ε that is described below; and $\{\rho_n\}$ is a sequence of positive constants that satisfies $\rho_n \to \infty$, $\rho_n \sup_z |\hat{f}_\varepsilon(z) - f_\varepsilon(z)| \to 0$, and $\rho_n \sup_z |\hat{f}_U(z) - f_U(z)| \to 0$ as $n \to \infty$.

To estimate F_ε, observe that since the distribution of ε is symmetrical around 0 by assumption

$$F_\varepsilon(z) = 0.5 + 0.5[F_\varepsilon(z) - F_\varepsilon(-z)]$$

$$= 0.5 + (2/\pi)\int_0^\infty \tau^{-1} \sin(z\tau) h_\varepsilon(\tau) d\tau.$$

Using arguments similar to those used to prove theorem 1, it can be shown that $F_\varepsilon(z)$ is estimated consistently uniformly over z by

$$_\varepsilon(z) = 0.5 + (2/\pi)\int_0^\infty \tau^{-1} \sin_\eta(z\tau) |{}_{n\varepsilon}(\tau)|^{1/2} g(\lambda \tau) d\tau.$$

This is the estimator of F_ε that we use in (3.12).

d. A Small-Sample Correction

The results of the Monte Carlo experiments described in Section 4 show that $_\varepsilon$, $_U$, and $(\theta | y_1, y^*, x)$ can be seriously biased in samples of practical size. We now describe small-sample corrections for $_\varepsilon$ and $_U$ that remove part of these biases.[3] The arguments leading to the corrections are identical for $_\varepsilon$ and $_U$, so we discuss only $_\varepsilon$.

To derive the correction to $_\varepsilon$, observe that by (3.7),

$$_\varepsilon(z) - f_\varepsilon(z) = \Delta_{n1}(z) + \Delta_{n2}(z),$$

where

$$\infty$$

$$\Delta_{n1}(z) = (1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}\left[|\hat{h}_\eta(\tau)|^{1/2} - |h_\varepsilon(\tau)| \right] g(\lambda_{n\varepsilon}\tau)d\tau$$

and

$$\Delta_{n2}(z) = (1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}[g(\lambda_{n\varepsilon}\tau) - 1]h_\varepsilon(\tau)d\tau.$$

Note that Δ_{n2} is nonstochastic. In a finite sample, neither $E\Delta_{n1}(z)$ nor $\Delta_{n2}(z)$ is zero in general, so $_\varepsilon(z)$ is biased. $E\Delta_{n1}(z)$ is the component of bias caused by estimation of h_ε, and $\Delta_{n2}(z)$ is the component of bias caused by smoothing the empirical distribution of η. The small-sample correction described here removes the second component of bias through order $\lambda_{n\varepsilon}^2$.

To derive the correction, recall that $g(\lambda_{n\varepsilon}\tau)h_\varepsilon(\tau)$is the characteristic function of the random variable $\varepsilon + \lambda_n\zeta$, where ζ is the random variable whose characteristic function is g. Therefore, $\Delta_{n2}(z)$ is the difference between the probability density of $\varepsilon + \lambda_n\zeta$ and the probability density of ε. Let ψ denote the density of ζ. Then

$$\Delta_{n2}(z) = \int_{-\infty}^{\infty} f_\varepsilon(z - \lambda_{n\varepsilon}\tau)\psi(\tau)d\tau - f_\varepsilon(z).$$

A Taylor series expansion of $f_\varepsilon(z - \lambda_{n\varepsilon}\tau)$ about $\lambda_{n\varepsilon} = 0$ and application of the dominated convergence theorem yield

$$\Delta_{n2}(z) = (1/2)\lambda_{n\varepsilon}^2 f_\varepsilon''(z)\sigma_\zeta^2 + o(\lambda_{n\varepsilon}^2), \qquad (3.13)$$

where f_ε'' denotes the second derivative of f_ε and σ_ζ^2 is the variance of ζ. The first term on the right-hand side of (3.13) is the smoothing bias in $_\varepsilon(z)$ through $O(\lambda_{n\varepsilon}^2)$. This bias can be removed by estimating $f_\varepsilon(z)$ with

$$_\varepsilon(z) = {}_\varepsilon(z) - (1/2)\lambda_{n\varepsilon}^2 \, {}_\varepsilon''(z)\sigma_\zeta^2 ,$$

where $_\varepsilon''(z)$ is a consistent estimator of $f_\varepsilon''(z)$.

A consistent estimator of $f_\varepsilon''(z)$ can be obtained by differentiating the right-hand side of (3.7) with respect to z and replacing $\lambda_{n\varepsilon}$ with a bandwidth $\gamma_{n\varepsilon}$ that converges to 0 at a sufficiently slow rate. This result is given formally in the following theorem.

<u>Theorem 3</u>: *Let A1-A4 hold. Assume that* f_ε'' *is Lipschitz continuous. Let* $\{\gamma_{n\varepsilon}\}$ *be a sequence satisfying* $\gamma_{n\varepsilon} \to 0$, $B_{n\varepsilon}/\gamma_{n\varepsilon}^3 \to 0$, *and* $A_{n\varepsilon}/\gamma_{n\varepsilon} = O(1)$ *as* $n \to \infty$. *Define*

$$_\varepsilon''(z) = -(1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}\tau^2 \,\big|\, {}_\eta(\tau)\big|^{1/2} g(\gamma_{n\varepsilon}\tau)d\tau.$$

Then

$$\operatorname*{plim}_{n\to\infty} \sup_z \big| {}_\varepsilon''(z) - f_\varepsilon''(z)\big| = 0.$$

A similar procedure can be used to remove smoothing bias from $_U$ through $O(\lambda_{nU}^2)$. The resulting estimator of f_U is

$$_U(z) = {}_U(z) - (1/2)\lambda_{nU}^2 \, {}_U''(z)\sigma_\zeta^2,$$

where

$$''_U(z) = -(1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}\tau^2 {}_\nu(\tau)\,|{}_\eta(\tau)|^{-1/2}\, g(\gamma_{nU}\tau)\,d\tau.$$

A proof identical to that of theorem 3 shows that $_U''$ is a uniformly consistent estimator of f_U if f_U'' is Lipschitz continuous, $\gamma_{nU} \to 0$, $B_{nU}/\gamma_{nU}^3 \to 0$, and $A_{nU}/\gamma_{nU}^3 = O(1)$ as $n \to \infty$. Section 4 presents Monte Carlo evidence on the extent of the bias reduction obtained by using $_\varepsilon$ and $_U$ instead of $_\varepsilon$ and $_U$.

4. Monte Carlo Experiments

This section presents the results of Monte Carlo experiments aimed at investigating whether $_\varepsilon$, $_\varepsilon$, $_U$, $_U$, and $(\theta|y_1,y^*,x)$ can provide useful information about f_ε, f_U, and $P(\theta|y_1,y^*,x)$ in samples of moderate size.

Data for the experiments were generated by simulation from the model

$$Y_{it} = {}_i\ U_{it} + \varepsilon\ ; \quad i = 1,...,1000; \quad t = 1,2.$$

Thus, the simulated data correspond to a panel of length T = 2 composed of n = 1000 individuals. The distribution of U is N(0,1). The distribution of ε is N(0,1) in one set of experiments and the mixture 0.9N(0,1) + 0.1N(0,16) in another set. The mixture distribution has tails that are thicker than the normal and, in this respect, is similar to the distribution of ε in the model described in Section 2. In both sets of experiments the fastest possible rate of convergence in probability of estimators of f_ε and f_U is $(\log n)^{-1}$, so the experiments address situations in which the estimators converge slowly.

The smoothing function, g, used in the experiments is the fourfold convolution of the uniform density with itself. This is the characteristic function of the density $c[(\sin x)/x]^4$, where c is a normalization constant. The density f_v was estimated using a kernel estimator with the standard normal density as the kernel. In estimating $P(\theta | y_1, y^*,)$, we set $y_1 = -1$, $y^* = 1$, and $\theta = 3, 5, 7, 9$, and 11.[4] Since we have no formal theory of how the bandwidths $\lambda_{n\varepsilon}$ and λ_{nU} and ρ_n should be selected in finite samples, we used informal graphical methods. We found through experimentation that apart from small wiggles, $_{\varepsilon}(\tau) = 0$ for $\tau > 6$, $_{U}(\tau) = 0$ for $\tau \geq 5$, and the integrand in the numerator of $(\theta | y_1 = -1, y^* = 1)$ is zero for $\tau \geq 4$. Accordingly, we set $\lambda_{n\varepsilon} = 0.18$, $\lambda_{nU} = 0.20$, and $\rho_n = 4$. Experimentation with other bandwidths showed that use of moderately larger values of $\lambda_{n\varepsilon}$ and λ_{nU} gave results similar to those reported here but that use of moderately smaller or very much larger values produced badly biased results. We set $\gamma_{n\varepsilon} = \lambda_{n\varepsilon}^{1/3}$ and $\gamma_{nU} = \lambda_{nU}^{1/3}$. The experiments were carried out with a program written in GAUSS using GAUSS random number generators. There were 100 replications per experiment.

Figure 3 shows graphs of the functions $_{\varepsilon}$, $_{\varepsilon}$, $_{U}$, and $_{U}$ obtained from the first 19 replications of the experiment with $\varepsilon \sim N(0,1)$. The results of the other replications are similar but are not shown because doing so would make the plots so dense that individual estimates could not be distinguished. Figure 4 shows similar graphs for the experiment with $\varepsilon \sim 0.9N(0,1) + 0.1N(0,16)$. In each figure, the solid lines are estimates and the lines with dots are the true densities It can be seen that most of the estimates are qualitatively similar to the corresponding true densities, although some estimates of f_U are notably wiggly and the estimates tend to be too flat. The main difference between the estimates with and without bias correction is that the corrected estimates are shifted upward relative to the uncorrected ones. As a result, the corrected estimates fit the centers of the true distributions better than the uncorrected estimates do.

The results of estimating $P(\theta | y_1=-1, y^*=1)$ are shown in Table 2. The first 5 columns show the true values of $P(\theta | y_1=-1, y^*=1)$ and the means of the estimates $(\theta | y_1=1, y^*=-1)$. The estimates based on $_{\varepsilon}$ and $_{U}$ are biased downward by 12-20%, depending on the distribution of ε and the value of θ. The downward bias increases with increasing θ. When the bias-corrected density estimates $_{\varepsilon}$ and $_{U}$ are used to compute $(\theta | y_1=-1, y^*=1)$, the downward bias is reduced to 1-13%. Thus, the bias correction removes 35% to virtually all of the bias of $(\theta | y_1=-1, y^*=1)$, depending on the distribution of ε and the value of θ.

The last column of Table 2 shows the means of the estimates of $P(\theta | y_1=-1, y^*=1)$ that are obtained by assuming that ε is normally distributed when, in fact, it has the mixture distribution 0.9N(0,1) + 0.1N(0,16). These estimates were obtained from (3.11) by assuming that ν, ε, and U are normally distributed with means of zero and variances estimated from the simulated samples. Specifically, σ_{ε}^2 is estimated by the sample variance of $(Y_2 - Y_1)$, σ_{ν}^2 is estimated by the sample variance of the pooled Y's, and σ_U^2 is estimated by $_{\nu}^{2}$ - $_{\varepsilon}^{2}$, where hats denote estimated values. It can be seen that the erroneous assumption of normality of ε produces estimates that are biased downward by 12-36%, whereas the downward bias is only 1-13% when the bias-corrected nonparametric density estimators are used. The Monte Carlo estimates of the variances of $(\theta | y_1\text{-}1, y^*=1)$ based on

$_\varepsilon$ and $_U$ are all below 0.0016, so this estimator has a smaller mean square error than does the erroneous parametric estimator as well as a smaller bias.

There is a simple intuitive explanation for the severe downward bias of the parametric estimator of $P(\theta | y_1=-1,y^*=1)$. The mixture distribution used in the experiments has less probability in its tails than does a normal distribution with the same variance. Therefore, the normal distribution has a higher probability of a transition from one tail to another than does the mixture distribution. Since $P(\theta | y_1=-1,y^*=1)$ is the probability that a transition between tails does not occur, the probabilities obtained from the normal distribution are too low.

Of course, one cannot draw general conclusions from a small set of Monte Carlo experiments. However, the evidence presented here indicates that the bias correction described in Section 3d is useful and that the bias-corrected nonparametric density estimators are capable of providing useful information about the densities of ε and U in samples of moderate size. The evidence also indicates that the nonparametric estimators can yield estimates of first-passage probabilities that are considerably more accurate than ones obtained from a misspecified parametric model.

5. An Application to Estimation of Earnings Mobility

In this section, we illustrate the use of the nonparametric estimators of f_ε and f_U in an application that consists of estimating indicators of the earnings-mobility of individuals. We consider an individual whose earnings are 100(1 - α) percent of median earnings of individuals with the same age, education and marital status, where α = 0.10, and 0.20. We estimate the probability that the individual's earnings never exceed 100(1 + α) percent of the median in any of the subsequent 2, 4, 6, 8, or 10 years. This corresponds to estimating $P(\theta|y_1,y^*,x)$, where θ = 3, 5, 7, 9, or 11. The variable x specifies age, education and marital status; and y_1 and y^*, respectively, are 100(1 - α) percent and 100(1 + α) percent of median earnings conditional on x. We compare the estimates of $P(\theta|y_1,y^*,x)$ obtained using nonparametric estimates of f_ε and f_U with estimates obtained under the assumption that ε and U are normally distributed.

The estimates are based on model (2.1) with Y_{jt} equal to the logarithm of real annual earnings of individual j in year t. The explanatory variables are those listed in Table 1 plus an intercept. The data are described in Section 2b. We assume that U and X are independent and estimate β by generalized least squares (see, e.g., Hsiao, 1986, pp. 34-38). Table 3 shows the estimates of β, σ_ε^2, and σ_U^2. We used the bias-corrected nonparametric estimators of f_ε and f_U and the same smoothing function, g, as was used in the Monte Carlo experiments. The bandwidths $\lambda_{n\varepsilon}$, λ_{nU}, and ρ_n were obtained using the informal graphical procedure described in Section 4, and $\gamma_{n\varepsilon}$ $(\gamma_{nU}) = \lambda_{n\varepsilon}^{1/3}$ $(\lambda_{nU}^{1/3})$.

As is discussed in Section 2, there is strong evidence that the distribution of ε is not normal. A plot analogous to Figure 1 cannot be made for U because realizations of U are not observed. However, an informal graphical test of normality of U can be obtained by plotting $\log[_U(\tau)]$ against $-\tau^2$. If U is normally distributed, $h_U(\tau) = \exp(-0.5\sigma_U^2\tau^2)$, so the plot will consist of scatter around a straight line. Figure 5 shows the plot (solid line) and a straight line (dashes). The plot suggests that any departure of the distribution of U from normality is mild. In the data used here, only ε has a distribution that is distinctly non-normal.

Table 4 shows the estimates of $P(\theta | y_1, y^*, x)$. Depending on the values of α and θ, the estimates obtained from the nonparametric estimators of f_ε and f_U are 15-100 percent higher than those obtained by assuming that ε and U are normally distributed. Thus, the assumption that ε is normally distributed leads to substantial overestimation of the probability that an individual with low earnings will become a high earner in future. This finding is consistent with the results of the Monte Carlo experiments, which showed that the probabilities of transitions from low to high values of Y are overestimated if ε has a thick-tailed distribution but is assumed to be normally distributed.

6. Conclusions

This paper has shown how to carry out nonparametric estimation of the densities of the error components in a regression model for panel data. The usefulness of the nonparametric estimators has been illustrated through Monte Carlo experiments and an application to estimating the earnings-mobility of individuals. The estimates of earnings mobility obtained by using the nonparametric estimators are considerably lower than those obtained under the assumption that the error components of the earnings model are normally distributed. The nonparametric estimators converge slowly, but slow convergence is intrinsic to the deconvolution problem that must be solved to estimate the densities of the error components. Alternative estimation approaches, such as attempting to find a parametric model that fits the data, cannot produce faster-converging estimators. In further research it would be useful to find systematic methods for selecting the bandwidths needed to implement the nonparametric density estimators.

APPENDIX: PROOFS OF THEOREMS

A1. Consistency of the Density Estimators

Throughout this appendix, λ_n refers to either $\lambda_{n\varepsilon}$ or λ_{nU}. Lemmas 1-4 below provide results that are used in proving Theorem 1.

Lemma 1: *Under assumptions A1-A3 and as* $n \to \infty$,

$$\sup_{|\tau| \le 1/\lambda_n} |\ _\eta(\tau) - h_\eta(\tau)| = o_p[(\log n)^{1/2}/n_p^{1/2}] + O_n[1/(n\ \ \lambda)]. \quad (A.1)$$

Proof: By the mean value theorem of differential calculus

$$_\eta(\tau) = [n(T-1)] \sum_{j=1}^{n} \sum_{t=2}^{T} \exp(i\tau\eta_{jt})\exp[i\tau(\ _{jt} - \eta_{jt})]$$

$$= [n(T-1)] \sum_{j=1}^{n} \sum_{t=2}^{T} \exp(i\tau\eta_{jt})\{1 + O[\tau(\ _{jt} - \eta_{jt})]\}.$$

By assumptions A2 and A3

$$\sup_{j,t} |\ _{jt} - \eta_{jt}|_p = O(n^{-1/2}).$$

Therefore,

$$_\eta(\tau) = [n(T-1)] \sum_{j=1}^{n} \sum_{t=2}^{T} \exp(i\tau\eta_{jt})_p^{1/2} + O_n[1/(n\ \ \lambda)] \qquad (A.2)$$

uniformly over $|\tau| \le 1/\lambda_n$. The class of functions $\exp(i\tau\eta)$, considered as functions of η indexed by τ for $-\infty < \tau < \infty$, satisfies the assumptions of theorem 2.37 of Pollard (1984). By this theorem

$$\sup_{\tau} \left| [n(T-1)]^{-1} \sum_{j=1}^{n} \sum_{t=2}^{T} [\exp(i\tau\eta_{jt}) - E\exp(i\tau\eta)] \right| = o[(\log n)/n^{1/2}] \quad (A.3)$$

almost surely. Since $E\exp(i\tau\eta) = h_\eta(\tau)$, the lemma follows by combining (A.2) and (A.3). Q.E.D.

Lemma 2: *Under assumptions A1-A4 and as* $n \to \infty$,

$$\sup_{|\tau| \le 1/\lambda_n} |\hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau)| = o_p(A_n) + O_p(B_n)$$

Proof: Note that $h_\varepsilon = h_\eta^{1/2}$. Therefore, by assumption A4 and lemma 1

$$|\hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau)| = h_\varepsilon(\tau)\left\{\left|1 + \frac{\hat{h}_\eta(\tau) - h_\eta(\tau)}{h_\eta(\tau)}\right|^{1/2} - 1\right\}$$

$$= h_\varepsilon(\tau)\left|\frac{\hat{h}_\eta(\tau) - h_\eta(\tau)}{h_\eta(\tau)}\right|[1/2 + o_p(1)]. \quad (A.4)$$

The lemma follows by substituting (A.1) into (A.4). Q.E.D.

Lemma 3: *Under assumptions A1-A3 and as* $n \to \infty$,

$$\sup_{|\tau| \le 1/\lambda_n} |\nu(\tau) - h_\nu(\tau)| = o_p[(\log n)/n]^{1/2} + O_p[1/(n^{1/2}\lambda_n)].$$

Proof: Identical to the proof of lemma 1. Q.E.D.

Lemma 4: *Under assumptions A1-A4 and as* $n \to \infty$

$$\varepsilon(\tau)/h_\varepsilon(\tau) - 1 = o_p(A_n) + O_p(B_n).$$

uniformly over $|\tau| \le 1/\lambda_n$.

Proof: This is an immediate consequence of (A.1) and (A.4). Q.E.D.

Lemma 5: *Under assumptions A1-A4 and as* $n \to \infty$,

$$\sup_{|\tau| \le 1/\lambda_n} |U(\tau) - h_U(\tau)| = o_p(A_n) + O_p(B_n)$$

Proof: Some algebra yields

$$U(\tau) - h_U(\tau) = \frac{\{[\nu(\tau) - h_\nu(\tau)]/h_\varepsilon(\tau)\} - h_U(\tau)[\varepsilon(\tau)/h_\varepsilon(\tau) - 1]}{1 + [\varepsilon(\tau)/h_\varepsilon(\tau) - 1]} \quad (A.5)$$

The lemma follows by substituting the results of lemmas 3 and 4 into the right-hand side of (A.5).

Proof of Theorem 1: To prove (3.9), write $\varepsilon(z)$ in the form

$$\varepsilon(z) = I_{n1}(z) + I_{n2}(z),$$

where

$$\infty$$

$$I_{n1}(z) = (2\pi)^{-1}\int_{-\infty}^{\infty} [\phi_\varepsilon(\tau) - h_\varepsilon(\tau)]g(\lambda_n \tau)\exp(-iz\tau)d\tau$$

and

$$I_{n2}(z) = (2\pi)^{-1}\int_{-\infty}^{\infty} \phi_\varepsilon h_n(\tau)g(\lambda_n \tau)\exp(-iz\tau)d\tau.$$

Define

$$R_{n\varepsilon} = \sup_{|\tau| \le 1/\lambda_n} |\phi_\varepsilon(\tau) - h_\varepsilon(\tau)|.$$

Then

$$|I_{n1}(z)| \le (2\pi)^{-1}\int_{-\infty}^{\infty} R_{n\varepsilon} |g(\lambda_n \tau)|d\tau$$

$$= (2\pi)^{-1}(R_{n\varepsilon}/\lambda_n)\int_{-1}^{1} |g(\tau)|d\tau$$

$$= o_p(A_n/\lambda_n) + O_p(B_n/\lambda_n) \qquad \text{(A.6)}$$

as $n \to \infty$ uniformly over z by lemma 2, A4 and the boundedness of g. Now consider $I_{n2}(z)$. Let K denote the probability density whose characteristic function is g. Observe that I_{n2} is the density of $\varepsilon + \lambda_n\zeta$, so

$$I_{n2}(z) = \int_{-\infty}^{\infty} f_\varepsilon(z - \lambda_n w)K(w)dw.$$

A Taylor series expansion of $f_\varepsilon(z - \lambda_n w)$ about $\lambda_n = 0$ together with symmetry of K yield

$$I_{n2}(z) = f_\varepsilon(z) + (1/2)\lambda_n^2 \int_{-\infty}^{\infty} w^2 f_\varepsilon''(\zeta_n)K(w)dw, \qquad (A.7)$$

where ζ_n is between $z - \lambda_n w$ and z. Since f_ε'' is uniformly bounded by A1 and $\int w^2 K(w)dw < \infty$ (because g is twice differentiable), $I_{n2}(z) = f_\varepsilon(z) + O(\lambda_n^2)$ uniformly over z. Equation (3.9) follows by combining this result with (A.6).

To prove (3.10), write $_U(z)$ in the form

$$_U(z) = J_{n1}(z) + J_{n2}(z),$$

where

$$J_{n1}(z) = (2\pi)^{-1} \int_{-\infty}^{\infty} [\;_U(\tau) - h_U(\tau)]g(\lambda_n \tau)\exp(-iz\tau)d\tau$$

and

$$J_{n2}(z) = (2\pi)^{-1} \int_{-\infty}^{\infty} h_U(\tau)g(\lambda_n \tau)\exp(-iz\tau)d\tau.$$

Define

$$R_{nU} = \sup_{|\tau| \leq 1/\lambda_n} |\;_U(\tau) - h_U(\tau)|.$$

Then

$$|J_{n1}(z)| \leq (2\pi)^{-1} R_{nU} \int_{-\infty}^{\infty} |g(\lambda_n \tau)| d\tau$$

$-\infty$

$$= (2\pi)^{-1}(R_{nU}/\lambda_n)\int_{-1}^{1} |g(\tau)| d\tau$$

$$= o_p(A_n/\lambda_n) + O_p(B_n/\lambda_n) \qquad \text{(A.8)}$$

as $n \to \infty$ uniformly over z, where the last line follows from lemma 5 and the boundedness of g. Finally, by arguments identical to those used in obtaining (A.7)

$$J_{n2}(z) = f_U(z) + (1/2)\lambda_n^2 \int_{-\infty}^{\infty} w^2 f_U''(\zeta_n)K(w)dw,$$

where ζ_n is between z - λ_nw and z. Since f_U'' is uniformly bounded by A1 and $\int w^2K(w)dw < \infty$, $I_{n2}(z) = f_U(z) + O(\lambda_n^2)$ uniformly over z. Equation (3.11) follows by combining this result with (A.8). Q.E.D.

A.2 Proof of Theorem 2

Set $\eta = \varepsilon^2$. Then a straightforward calculation gives

$$\text{ISE} = (2\pi)^{-1}\int_{-\infty}^{\infty} |_{\eta} [(\tau)_{\eta}^{1/2} - h(\tau)^{1/2}]^2 d\tau$$

$$\geq (2\pi)^{-1}\int_{-\infty}^{\infty} |_{\eta} [(\tau)^{1/2} - h_{\eta}(\tau)^{1/2}]^2 g(\lambda_n\tau)^4 d\tau \qquad \text{(A.9)}$$

for any positive sequence $\{\lambda_n\}$. Choose $\{\lambda_n\}$ such that as $n \to \infty$,

p

$$\sup_{|\tau| \le 1/\lambda_n} |{}_\eta(\tau) - h_\eta(\tau)|/h_\eta(\tau) \to 0$$

Then

$$[{}_\eta(\tau)^{1/2} - h_\eta(\tau)^{1/2}]^2 = (1/4)\frac{[{}_\eta(\tau) - h_\eta(\tau)]^2}{h_\eta(\tau)}[1 + o_p(1)]$$

uniformly over $|\tau| \le 1/\lambda_n$. It follows that the rate of convergence in probability of the right-hand side of (A.9) is the same as the rate of convergence of

$$J_n \equiv \int_{-\infty}^{\infty} \frac{[{}_\eta(\tau) - h_\eta(\tau)]^2}{h_\eta(\tau)} g(\lambda_n\tau)^4 \, d\tau$$

$$= \int_{-\infty}^{\infty} [\exp(\sigma_\varepsilon^2\tau^2)][{}_\eta(\tau) - h_\eta(\tau)]^2 g(\lambda_n\tau)^4 \, d\tau$$

if $\varepsilon \sim N(0,\sigma_\varepsilon^2)$. Let ${}_\eta^*(\tau) = {}_\eta(\tau)g(\lambda_n\tau)^2$ and $h_\eta^*(\tau) = h_\eta(\tau)g(\lambda_n\tau)^2$. Then $E(J_n) = J_{n1} + J_{n2}$, where

$$J_{n1} = \int_{-\infty}^{\infty} [\exp(\sigma_\varepsilon^2\tau^2)]E[{}_\eta^*(\tau) - E{}_\eta^*(\tau)]^2 \, d\tau$$

and

$$J_{n2} = \int_{-\infty}^{\infty} [\exp(\sigma_\varepsilon^2\tau^2)][E{}_\eta(\tau) - h_\eta^*(\tau)]^2 \, d\tau.$$

By Jensen's inequality

$$J_{n1} \geq \int_{-\infty}^{\infty} E[\eta^*(\tau) - E\eta^*(\tau)]^2 \, d\tau \bullet \exp\left\{ \sigma_\varepsilon^2 \frac{\int_{-\infty}^{\infty} \tau^2 E[\eta^*(\tau) - E\eta^*(\tau)]^2 \, d\tau}{\int_{-\infty}^{\infty} E[\eta^*(\tau) - E\eta^*(\tau)]^2 \, d\tau} \right\}$$

$$= \int_{-\infty}^{\infty} \mathrm{Var}[\eta(z)]dz \bullet \exp\left\{ \sigma_\varepsilon^2 \frac{\int_{-\infty}^{\infty} \mathrm{Var}[\eta'(z)]dz}{\int_{-\infty}^{\infty} \mathrm{Var}[\eta(z)]dz} \right\}$$

$$= \frac{c_{1n}}{n a_n} \exp\left(\frac{\sigma_\eta^2 c_{2n}}{c_{1n} a_n^2} \right) \qquad \text{(A.10)}$$

In addition

$$J_{n2} \geq \int_{-\infty}^{\infty} [E\eta^*(\tau) - h_\eta^*(\tau)]^2 \, d\tau$$

$$= \int_{-\infty}^{\infty} [E\eta(z) - f_\eta^*(z)]^2 \, dz$$

$$= c_{3n} a_n^4 \, . \qquad \text{(A.11)}$$

The right-hand side of (A.10) does not converge to 0 unless

$$\left(\frac{\sigma_\varepsilon^2 c_{2n}}{\quad} \right)^{1/2}$$

$$a_n > \left(c_{1n} \log n\right)$$

for all sufficiently large n, in which case (A.11) yields

$$J_{n2} \geq c\left(\frac{\sigma_\varepsilon^2 c_{2n}^2}{3n\left(c_{1n} \log n\right)}\right).$$

Therefore,

$$EJ_n \geq c\left(\frac{\sigma_\varepsilon^2 c_{2n}^2}{3n\left(c_{1n} \log n\right)}\right).$$

It follows from Markov's inequality that the rate of convergence in probability of ISE does not exceed $(\log n)^{-2}$. Q.E.D.

<u>Proof of Theorem 3</u>: Define $_\varepsilon(\tau) \equiv \left|_\eta(\tau)\right|^{1/2}$. Then

$$''_\varepsilon(z) - f_\varepsilon(z) = I_{n1}(z) + I_{n2}(z),$$

where

$$I_{n1}(z) = -(1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}\tau^2 {}_\varepsilon[\,(\tau)_\varepsilon - h_{n\varepsilon}(\tau)]g(\gamma\ \tau)d\tau$$

and

$$I_{n2}(z) = -(1/2\pi)\int_{-\infty}^{\infty} e^{-iz\tau}\tau^2{}_{n\varepsilon}[g(\gamma\ \tau)_\varepsilon - 1]h(\tau)d\tau.$$

Define

$$R_{n\varepsilon} = \sup_{|\tau| \le 1/\gamma_{n\varepsilon}} \left| {}_{\varepsilon}(\tau) - h(\tau) \right| .$$

Then

$$|I_{n1}(z)| \le (2\pi)^{-1} R_{n\varepsilon} \int_{-\infty}^{\infty} |\tau|^2 |g(\gamma_{n\varepsilon}\tau)| d\tau$$

$$= (2\pi)^{-1} (R_{n\varepsilon}/\gamma_{n\varepsilon}^3) \int_{-1}^{1} |\tau|^2 |g(\tau)| d\tau$$

$$= o_p(A_{n\varepsilon}/\gamma_{n\varepsilon}^3) + O_p(B_{n\varepsilon}/\gamma_{n\varepsilon}^3)$$

$$= o_p(1) \qquad \text{(A.12)}$$

as $n \to \infty$ uniformly over z by lemma 2, the boundedness of g, and the assumed behavior of $\{\gamma_{n\varepsilon}\}$. Now consider $I_{n2}(z)$. Observe that I_{n2} is the difference between the second derivative of the density of $\varepsilon + \gamma_{n\varepsilon}\zeta$ and f_ε, so

$$I_{n2}(z) = \int_{-\infty}^{\infty} [f_\varepsilon''(z - \gamma_{n\varepsilon} w) - f_\varepsilon''(z)] K(w) dw.$$

$$= o(1) \qquad \text{(A.13)}$$

uniformly over z by Lipschitz continuity of f_ε. The theorem follows by combining (A.12) and (A.13). Q.E.D.

REFERENCES

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density, *Journal of the American Statistical Association*, 83, 1184-1186.

Chamberlain, G. (1984). Panel Data, in *Handbook of Econometrics*, vol 2, Z. Griliches and M. Intrilligator, eds., Amsterdam, North-Holldand Publishing Company, pp. 1247-1318.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems, *The Annals of Statistics*, 19, 1257-1272.

Hsiao, C. (1986). *Analysis of Panel Data*, Cambridge, Cambridge University Press.

Pollard, D. (1984). *Convergence of Stochastic Processes*, New York, Springer-Verlag.

Stefanski, L.A. and Carroll, R.J. (1991). Deconvolution-based score tests in measurement error models, *The Annals of Statistics*, 19, 249-259.

Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators, *Statistics*, 2, 169-184.

TABLE 1: VARIABLES OF THE EARNINGS MODEL

Variable	Definition
AGE	Age of the individual in years
AGE2	Square of AGE
EDUC	Education of the individual in years
EDUC2	Square of EDUC
HI	1 if the individual has completed high school (EDUC $\geq$ 12), 0 otherwise.
COL	1 if the individual has completed college (EDUC $\geq$ 16), 0 otherwise.
MAR	1 if the individual is married, 0 otherwise

TABLE 2: RESULTS OF MONTE CARLO EXPERIMENTS WITH ESTIMATOR OF $P(\theta \mid y_1 = -1, y^* = 1)$

Distr. of ε	θ	True Probability	Mean of Estimates: With Bias Corr.	Mean of Estimates: Without Bias Corr.	Mean of Estimates: Assuming Normal ε
Normal	3	0.89	0.88	0.78	
	5	0.81	0.77	0.69	
	7	0.74	0.69	0.62	
	9	0.69	0.62	0.56	
	11	0.64	0.57	0.51	
Mixture of Normals	3	0.86	0.83	0.74	0.76
	5	0.76	0.71	0.63	0.60
	7	0.67	0.61	0.55	0.49
	9	0.60	0.54	0.49	0.41
	11	0.55	0.48	0.44	0.35

TABLE 3: RESULTS OF ESTIMATING THE EARNINGS MODEL

Variable	Estimate	Std. Error
Intercept	6.089	0.276
AGE	0.125	0.00933
AGE2	-0.00136	0.000113
EDUC	0.109	0.0403
EDUC2	-0.00390	0.00187
HI	0.235	0.0692
COL	0.373	0.0869
MAR	0.308	0.0364
σ_ε^2	0.195	
σ_U^2	0.302	

TABLE 4: SEMIPARAMETRIC AND PARAMETRIC ESTIMATES OF $P(\theta \mid y_1, y^*, x)$ FOR THE EARNINGS MODEL

		Estimate of P Based on	
α	θ	Nonparametric Density Estimates	Assumption that ε and U Are Normally Distributed
0.10	3	0.70	0.61
	5	0.58	0.43
	7	0.50	0.33
	9	0.43	0.26
	11	0.37	0.22
0.20	3	0.94	0.72
	5	0.86	0.56

7	0.79	0.46
9	0.73	0.39
11	0.68	0.34

FOOTNOTES

1. Since every individual's age changes by 1 year between 1986 and 1987, the age variable is equivalent to an intercept term.

2. The cumulative standard Cauchy distribution function is $P(x) = \pi^{-1}\tan^{-1}x + 0.5$. Thus, $\tan\{\pi[F(v) - 0.5]\}$ is the inverse Cauchy distribution function evaluated at F(v).

3. The biases of interest in this section are those of estimators of the stated functions, not of estimators of means or other parameters of the distributions of ε and U.

4. Since there are no explanatory variables in the Monte Carlo experiments, we drop x from $P(\theta | y_1, y^*, x)$ in this section.