Average Difference Estimation of Nonlinear Pricing Models

Mark Coppejans and Holger Sieg*

Duke University

Department of Economics

Box 90097

Durham, N.C. 27708

December 13, 2000

*We would like to thank Peter Arcidiacano, Ron Gallant, Burton Hollifield, Armando Levy, Tom Mroz, George Tauchen, and Francis Vella for comments and suggestions. Derek Brown provided valuable research assistance. Financial support by the William Larimer Mellon Foundation is gratefully acknowledged.

Abstract

In this paper, we derive a nonparametric test for nonlinear pricing. Our application focuses on labor markets in which nonlinear pricing arises naturally in a contractual framework. To implement our testing procedure we build on the recent literature on average derivative estimators. Our empirical results suggest that average derivative estimators are sensitive to the choice of the bandwidth parameter and yield implausible estimates of the average marginal compensation. The sensitivity arises because the main covariate (hours) is measured in discrete increments. To overcome this drawback, we develop a new estimator, the average difference estimator. We show that this estimator is consistent and root-N asymptotically normally distributed. Furthermore, the average difference estimator converges to the average derivative estimator as the increment used to compute the difference converges to zero. We thus extend the average derivative estimation framework in a number of important directions that should significantly improve its applicability. We implement the approach using repeated cross-sectional data from the CPS for a number narrowly defined occupations. We find that the average difference estimator yields quite plausible estimates for the average marginal compensation in all our samples.

KEYWORDS: Compensation, Contracts, Nonlinear Pricing, Nonparametric Estimation, Average Derivative Estimation, Average Difference Estimation, Boundary Kernels.

JEL classification: C14, D4, J3.

1 Introduction

In this paper, we derive a nonparametric test for nonlinear pricing. Our empirical application focuses on labor markets in which nonlinear pricing arises naturally in a contractual framework. To implement our testing procedure we build on the recent literature on average derivative estimators. Our empirical results suggest that average derivative estimators are sensitive to the choice of the bandwidth parameter and yield implausible estimates of the average marginal compensation. The sensitivity arises because the main covariate (hours) is measured in discrete increments. To overcome this drawback, we develop a new estimator, the average difference estimator. We show that this estimator is consistent and root-N asymptotically normally distributed. It also nests the average derivative estimator as a special case. We implement the approach using repeated cross-sectional data from the CPS for a number narrowly defined occupations. We find that the average difference estimator yields quite plausible estimates of the average marginal compensation in all our samples.

Our econometric analysis of nonlinear pricing is motivated by the recent literature in labor economics which considers contracts as an alternative mechanism for allocating hours and earnings.[1] A contract usually results from bilateral negotiations between employers and employees. Given the constraints characterizing the economic environment, the optimal contract can typically be computed by maximizing the profits of the firms subject to the constraint that workers receive at least their reservation utility. It specifies up front exactly how many hours a worker must supply and what compensation he or she will obtain in different states of the world. When the state is realized, there is no further scope for free choice at some external wage rate. Instead, the worker supplies the predetermined amount of labor and receives the corresponding compensation.[2]

While there is a large theoretical literature on contracts in labor economics, the em-

[1] Alternatively, labor market outcomes have often been modelled as if they were determined by competitive markets in which workers choose hours given the equilibrium wage rate. See, for example,Heckman and MaCurdy (1980), MaCurdy (1981), Browning, Deaton, and Irish (1985), Altonji (1986), Blundell and Walker (1986), Mroz (1987), Altug and Miller (1990), Arellano and Meghir (1992), Blundell, Duncan, and Meghir (1998) and Sieg (2000).

[2] See, for example, Holmstrom (1979) and Grossman and Hart (1983).

pirical evidence is limited. A promising approach for empirical analysis is to specify the economic environment, and to impose some functional assumptions on the production technology, the preferences of the workers, and the distribution of the unobservables.[3] One can then solve the model numerically and estimate the parameters using a full information maximum likelihood estimator. Ferrall and Shearer (1999) implement this approach and analyze payroll data of a mining company and study whether observed compensation schemes differ systematically from the optimal contract, and thus quantify efficiency losses due to deviations from the optimal.[4] There are, however, a number of difficulties encountered in estimating a structural model of contracts. The main requirement is that researchers have access to firm level data and can combine these data with detailed information about the contractual arrangements within that firm. In the absence of such detailed information, it may be hard to specify correctly the relevant economic model. It is, therefore, useful to search for alternative procedures that apply quite broadly and can be implemented with less stringent data requirements. It is desirable to develop an empirical procedure that does not require researchers to fully specify the economic environment and to use full solution algorithms to compute the optimal contract. At the same time, one would like to have a test procedure that allows researchers to investigate the hypothesis that observed outcomes are determined by contracts against the well defined alternative that outcomes are determined in completive spot markets

The approach proposed in this paper addresses these issues by focusing directly on the shape of the earnings function. In the competitive market framework, total compensation is linear in hours worked and equals the product of equilibrium wage times hours. However, if hours and earnings are allocated simultaneously through contracts instead of markets, then total compensation does not necessarily have to satisfy this linearity condition. In fact Rosen (1985) points out that "the solution of the problem (of computing optimal contracts)

[3] An alternative approach is given by Abowd and Card (1987) and Card (1990) who study the co-movement of labor supply and earnings. This approach focuses on first order conditions that earnings and hours need to satisfy under the competing hypotheses.

[4] Closely related to these articles is the empirical work on managerial compensation pioneered by Jensen and Murphy (1990), Haubrick (1994) and Margiotta and Miller (2000).

is formally identical to the theory of nonlinear pricing" (p. 1167).

The discussion above suggests a simple test to distinguish between the two competing theories. Under the null hypothesis of competitive markets, the average difference between marginal and average compensation should converge to zero. Under the alternative hypothesis, that compensation is determined by contracts, we would not expect this. The key problem encountered in implementing this approach is that marginal compensation is not directly observed by the econometrician. However, flexible nonparametric methods can be used to estimate the marginal compensation scheme. As a first approach, we adapt average derivative estimators (ADE) proposed by Härdle and Stoker (1989) and Powell, Stock, and Stoker (1989). We modify the ADE framework to account for problems that arise because the support of the main covariate, hours worked, is bounded. We use boundary kernels which have recently been used by Zhang and Karuhamuni (2000). We implement the ADE using a special sample of narrowly defined occupations based on the Current Population Survey (CPS), the best available data in the absence of firm level data. We find that estimates of the derivative of the earnings function are highly sensitive to the choice of the bandwidth parameter.

The sensitivity of the ADE is due to the fact that continuous covariates are often measured in practice in discrete increments. For example, weekly hours are typically measured in increments of one hour. Additional complications arise in our application because the empirical distribution of hours has spikes at 20, 30, 40 and 50 hours. In a contractual framework, spikes often arise because of fixed costs or adjustment costs in production functions. Consequently estimates of the density and the derivative of the density function of hours worked, key components of the ADE, are sensitive to the choice of the bandwidth. We, therefore, explore an alternative method that yields more robust results. We develop a type of difference estimator based on difference approximations of the slope of the earnings function. We show that this estimator is consistent and root-N asymptotically normally distributed. Furthermore, the average difference estimator converges to the average derivative estimator as the increment used to compute the difference converges to zero.[5] Our empirical results suggest that the average difference estimator is much better behaved in our application than the average derivative estimator. The point estimates for the average slope of the earnings functions are quite plausible for a large set of bandwidth choices. Our findings also indicate that there are substantial nonlinearities in the earnings functions for two of the three occupations analyzed in this paper. These nonlinearities suggest that marginal compensation does not necessarily equal average compensation, which supports our conjecture that nonlinear pricing due to contractual arrangements in labor market can be quite prevalent.

The rest of the paper is organized as follows. Section 2 formalizes the idea on which our test procedure is based. It also discusses econometric problems which are encountered in the implementation of the test procedure. Section 3 provides summary statistics of the data set used in the analysis. Section 4 reports the empirical results obtained in the analysis. Section 5 gives some concluding remarks.

2 The Econometric Approach

2.1 The Basic Idea

We assume that labor market allocations are determined by contracts that are negotiated between employers and employees. The procedure developed in this paper builds on Rosen's (1985) insight that the earnings function which decentralizes the optimal contract between firms and workers is not constrained to be linear in total hours. Let $y_i = m(l_i)$ denote earnings of individual i who supplied l_i units of labor. Nonlinear pricing implies that, in

[5]Closely related to our approach is recent work by Roehrig (1988) and Newey, Powell, and Vella (1999) that uses nonparametric methods to estimate structural equations. An important difference is that our method is purely nonparametric.

general, average compensation does not equal marginal compensation:

$$\frac{m(l_i)}{l_i} \neq \frac{\partial m(l_i)}{\partial l_i}. \tag{2.1}$$

In contrast, total earnings are the product of total hours worked and the market wage, denoted by w_i, if markets are competitive. This implies that average earnings equal marginal earnings, which are given by the equilibrium wage:

$$\frac{\partial m(l_i)}{\partial l_i} = w_i = \frac{y_i}{l_i}. \tag{2.2}$$

Given a sample of size N, we can then define a test statistic, T_N, as the sample average of the difference between marginal and average compensation:

$$T_N = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial m(l_i)}{\partial l_i} - \frac{y_i}{l_i}\right]. \tag{2.3}$$

Under the null hypothesis of competitive markets, the average difference between marginal and average compensation, T_N, converges to zero. Under the alternative hypothesis, that compensation is determined by contracts, T_N will converge to the difference in the population means of average and marginal compensation, which typically is not zero. The key problem encountered in implementing this approach is that marginal compensation is not directly observed by the econometrician. However, flexible nonparametric methods can be used to estimate the marginal compensation scheme.

2.2 Average Derivative Estimators

The implementation of our test procedure crucially depends on the ability to estimate the contractual earnings function and its first derivative. Earnings, y_i, of agent i can be expressed as

$$y_i \quad = \quad m(l_i) + \epsilon_i, \qquad i = 1, \ldots, N, \tag{2.4}$$

where $m(l) \equiv E[y|l]$ and $\epsilon \equiv y - E[y|l]$ given that $E|y| < \infty$. Assuming that $m(\cdot)$ is smooth, the statistic of interest is the average marginal compensation,

$$\delta \quad = \quad E[m'(l)], \tag{2.5}$$

where $m'(l) = (\partial/\partial l)m(l)$. Härdle and Stoker (1989) propose the following estimator of δ:

$$\hat{\delta}_N \quad = \quad \frac{-1}{N} \sum_{i=1}^{N} y_i \, \frac{\hat{f}_i'(l_i)}{\hat{f}_i(l_i)} \, 1\{\hat{f}_i(l_i > b_N)\}, \tag{2.6}$$

where $\hat{f}_i(\cdot)$ and $\hat{f}_i'(\cdot)$ are here boundary adjusted and leave-one-out nonparametric kernel estimates of the density, $f(l)$, and its derivative, $f'(l)$:

$$\hat{f}_i(l) \quad = \quad \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{h} K\left(\frac{l - l_j}{h}\right), \tag{2.7}$$

$$\hat{f}_i'(l) \quad = \quad \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{h^2} K'\left(\frac{l - l_j}{h}\right). \tag{2.8}$$

In addition, $1\{\cdot\}$ denotes the indicator function, and b_N is a sequence of truncation values which converges to zero. Under the appropriate assumptions, Theorem 3.1 in Härdle and Stoker (1989) establishes that

$$\sqrt{N}(\delta \; - \; \hat{\delta}_N) \xrightarrow{d} N(0, V), \tag{2.9}$$

where V is the variance of

$$r(l,y) = m'(l) + [y - m(l)]\frac{-f'(l)}{f(l)}. \tag{2.10}$$

A nice property of this estimator is that it converges at a rate of $\sqrt{N}$ despite the fact that the nonparametric estimator of the individual marginal compensation converges at a much slower rate. An additional advantage is that the statistic, δ, is a scalar, making reporting and comparisons with the sample averages easy. An alternative approach would be to nonparametrically estimate $m'(l)$ and then draw inference from this. However, the pointwise derivative converges at a relatively slow rate, $n^{-(q-1)/[2(q-1)+1]}$, where $m(l)$ is q times continuously differentiable. Moreover, reporting and drawing comparisons are more difficult given that $m'(l)$ is a function.

It is straight forward to show that, under the null hypothesis that marginal compensation equals average compensation, $\sqrt{N}T_N$ converges to a normal random variable with mean zero and variance Σ, where Σ is the variance of

$$m'(l) \ + \ [y - m(l)]\frac{-f'}{f} \ - \ \frac{y}{l} \ \ = \ \ r(l,y) \ - \ \frac{y}{l}. \tag{2.11}$$

A crucial condition used in Härdle and Stoker (1989) proof of $\sqrt{N}$ asymptotic normality is that the bias is $o(N^{-1/2})$. In the setup here, the support of the covariate, hours worked, is naturally bounded, and this causes a slight technical difficulty with respect to bounding the bias. To illustrate this, denote the bounds of the support of l as L and U, $l \in [L,U]$. Also let $\mathcal{K}(\cdot)$ be a typical order p kernel with support on $[-1,1]$. For example, $1\{-1 \leq x \leq 1\}0.75(1-l^2)$ is an order two kernel with support on $[-1,1]$. Hence the standard kernel estimator of $f(l)$ is $\bar{f}(l) = (1/Nh)\sum_{i=1}^{N}\mathcal{K}((l-l_i)/h)$. Assuming that $f(\cdot)$ is p times continuously differentiable on $[L,U]$, note that for $l > -h + U$ (i.e. points near the upper bound of the support),

$$
\begin{aligned}
E\left[\bar{f}(l)\right] &= E\left[\frac{1}{h}\mathcal{K}\left(\frac{l-x}{h}\right)\right] \\
&= \int_L^U \frac{1}{h}\mathcal{K}\left(\frac{l-x}{h}\right) f(x)\,dx \\
&= \int_{(l-U)/h}^{1} \mathcal{K}(u) f(l-hu)\,du \\
&= f(l)\int_{(l-U)/h}^{1} \mathcal{K}(u)\,du \;-\; f^{(1)}(l)\int_{(l-U)/h}^{1} u\mathcal{K}(u)\,du \\
&\quad + \cdots + \frac{(-h)^p}{p!} f^{(p)}(l)\int_{(l-U)/h}^{1} u^p\mathcal{K}(u)\,du + o(h^p),
\end{aligned}
\tag{2.12}
$$

where $f^{(p)}(l)$ denotes $(\partial^p/\partial l^p)f(l)$. The third line follows by change of variables, $u = (l-x)/h$. To better understand the range of integration in the third equation above, note that by a straight change of variables we would get that the range of integration is $[(l-L)/h, (l-U)/h]$. However, the support of the kernel is $[-1, 1]$, implying that the above range can be replaced with 1 when $1 \le (l-L)/h$ and -1 when $-1 \ge (l-U)/h$. Given that $l > -h + U$, the latter case, however, can not occur. Unless suitable modifications to the kernel are made,

$$
\int_{(l-U)/h}^{1} \mathcal{K}(u)\,du \neq 1 \tag{2.13}
$$

and

$$
\int_{(l-U)/h}^{1} u^j\mathcal{K}(u)\,du \neq 0, \qquad j = 1, \ldots, p-1. \tag{2.14}
$$

Hence the bias,

$$
E\left[\frac{1}{h}\mathcal{K}\left(\frac{l-x}{h}\right)\right] - f(l), \tag{2.15}
$$

is generally $O(1)$. Assuming more smoothness on the underlying function or letting the bandwidth parameter tend to zero at a faster rate does not decrease the order of bias for points near the boundary. An analogous result holds for $l < h + L$.

To correct for the bias near the boundary, we use the boundary kernel, $K(l)$, derived in Zhang and Karuhamuni (2000). The basic idea of this correction technique is to rescale the kernel near the boundary such that each term of the bias vanishes. In general, there are an infinite number of choices for boundary kernels. We chose kernels which have two appealing properties: they are based on the optimal kernel and they can be written in closed form.[6]

2.3 Average Difference Estimators

While it is reasonable to treat hours of work conceptually as a continuous random variable, hours are often measured in practice in discrete increments. For example, weekly hours are typically measured in increments of one hour. Additional complications arise in our application because the empirical distribution of hours has spikes at 20, 30, 40, and 50 hours. In a contractual framework, spikes often arise because of fixed costs or adjustment costs in production functions. This implies that estimates of the density and the derivative of the density function of hours are sensitive to the choice of the bandwidth (Coppejans, 2000). It is not surprising that the ADE also suffers from these drawbacks. As a remedy, we propose a difference estimator which nests the ADE as a special case. For some $\Delta > 0$, define the object of interest as

$$\gamma = \frac{1}{\pi} E[(m(l+\Delta) - m(l-\Delta))1\{L + \Delta \leq l \leq U - \Delta\}], \tag{2.16}$$

where $\pi = 2\Delta P(L + \Delta \leq l \leq U - \Delta)$. Under mild assumptions, as $\Delta \to 0$, γ converges to δ, which gives further support for the form of γ. However, the results below are for the case of when Δ is held fixed (i.e. it is independent of the sample size). The estimator of γ

[6] The details are given in Appendix B.

is defined as

$$\hat{\gamma}_N = \frac{1}{\hat{\pi}_N N} \sum_{i=1}^{N} [\hat{m}_i(l_i + \Delta) - \hat{m}_i(l_i - \Delta)]\, \hat{I}_i, \tag{2.17}$$

where $\hat{m}_i(\cdot)$ is the typical leave-one-out kernel estimator of $m(l)$:

$$\hat{m}_i(l_i) = \frac{\frac{1}{(N-1)h} \sum_{\substack{j=1 \\ j\neq i}}^{N} K\left(\frac{l_i - l_j}{h}\right) y_j}{\hat{f}_i(l_i)} \tag{2.18}$$

and

$$\hat{\pi}_N = \frac{2\Delta}{N} \sum_{i=1}^{N} 1\{L + \Delta \leq l_i \leq U - \Delta\}, \tag{2.19}$$

$$\hat{I}_i = 1\{\hat{f}_i(l_i + \Delta) > b_N, \hat{f}_i(l_i - \Delta) > b_N, L + \Delta \leq l_i \leq U - \Delta\}. \tag{2.20}$$

The following assumptions are used in order to ensure asymptotic normality of $\hat{\gamma}_N$.

Assumption 1 *The data $\{y_i, l_i\}$, $i = 1, \ldots, N$, are i.i.d., and $E|y|^\alpha < \infty$, $\alpha \geq 2$.*

Assumption 2 *The support $[L, U]$ of $f(l)$ is bounded with $f(l) > 0$ for all $l \in [L, U]$.*

Assumption 3 *The functions $f(l)$ and $m(l)$ are $q \geq 2$ times continuously differentiable.*

Assumption 4 *The integrable boundary kernel $K(\cdot)$, as defined in Appendix B, is of order $(0, p)$.*

The assumptions are extensions of those in Powell et al. (1989) and Härdle and Stoker (1989) with one key exception. In Assumption 2, it is required that $f(l) > 0$, which is common in nonparametric regression (Bierens, 1987). This type of restriction is not unreasonable here, since for most occupations, it is likely that every possible hour worked may be chosen with positive probability. In contrast, Härdle and Stoker (1989) and Powell et al. (1989) assume $f(L) = f(U) = 0$, which is used in their integration by parts calculations. In the former paper, for example, this is required to equate $E[m'(l)]$ with $E[-yf'(l)/f(l)]$. Our estimator is more direct, and as a result, this assumption is not needed.

Clearly $\hat{\pi}_N \rightarrow \pi$ in probability. We also show that $\sqrt{N}(\hat{\pi}_N\hat{\gamma}_N - \pi\gamma)$ has a normal limiting distribution with zero mean. Define $\phi > 0$ so that it is the smallest value such that $\phi\alpha > 1$ for α given in Assumption 1 (if the smallest value does not exist, define it as $\phi = [1+\eta]/\alpha$ for arbitrarily small $\eta > 0$).

Theorem 1 *Suppose that Assumptions 1–4 hold,* $b_N = o(1)$, $N^{\phi+2\epsilon-1}h^{-2} = o(1)$, *and* $h^p = o(N^{-1/2})$, $p \leq q$, *where* $\epsilon > 0$ *is arbitrarily small. Then*

$$\sqrt{N}(\hat{\pi}_N\hat{\gamma}_N \; - \; \pi\gamma) \;\; \stackrel{d}{\rightarrow} \;\; N(0,\Omega), \tag{2.21}$$

where Ω *is the covariance matrix of*

$$\begin{aligned} s(l,y) \;\; = \;\; & [m(l+\Delta) - m(l-\Delta)]1\{L+\Delta \leq l \leq U-\Delta\} \\ & + [y-m(l)]\frac{f(l-\Delta)}{f(l)}1\{l \geq L+2\Delta\} \\ & - [y-m(l)]\frac{f(l+\Delta)}{f(l)}1\{l \leq U-2\Delta\}. \end{aligned} \tag{2.22}$$

Note also that under reasonable assumptions, $s(l,y)/\pi \rightarrow r(l,y)$ as $\Delta \rightarrow 0$. There is also a tradeoff between α and q. That is, the smoother the underlying functions, the less restrictive are the conditions on the moments of y. The parameters α and ϕ are used directly in Lemma 1 in the Appendix A, which bounds the rate at which $\hat{m}_i(l)\hat{f}_i(l)$ uniformly converges to $m(l)f(l)$.

Another nice feature about the average difference estimator is that asymptotic normality can be obtained with only a second order kernel. In comparison, the average derivative estimator requires at least a third order kernel. As is well know, kernels of order higher than two are not everywhere non-negative. As a result, we would expect that higher order kernels are more sensitive to the truncation parameter and in general to have inferior small sample properties.

2.4 Extensions

In practice, we would often like to condition our analysis on an observed vector of covariates, x_i. For expositional and notational simplicity, we did not condition on x_i above. However, it is straight forward to extend the framework present in this paper to accommodate for this form of observed heterogeneity. Below, we list three different methods for controlling for observed heterogeneity.

To begin, we can define $m(l,x)$ as $E[y|l,x]$, and likewise extend the average derivative and average difference estimator with respect to l accordingly. Note that in the average derivative case, this technique has already been documented by Härdle and Stoker (1989). The results in the Appendix also cover this multidimensional case for the average difference estimator. Interestingly, the statistic still converges at a rate of root-N. Also note that when x is discrete, taking on only a few different values, for example, say $x = 1$ or $x = 2$, then the analysis from the previous subsections extend directly here by breaking the data into two sets. That is, we can construct the average difference estimator associated with $E[y|l, x = 1]$ by using only those hours worked for individuals with $x = 1$.

Even though the above extension guarantees root N asymptotics, for interpretation reasons or for small sample performance, the researcher may prefer one of the following two models: additive and partial linear additive models. The former model takes the form of $m(l,x) = m_0(l) + \sum_{j=1}^{k} m_j(x_j)$, and the latter is $m(l,x) = m_0(l) + x'\beta$, where $x \in \Re^k$. Linton (1997) and Robinson (1988), respectively, have shown how to estimate the models by kernels. Moreover, because hours worked enters linearly into both models, the average derivative and difference estimator depends on only $m_0(l)$.[7]

[7] Controlling for observed heterogeneity is less important in our application since we are looking at narrowly defined occupations in which all workers are likely to have similar educational backgrounds and work experience. Nonetheless, in the empirical analysis, we conduct some sensitivity analysis to evaluate whether our main findings are affected by conditioning on additional observed characteristics.

3 Data

The data set we use in this paper is from the 1989, 1994, and 1999 March supplements of the Current Population Survey (CPS).[8] The large sample size of the CPS allows us to analyze earnings function for narrowly defined occupational group, making it ideal for this study. Initially we explored a sample of 12 occupation specific subgroups. In this paper, we report results from the following three groups: truckers, wholesale sales representatives, and cooks.[9] These occupations were chosen mainly because of the relatively large number of respondents.

For earnings and hours worked, the supplement contains questions pertaining to jobs held "last year". We need to isolate earnings, hours and weeks worked for *specific jobs* worked last year. The CPS distinguishes two kinds of wage and salary income from jobs held last year: earnings from the "longest job held last year," and earnings "from work other than longest job held last year."[10] Hours and weeks worked last year are only collected at the aggregate level. We know only the average number of hours worked per week and the total number of weeks worked *at all jobs*. In other words, it is impossible to identify job-specific hours and weeks if an individual held more than one job in the past year. We thus exclude from the analysis anyone who held multiple jobs during the year priors to the survey.

The years, 1989, 1994 and 1999 are chosen for practical reasons. Prior to 1988, wage and salary income is also reported as an aggregate for all jobs, and the CPS does not

[8] The CPS is collected monthly by the US Bureau of the Census for the Bureau of Labor Statistics. The survey consists of a core questionnaire and for most months, a supplement. See http://www.bls.census.gov/cps/ for details. The March supplement of the CPS contains detailed work and income information at both the household and the individual level.

[9] Results for the other nine groups which include carpenters, construction laborers, cashiers, bank tellers, securities and financial services sales occupations, real estate sales occupations, auto mechanics, registered nurses, and receptionists, are available upon request from the authors.

[10] Hourly earnings and whether a worker was paid hourly or not are both collected only for the "earner study" rotational groups, a small subset of the March CPS. See the CPS web page for more details. The earner study groups are selected individuals who are asked some additional employment questions. Respondents are asked a few additional employment questions, making it useful for certain kinds of studies but not for ours. Aside from the hourly earnings variables and union status, no further information about earnings, hours, or weeks worked is collected in the earner study.

Table 1: Average Income, Hours, and Weeks Worked

	Year	No. of Obs.	Income Ave.	Income SD	Hours Ave.	Hours SD	Weeks Ave.	Weeks SD
Truckers	1999	511	33403.65	15955.93	47.32	10.73	50.28	5.07
	1994	435	29269.80	14366.13	49.35	12.08	50.10	5.24
	1989	394	25570.39	12090.21	50.98	12.42	49.78	5.43
Wholesale	1999	368	44909.01	25768.35	44.19	8.44	51.20	3.79
	1994	380	36201.51	19570.81	45.48	8.93	50.98	4.49
	1989	434	29276.92	15512.18	44.45	8.57	51.00	3.84
Cooks	1999	599	13178.05	9703.56	38.13	10.26	48.80	7.23
	1994	552	10752.23	7592.72	37.34	10.37	47.72	8.19
	1989	449	8819.16	6088.71	36.12	8.98	48.16	7.38

ask if an individual worked more than one job. To minimize outliers and keep our groups homogeneous, we deleted any observation where less than 26 weeks of employment during the past year is reported, or where less than 20 hours of work per week is reported. We also exclude a small number of individuals who reported non-incorporated self-employment as their class of employment. Income and hours worked in the CPS is top-coded (truncated) for individuals who report large values. The number of hours worked is top-coded at 99 hours. With respect to income, top-coding has varied according to race, sex, and worker status since 1996. To simplify the analysis, we deleted these observations from our sample. The number of observations deleted is small, no more than about a fifteen for any group.

Table 1 provides summary statistics for income, and number of hours and weeks worked by group. Average income is lowest for retail cooks in all years; income is highest for wholesale representatives sales. The number of weeks worked last year is nearly constant in each group across the three years. The survey question includes paid vacation and leave, so it is not surprising to find 50 to 51 weeks of work reported after we discard observations who report less than 26 weeks of work. After discarding individuals who worked less than

Table 2: Variation in Hours Worked

	Year	% Hours/Week 20	30	40	50	60	Min	Max
Truckers	1999	0.98	1.17	38.94	15.26	11.35	20	96
	1994	0.23	1.38	32.87	16.32	13.79	20	96
	1989	0.76	1.27	31.98	18.27	10.15	20	96
Wholesale	1999	2.72	1.09	50.00	18.21	6.79	20	80
	1994	0.79	1.84	45.53	18.42	7.89	20	90
	1989	1.61	0.69	48.39	16.36	4.84	20	85
Cooks	1999	9.35	6.34	46.08	2.67	2.17	20	84
	1994	8.51	8.88	35.69	4.71	2.90	20	96
	1989	10.24	10.47	40.09	5.35	0.67	20	80

20 hours per week on average, we find that several blue collar workers still report less than 40 hours per week worked. Truckers report the highest number of average hours per week in all years; cooks report the shortest average work-week.

Table 2 provides more detailed information about the variation in hours worked. Since many individuals report round numbers for average hours worked per week, we provide the main "cluster points" of 20, 30, 40, 50, and 60. Some of this variation is certainly due to institutional characteristics of selected occupations. The minimum and maximum number of hours for each group is also provided.

Table 3 provides some information on demographics and age. Our demographic variables are standard measures collected in the CPS. Race is delineated as white, black, or other. Educational attainment (highest grade attended and highest completed, through two or more years of college) and age control for potential work experience.

Table 3: Demographic Summary Statistics

	Year	Ave. Age	% Male	% White	Education Level[a] <HS	=HS	>HS	% in Census Region[b] NE	MW	S	W
Truckers	1999	43	96	85	21	54	25	18	23	36	23
	1994	42	97	90	20	56	25	23	26	33	18
	1989	40	99	89	26	54	20	24	30	34	13
Wholesale	1999	41	76	95	5	24	71	20	24	30	27
	1994	41	81	96	3	28	69	22	28	33	17
	1989	39	82	97	4	30	65	26	26	33	15
Cooks	1999	34	64	80	45	35	20	19	16	32	32
	1994	33	63	76	38	40	22	21	22	31	27
	1989	32	56	78	31	52	17	18	29	31	23

[a] HS = High School.

[b] NE=New England and Middle Atlantic, MW=East & West North Central, S=East & West South Central and South Atlantic, and W=Mountain and Pacific.

4 Empirical Results

4.1 Implementation

In order to implement the estimators, the truncation, b_N, and the bandwidth, h, parameters need to be determined. From the results in Härdle and Stoker (1989), asymptotic normality holds in the average derivative case if we assume $q \geq 4$ and set $p = 4$, $b_N = c_b N^{-1/12+0.0002}$, and $h = c_h N^{-1/6-0.0001}$, where c_b and c_h are some positive constants. Of course, finite sample results will be sensitive to the choices of these constants. It is therefore useful to review the rationale for determining b_N and h below. For expositional simplicity, we only discuss the case of choosing the bandwidth for the density estimators and not for the average derivative or average difference estimator.

Define $\hat{H}_t$ as the histogram estimate of $f(l)$, where the bin width is fixed at one hour, $t = L, L+1, \ldots, U$, $L = 20$, $U = 89$, $\hat{H}_t = n_t/N$, and n_t is the number of individual who have worked t hours. Let $\hat{H} = \min_t \hat{H}_t$. Then we set c_b equal to $\max\{\hat{H}, N^{-1}\}$, where the latter term ensures that $c_b > 0$. The rationale for using this stems from the fact that $\hat{H} \to H^*$ at a rate $\sqrt{N}$, where

$$H^* = \min_{\{t:t=L,\ldots,U\}} \int_{t-1/2}^{t+1/2} f(l)\,dl, \tag{4.1}$$

and $H^* > 0$, since by assumption, $f(l) > 0$. Because it is non-binding, CH^* for any constant C, $0 < C < 1$, would be a valid truncation rule in this case, which occurs with high probability given our choice of b_N.

In the literature, there are fundamentally two different data driven approaches for determining the bandwidth parameter, h: cross-validation and plug-in. Given any kernel estimator $\hat{f}(\cdot)$ of a density $f(\cdot)$, cross-validation selects h by minimizing, with respect to h, an estimate of

$$\int [\hat{f}(l) - f(l)]^2\,dl. \tag{4.2}$$

Note that this is equivalent to minimizing

$$\int \hat{f}(l)^2\,dl \; - \; 2\int \hat{f}(l)f(l)\,dl. \tag{4.3}$$

The first term in the above equation is relatively easy to compute, but the second term is more difficult, at least conceptually, because it depends on the unknown $f(l)$. The most common solution is to approximate the term by $-2/N\sum_{i=1}^{n}\hat{f}_i(l_i)$. But when the data is very discrete, we have for small h and large N that this sum is approximately $[1/N(N-1)h]K(0)\sum_{t=L}^{U} n_t(n_t-1)$. That is, only those observations for which $l_i = l_j, j \neq i$, have positive weight (e.g. if $l_i = 40$ hours and $l_j = 41$ hours, then $K((l_i - l_j)/h) = K(1/h)$ is arbitrarily small or zero if $K(\cdot)$ has compact support). But $\int \hat{f}(l)^2\,dl$ is approximately $(1/N^2h)[\int K(u)^2\,du]\sum_{t=L}^{U} n_t^2$, where for most kernels, $\int K(u)^2\,du \leq 2K(0)$. Also note that $(1/N^2)\sum_{t=L}^{U} n_t^2 < [1/N(N-1)]\sum_{t=L}^{U} n_t(n_t-1)$. Hence, in terms of $h \to 0$, $\int \hat{f}(l)^2\,dl$ can tend to infinity slower than $2\sum_{i=1}^{n} \hat{f}_i(l_i)$. As a result, cross-validation often selects the degenerate value $h = 0$ when the data is too discrete, and this is precisely the problem here.[11]

In contrast, plug-in methods work off the approximate mean integrated squared error (AMISE). Using standard change of variables and Taylor expansion techniques, the AMISE for the density is

$$\frac{h^8}{(4!)^2}\left[\int u^4 K(u)\,du\right]^2 \int [f^{(4)}(l)]^2\,dl \;+\; \frac{1}{nh}\int K(u)^2\,du, \tag{4.4}$$

which is just the integrated bias squared and variance decomposition.[12] The difficulty with minimizing the AMISE with respect to h is that $f^{(4)}(l)$ is unknown. There are two common methods for overcoming this. The first is to nonparametrically estimate $f^{(4)}(l)$ using a different bandwidth parameter (which requires $f(\cdot)$ to be even smoother than four times continuously differentiable)—this is called the pilot estimate. As Loader (1999) has shown, the plug-in method is sensitive to the pilot estimate.[13] Clearly, since the density itself is difficult to estimate, higher order derivatives will be at least as difficult, so we did not adopt this approach here.[14]

The second method to implement the plug-in is more promising in our application. It uses a baseline density to calculate the AMISE. If $f(l)$ is normally distributed with a variance of σ_l^2, we can calculate $\int [f^{(4)}(l)]^2\,dl$. Given the kernel in Appendix B, we show that the optimal bandwidth parameter under this distributional assumption for the density estimator is $3.0295\sigma_l n^{-1/9}$. This leads to a natural choice of c_h as $3.0295\hat{\sigma}_l$, where $\hat{\sigma}_l$ is the sample standard deviation. Since standard methods to determine the bandwidth parameter are likely to be problematic in our application, we adapt a dual approach in this paper.

[11] See Silverman (1986), especially pp. 51–52, for a more rigorous argument, and Coppejans (2000) for more examples of this type of phenomenon.

[12] See Silverman (1986), pp. 38–40.

[13] See Loader (1999) for a more thorough critique of this technique in general.

[14] Härdle, Hart, Marron, and Tsybakov (1992) developed conditions under which the plug-in estimator is theoretically justified for the average derivative estimator. Powell and Stoker (1996) have a similar result for density weighted averages.

First, we report estimates which are based on the plug-in using normal approximations of the density. Second, we compute all estimators for a large number of bandwidth choices and evaluate how sensitive the empirical findings are with respect to the choice of the bandwidth parameter.

4.2 Average Derivative Estimation Results

Table 4 summarizes our findings using average derivative estimators. It reports our estimates for the sample average of average compensation and our estimate of average marginal compensation. The p-value in the last column corresponds to the null hypothesis that $T_N = 0$.[15]

Table 4: Average Derivative Estimates of Wages

	Year	No. of Obs.	Samp. Ave.	Ave. Der.	P-Value
Truckers	1999	511	14.2168	11.5409	0.2784
	1994	435	12.1197	19.1002	0.0054
	1989	394	10.4289	17.0400	0.0138
Wholesale	1999	368	19.9888	36.8189	0.0012
	1994	380	15.5954	-5.7072	0.2406
	1989	434	12.9959	17.1708	0.3922
Cooks	1999	599	6.8788	0.1622	0.7104
	1994	552	5.8430	-9.1711	0.0178
	1989	449	4.8365	6.7841	0.5771

P-Value is with respect to equality of the sample average and the average difference estimate, T_N.

In general, the results shown in Table 4 are not encouraging. The estimates of the average derivative of the earnings function do not show any clear pattern. Sometimes the

[15] We find that the estimated asymptotic standard errors, as compared to the estimates themselves, are overly sensitive to the choice of the truncation parameter. We therefore report standard errors that are calculated by bootstrapping 100 times from the initial sample.

estimates are negative, which is counterintuitive. In other cases, they are twice as large as the average compensation. Furthermore, estimates of the average derivative appear to be overly sensitive to the choice of the time period. For example, it is implausible that average marginal compensation of wholesale representatives changed from -5.1 in 1994 to 36.81 in 1999.

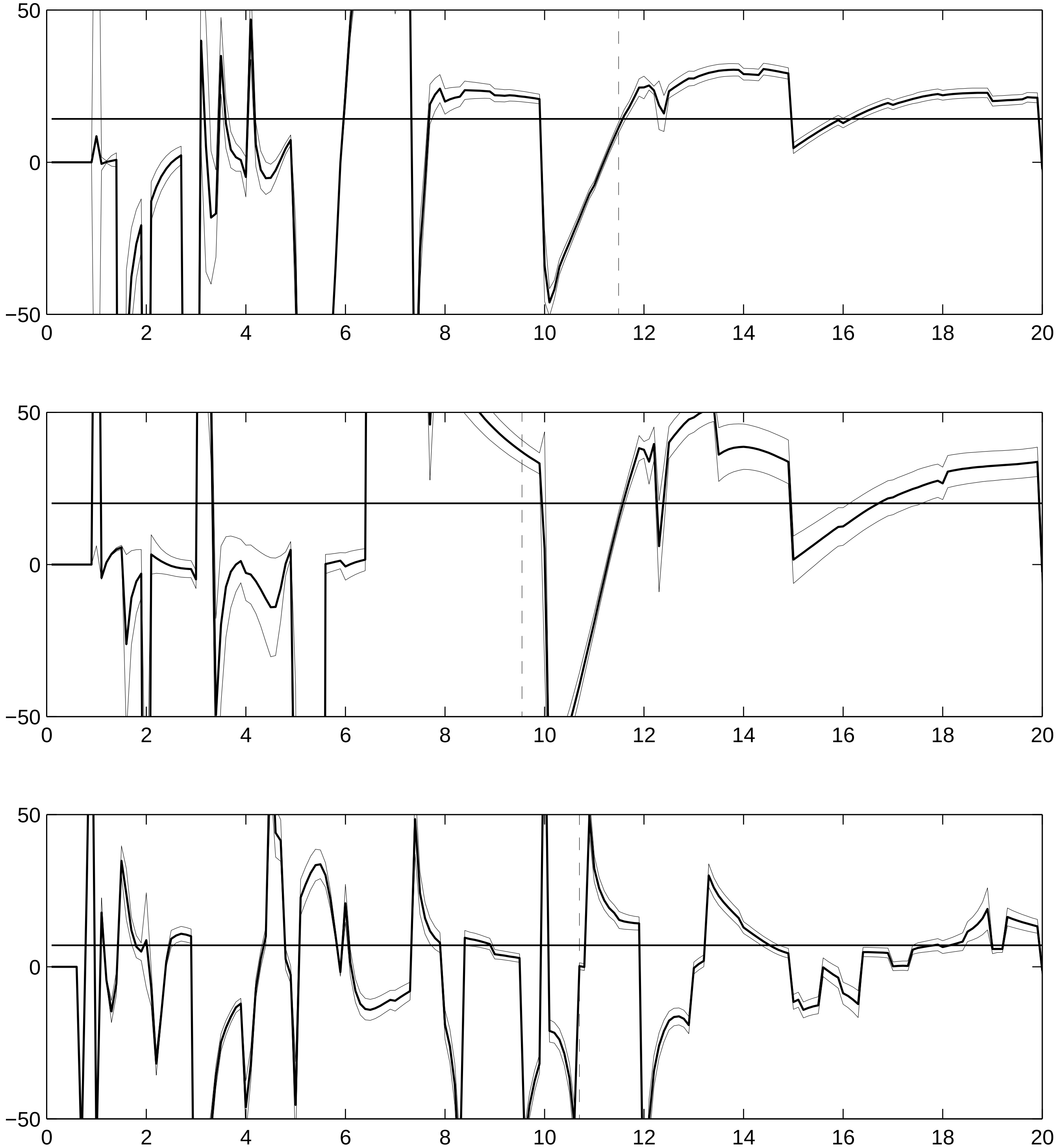
In order to gain additional insights into the small sample properties of this estimator, we plot the average derivative estimates for the year 1999 for bandwidths ranging from 0.1 to 20.0. The results are shown in Figure 1. Our findings suggest that the average derivative estimator is not stable in this application. The estimates depend crucially on the choice of the bandwidth parameter—a slight change to it can drastically affect the estimate—even so much as going from a large positive number to a large negative one. This instability is not due to the form of the kernel. We find qualitatively similar results across different kernels, for both boundary and standard ones, and stable across different values for b_n.

4.3 Average Difference Estimators

The unsatisfactory performance of the average derivative estimator is the main motivating factor behind the average difference estimator. For this estimator, we set Δ equal to five hours. The rationale is that the clustering appeared in mainly blocks of ten hours, as can be seen in Table 2, which corresponds to the number of hours over which we are approximating the derivative (i.e. $10 = 2\Delta = [l + \Delta] - [l - \Delta]$). Table 5 summarizes our findings using average difference estimators. It reports estimates for the sample average of average compensation and estimates of average marginal compensation using standard second order kernels and fourth order boundary kernels. The p-value in the last column corresponds to the null hypothesis that $T_N = 0$.

Several interesting patterns emerge from an analysis of Table 5. For truckers and wholesale sale representatives, the average difference estimates are below the average wage. This difference is usually significant, indicating the presence of nonlinearities in the earnings function. For cooks, it appears that the reverse ordering holds though the results are less

Figure 1: Average Derivative Estimator



The bold solid line is the average derivate estimate across different bandwidths, the light solid lines are the bootstrap standard errors, the solid line is the sample average hourly wage, and the dashed vertical line is the optimal bandwidth under a normal density. The top plot is truckers, the middle is wholesale, and the bottom is cooks for the year 1999.

Table 5: Average Difference Estimates of Wages

	Year	No. of Obs.	Samp. Ave.	Ave. Diff.[a]	P-Value	Ave. Diff.[b]	P-Value
Truckers	1999	511	14.2168	8.1981	0.0000	7.0279	0.0000
	1994	435	12.1197	5.3112	0.0000	3.7539	0.0000
	1989	394	10.4289	6.5757	0.0002	5.0236	0.0000
Wholesale	1999	368	19.9888	8.6154	0.0004	7.6216	0.0000
	1994	380	15.5954	10.5266	0.0138	9.4143	0.0000
	1989	434	12.9959	6.1748	0.0008	5.7622	0.0000
Cooks	1999	599	6.8788	8.2242	0.2302	6.4909	0.5633
	1994	552	5.8430	8.3537	0.0068	6.0018	0.5946
	1989	449	4.8365	6.4376	0.0368	5.3531	0.1867

P-Value is with respect to equality of the sample average and average derivative estimate, T_N.

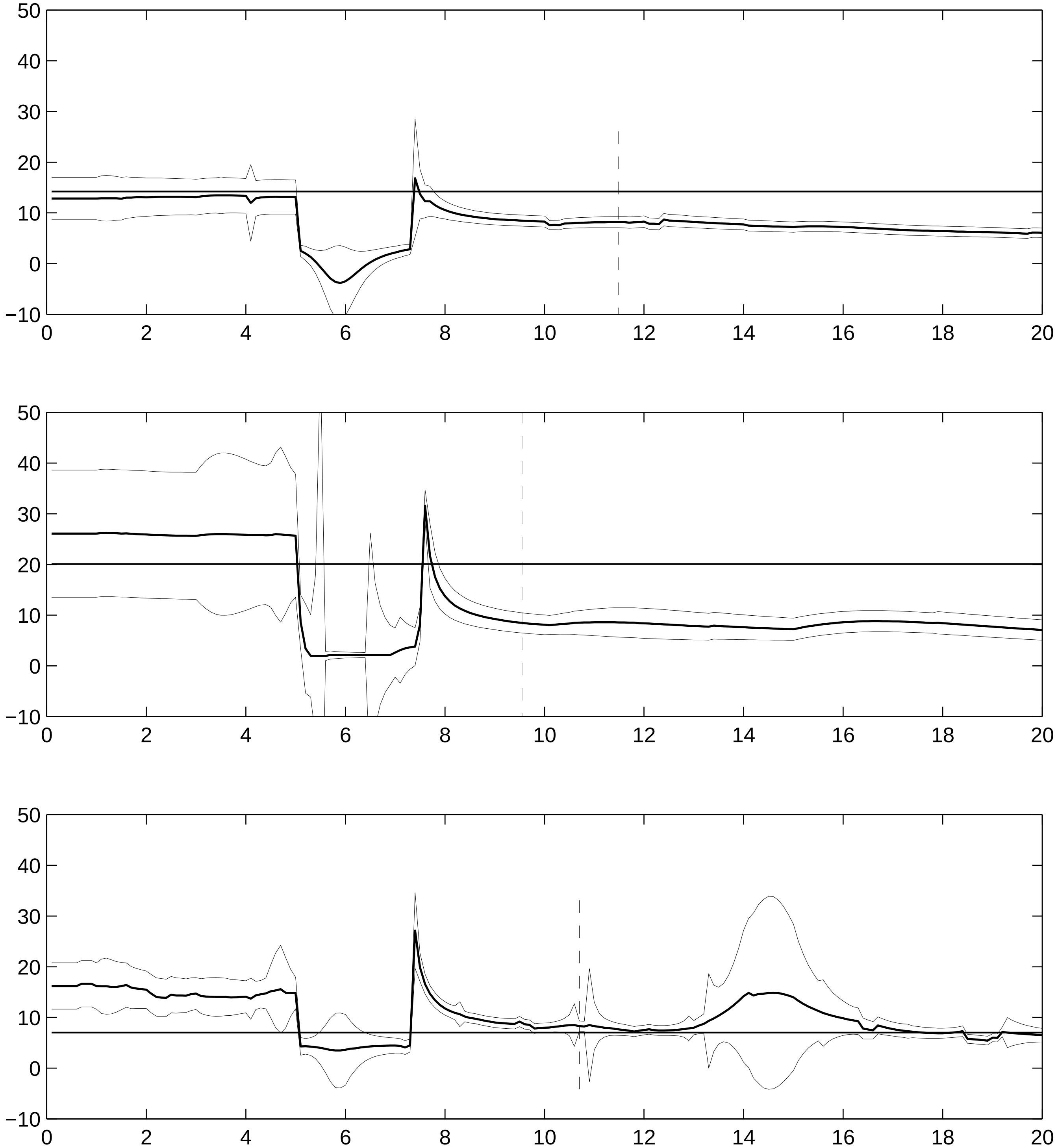
[a] Forth order boundary kernel.

[b] Second order standard kernel (not adjusted for points near the boundary).

significant. Comparing the estimates based on second order kernels to those obtained using fourth order boundary kernels, we find that the choice of the kernel does not affect our main conclusions. In summary, there appear to be significant nonlinearities for two of the three professions in our sample, as predicted by the contractual paradigm.

We investigate whether the main findings are due to demographic or experience effects. To control for this, we use age as a proxy for experience. We estimate the average difference for a sample that consists of individuals under the age of fifty and not from the South. Table 6 summarizes our findings using average difference estimators. Comparing the results shown in Table 6 with those reported in Table 5, we find that there are small quantitative differences both in our estimates for marginal and average compensation. However, these differences do not affect our main findings regarding nonlinearities in the earnings functions. We therefore conclude that our main results are not strongly influenced by controlling for this type of observed heterogeneity.

Figure 2: Average Difference Estimators—Fourth Order Boundary Kernel



The bold solid line is the average derivate estimate across different bandwidths, the solid lines are the bootstrap standard errors, the solid line is the sample average hourly wage, and the dashed vertical line is the optimal bandwidth under a normal density. The top plot is truckers, the middle is wholesale, and the bottom is cooks for the year 1999.

Table 6: Average Difference Estimates of Wages
Under 50 & Non-Southerner

	Year	No. of Obs.	Samp. Ave.	Ave. Diff.[a]	P-Value	Ave. Diff.[b]	P-Value
Truckers	1999	240	14.2107	10.4239	0.0845	7.9163	0.0000
	1994	215	12.3334	4.3666	0.0000	2.6720	0.0000
	1989	200	10.4044	6.2227	0.0068	3.2247	0.0000
Wholesale	1999	203	20.0911	5.4445	0.0010	4.2441	0.0000
	1994	199	15.1404	8.2580	0.0039	6.6208	0.0000
	1989	225	13.8261	4.6474	0.0018	5.0185	0.0000
Cooks	1999	365	7.0259	7.9336	0.4510	5.6990	0.0532
	1994	332	5.8537	7.0488	0.2168	4.7347	0.0394
	1989	267	4.8907	5.8276	0.2514	4.9900	0.7806

P-Value is with respect to equality of the sample average and average derivative estimate, T_N.

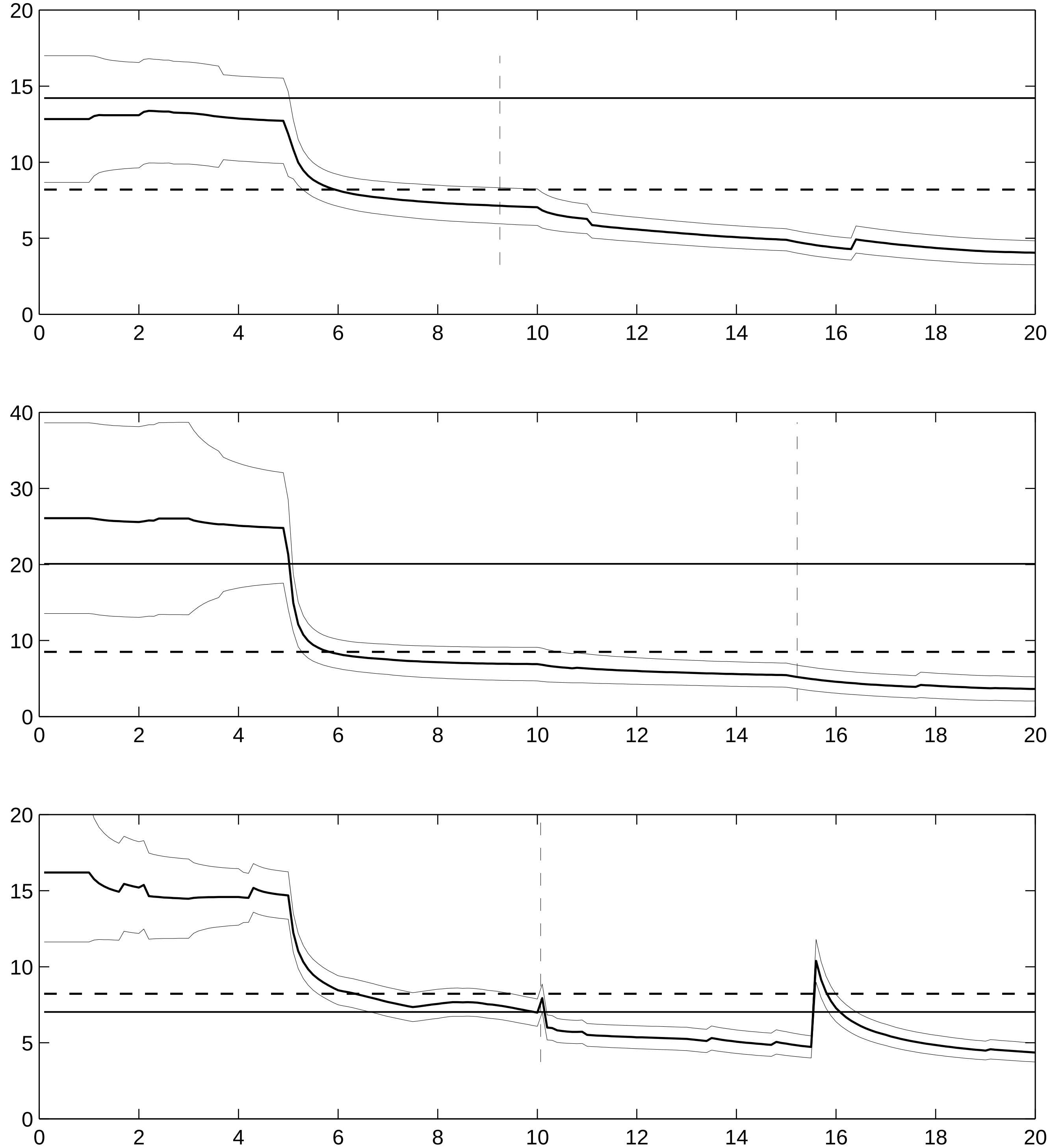
[a] Forth order boundary kernel.

[b] Second order standard kernel (not adjusted for points near the boundary).

To get a better idea of how sensitive our results are to the choice of the bandwidth parameter, we plot the average difference estimates for the year 1999 for bandwidths ranging from 0.1 to 20.0. The results of this exercise are shown in Figure 2. A comparison between Figure 1 and Figure 2 shows that the estimates based on the average difference estimator are much more stable than those obtained using the average derivative estimators. This is true even for small bandwidth parameters. The notable exception in all cases is when the bandwidth is between 5.0 and 8.0. The cause of this trough is specific to the choice of the kernel, Δ, and the discreteness of the data. For example, increasing Δ moves this trough to the right. All of the bandwidths used in this experiment were to the right of this trough; however, if $h \in (5, 8)$, then the researcher would have to decide whether or not to set $h = 5$ or $h = 8$, for example. This exercise also illustrates the importance of plotting out the estimates with respect to different choices of the bandwidth parameter.

We also computed the estimates for the average difference estimator when only a second

Figure 3: Average Difference Estimator—Second Order Boundary Kernel



The bold solid line is the average derivate estimate across different bandwidths, the solid lines are the bootstrap standard errors, the solid line is the sample average hourly wage, the dashed line is the chosen average difference estimate under a fourth order kernel, and the dashed vertical line is the optimal bandwidth under a normal density. The top plot is truckers, the middle is wholesale, and the bottom is cooks for the year 1999.

order boundary kernel is used. These results are shown in Figure 3. Not surprisingly, the plots in this case are more stable than when the fourth order kernel is used. In support of our methodology, we find that the two sets of estimates are quite similar under the above mentioned choices for the bandwidth.

Furthermore, we investigate whether the boundary kernel makes a difference. Given that the occupation cooks has the largest percentage of individuals working twenty hours, it would appear that the boundary kernel would have its greatest influence in this case. The plots in Figure 4 suggest that perhaps it even makes things worse. Hence the conclusion we draw is that discreteness appears to be a more serious problem than boundary effects.

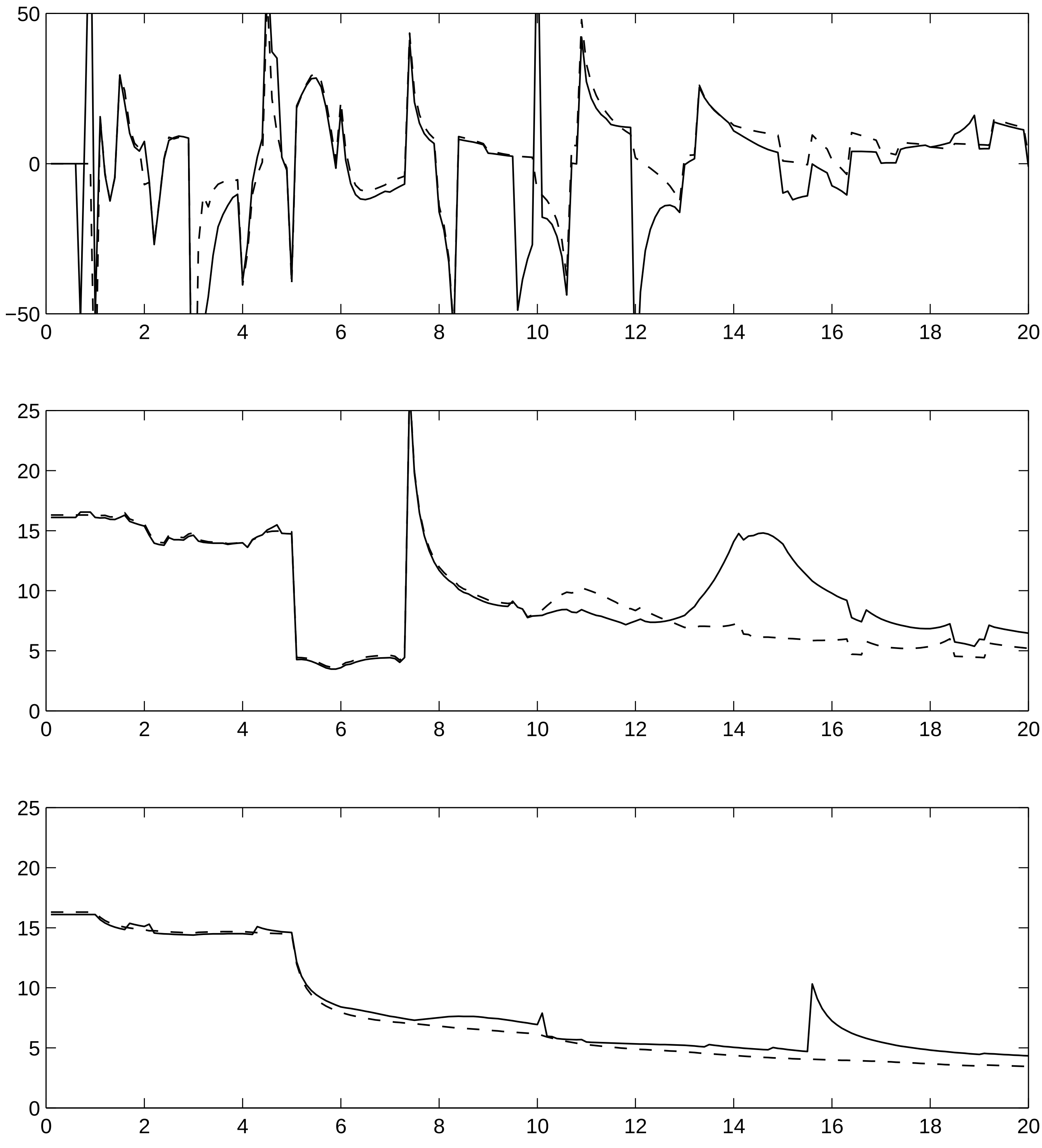
As a final robustness check, we regressed average compensation on hours using polynomial expansions. This analysis is in the spirit of Newey et al. (1999), but extended to the purely nonparametric case. Consider first the linear model:

$$\frac{y}{l} = \beta_0 + \beta_1 l + \epsilon. \tag{4.5}$$

The results for the linear model are reported in Table 7.

When the average difference estimate, as reported in Tables 5 and 6, is smaller than the sample average, we would expect the slope coefficient, β_1, to be negative on average. This relationship does indeed hold. The one main inconsistency with the previous results, however, is that the p-value for wholesale sales representatives is much larger here. Surprisingly, the regression is only stable in the linear case. When we add higher powers of l as regressors, such as just l^2, all of the estimated coefficients in the regression tended to be insignificant, even estimates of the constant and slope terms. This suggest that perhaps a polynomial series is not a good basis in the setting here.

Figure 4: Boundary vs. Standard Kernel



The solid line is the boundary kernel, and the dashed line is the standard kernel. The top plot is the average derivative, the middle is fourth order average difference, and the bottom is second order average difference. All plots are for cooks for the year 1999.

Table 7: Parametric Estimates

	Year	Entire Sample Cnst.	Entire Sample Slope	Entire Sample P-value	Under 50 & Non-Southerner Cnst.	Under 50 & Non-Southerner Slope	Under 50 & Non-Southerner P-Value
Truckers	1999	18.3416	-0.0872	0.0008	16.4990	-0.0492	0.2587
	1994	17.1549	-0.1020	0.0000	18.6117	-0.1256	0.0000
	1989	14.0424	-0.0723	0.0000	14.7920	-0.0851	0.0006
Wholesale	1999	26.3779	-0.1446	0.0530	27.5651	-0.1684	0.1171
	1994	17.4279	-0.0403	0.3756	19.6638	-0.0984	0.0464
	1989	14.5151	-0.0342	0.3301	16.0703	-0.0498	0.3270
Cooks	1999	5.6191	0.0330	0.1242	5.7209	0.0338	0.1790
	1994	5.1730	0.0180	0.1960	5.7317	0.0033	0.8566
	1989	2.6830	0.0596	0.0000	2.4199	0.0701	0.0000

P-Value is for the slope coefficient equal to zero. The standard errors for the constant term are relatively small and therefore the corresponding P-value is not given.

5 Conclusions

In this paper, we derive a simple nonparametric test for nonlinear pricing. Our application focuses on labor markets in which nonlinear pricing arises naturally in a contractual framework. To implement our testing procedure, we build on the recent literature on average derivative estimators. We extend the standard estimator to accommodate for boundary effects. Our results suggest that average derivative estimators are sensitive to the choice of the bandwidth parameter and yield implausible estimates of the average marginal compensation in all three subsamples. The sensitivity arises because hours is measured in discrete increments.

To overcome the drawbacks associated with the ADE, we develop a new estimator, the average difference estimator, which nests the ADE as a special case. The average difference estimator yields quite plausible estimates for the average marginal compensation in the the three subsamples. These results provide some convincing evidence in favor of nonlinear pricing. Our main empirical findings are robust with respect to modifications of our estimation strategy. Using second order kernels improves our estimates on the margin. Failing to control for boundary effects is on the other hand immaterial. Measuring continuous variable in discrete increments is much more problematic in nonparametric estimation than are boundary effects.

We have thus extended the average derivative estimation framework in a number of important ways. These extensions should significantly improve the applicability of this framework. We have also developed a general testing procedure for detecting nonlinear pricing. This testing procedure has broad applications outside of labor economics and contract theory. As an example, consider the market for higher education. One of the main issues in educational economics is to determine whether universities discriminate by income or ability when setting tuition policies. Determining the degree of price discrimination has significant welfare and policy implications. Since ability is also measured in discrete increments using, for example, SAT scores, the econometric techniques discussed in this paper should help to address this issue as well.

A Proof of Theorem 1

The outline of the proof follows the strategy used in Härdle and Stoker (1989). Let $g(l) = m(l)f(l)$, $\hat{g}(l) = \hat{m}(l)\hat{f}(l)$, and we generalize to the vector case, $l \in \Re^k$. We begin with the following Lemma used repeatedly below.

Lemma 1 *Suppose that the conditions in Theorem 1 hold. Then for any* $\epsilon > 0$,

$$\begin{aligned}
\sup_l |\hat{f}_i(l) - f(l)| &= O_p[(N^{1-\epsilon}h^k)^{-1/2} + h^p], \\
\sup_l |\hat{g}_i(l) - g(l)| &= O_p[(N^{1-\epsilon-\phi}h^k)^{-1/2} + h^p].
\end{aligned}$$

Proof: The proof of the first result can be found by slightly altering the results in Bhattacharya (1967) to the leave-one-out and multivariate case. The second result is not new, however we could not find an existing result in the literature in this stated form. The proof is composed of bounding the terms $\sup|\hat{g}_i(l) - E[\hat{g}_i(l)]|$ and $\sup|E[\hat{g}_i(l)] - g(l)|$. By well known techniques (e.g. change of variables and Taylor expansion), the latter piece is $O(h^p)$. Bounding the former piece requires an application of Hoeffding's inequality, which is only valid for bounded variables. To this end, denote $\mathcal{Y}_N$ as the event $\{|y_i| \leq N^{\phi}, i = 1, \ldots, N\}$ for some $\phi > 0$. Then conditioned on $\mathcal{Y}_N$, we have by Bhattacharya's (1967) Lemma 2 that $\sup|\hat{g}_i(l) - E[\hat{g}_i(l)]|$ is $o_p(N^{1-\epsilon}h^k N^{-\phi})^{-1/2}$. To finish the proof, we need to show that the compliment of $\mathcal{Y}_N$, $\mathcal{Y}_N^c$, is $o_p(1)$. Note that by Markov's inequality,

$$\begin{aligned}
P(\mathcal{Y}_N^c) &\leq NP(y \geq N^{\phi}) \\
&\leq N\frac{E|y|^{\alpha}}{N^{\phi\alpha}} \to 0.
\end{aligned}$$

Q.E.D.

There are four steps to the proof of Theorem 1. Without loss of generality, set $\hat{\pi}_N = \pi = 1$.

Step 1: Linearization. Denote $\hat{\gamma}_N$ with $\hat{I}_i$ replaced with I_i, where

$$\begin{aligned} I_i &= 1\{f_i(l_i+\Delta) > b_N, f_i(l_i-\Delta) > b_N, L+\Delta \le l_i \le U-\Delta\} \\ &= 1\{L+\Delta \le l_i \le U-\Delta\} \qquad \text{(for small } b_N\text{)}, \end{aligned}$$

as

$$\bar{\gamma}_N = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\hat{g}_i(l_i+\Delta)}{\hat{f}_i(l_i+\Delta)} - \frac{\hat{g}_i(l_i-\Delta)}{\hat{f}_i(l_i-\Delta)}\right] I_i.$$

Define a linearization of $\bar{\gamma}_N$ as

$$\tilde{\gamma}_N = \tilde{\gamma}_{0,N} + \tilde{\gamma}_{1,N} - \tilde{\gamma}_{2,N},$$

where

$$\begin{aligned} \tilde{\gamma}_{0,N} &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{g_i(l_i+\Delta)}{f_i(l_i+\Delta)} - \frac{g_i(l_i-\Delta)}{f_i(l_i-\Delta)}\right] I_i, \\ \tilde{\gamma}_{1,N} &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\hat{g}_i(l_i+\Delta)}{f_i(l_i+\Delta)} - \frac{\hat{g}_i(l_i-\Delta)}{f_i(l_i-\Delta)}\right] I_i, \\ \tilde{\gamma}_{2,N} &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{g_i(l_i+\Delta)\hat{f}_i(l_i+\Delta)}{f_i^2(l_i+\Delta)} - \frac{g_i(l_i-\Delta)\hat{f}_i(l_i-\Delta)}{f_i^2(l_i-\Delta)}\right] I_i. \end{aligned}$$

We will show that $\sqrt{N}(\bar{\gamma}_N - \tilde{\gamma}_N) = o_p(1)$. Observe that with high probability,

$$\begin{aligned} \sqrt{N}(\bar{\gamma}_N - \tilde{\gamma}_N) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Bigg\{ & \Bigg[\frac{[g_i(l_i+\Delta) - \hat{g}_i(l_i+\Delta)]\left[\hat{f}_i(l_i+\Delta) - f_i(l_i+\Delta)\right]}{f_i^2(l_i+\Delta)} \\ & - \frac{\hat{g}_i(l_i+\Delta)\left[\hat{f}_i(l_i+\Delta) - f_i(l_i+\Delta)\right]^2}{\hat{f}_i(l_i+\Delta)f_i^2(l_i+\Delta)}\Bigg] \\ & + \Bigg[\frac{[g_i(l_i-\Delta) - \hat{g}_i(l_i-\Delta)]\left[\hat{f}_i(l_i-\Delta) - f_i(l_i-\Delta)\right]}{f_i^2(l_i-\Delta)} \end{aligned}$$

$$- \frac{\hat{g}_i(l_i - \Delta)\left[\hat{f}_i(l_i - \Delta) - f_i(l_i - \Delta)\right]^2}{\hat{f}_i(l_i - \Delta) f_i^2(l_i - \Delta)} \Bigg] \Bigg\} I_i.$$

By repeated applications of Lemma 1, we get the desired result because the conditions in Theorem 1 ensure that

$$\sup_l |\hat{f}_i(l) - f(l)| \sup_l |\hat{g}_i(l) - g(l)| \quad = \quad o_p(N^{-1/2}).$$

Step 2: Asymptotic Normality of $\sqrt{N}(\tilde{\gamma}_N - E[\tilde{\gamma}_N])$. Because $\tilde{\gamma}_{0,N}$ is a sample average of i.i.d. observations with $E[g/f] = E[m]$, $E[m^2] < \infty$, and $b_N \to 0$, clearly

$$\sqrt{N}(\tilde{\gamma}_{0,N} - E[\tilde{\gamma}_{0,N}]) \quad = \quad \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [m(l_i + \Delta) - m(l_i - \Delta)] I_i + o_p(1).$$

To tackle $\tilde{\gamma}_{1,N}$ and $\tilde{\gamma}_{2,N}$, note that they can be written as U statistics,

$$U_a \quad = \quad \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j=i+1}^{N} p_{a,N}(v_i, v_j)$$

where $U_a = \tilde{\gamma}_{a,N}$, $a = 1, 2$, and the symmetric kernels $(p_{a,N}(v_i, v_j) = p_{a,N}(v_j, v_i))$ are defined as

$$\begin{aligned}
p_{1,N}(v_i, v_j) \quad &= \quad \frac{1}{2h^k} K\left(\frac{l_i + \Delta - l_j}{h}\right) \frac{y_j}{f(l_i + \Delta)} I_i \\
&\quad - \frac{1}{2h^k} K\left(\frac{l_i - \Delta - l_j}{h}\right) \frac{y_j}{f(l_i - \Delta)} I_i \\
&\quad + \frac{1}{2h^k} K\left(\frac{l_j + \Delta - l_i}{h}\right) \frac{y_i}{f(l_j + \Delta)} I_j \\
&\quad - \frac{1}{2h^k} K\left(\frac{l_j - \Delta - l_i}{h}\right) \frac{y_i}{f(l_j - \Delta)} I_j \\
&\equiv \quad \frac{1}{2} \sum_{a=1}^{4} q_{a,N}(v_i, v_j),
\end{aligned}$$

$$\begin{aligned} p_{2,N}(v_i, v_j) &= \frac{1}{2h^k} K\left(\frac{l_i + \Delta - l_j}{h}\right) \frac{m(l_i + \Delta)}{f(l_i + \Delta)} I_i \\ &\quad - \frac{1}{2h^k} K\left(\frac{l_i - \Delta - l_j}{h}\right) \frac{m(l_i - \Delta)}{f(l_i - \Delta)} I_i \\ &\quad + \frac{1}{2h^k} K\left(\frac{l_j + \Delta - l_i}{h}\right) \frac{m(l_j + \Delta)}{f(l_j + \Delta)} I_j \\ &\quad - \frac{1}{2h^k} K\left(\frac{l_j - \Delta - l_i}{h}\right) \frac{m(l_j - \Delta)}{f(l_j - \Delta)} I_j \\ &\equiv \frac{1}{2} \sum_{a=1}^{4} q_{4+a,N}(v_i, v_j). \end{aligned}$$

From Lemma 3.1 in Powell et al. (1989), if $E[||p_{a,N}||^2] = o(N)$, then $\sqrt{N}(\tilde{\gamma}_{a,N} - E[\tilde{\gamma}_{a,N}])$ has the same asymptotic distribution as $(1/\sqrt{N}) \sum \{\rho_{a,N}(v_i) - E[\rho_{a,N}(v_i)]\}$, where $\rho_{a,N}(v_i) = E[p_{a,N}(v_i, v_j)|v_i]$. Given that $E[y^2|l]$ and $E[m(l)^2]$ are bounded, the result follows by bounding terms of the form

$$\int \frac{1}{h^{2k}} K\left(\frac{l_i + \Delta - l_j}{h}\right)^2 \frac{f(l_i) f(l_j)}{f(l_i + \Delta)^2}\, dl_i dl_j.$$

By a change of variables, (l_i, l_j) to $(l_i, u = (l_i + \Delta - l_j)/h)$, the above equation reduces to

$$\int \frac{1}{h^k} K(u)^2 \frac{f(l_i) f(l_i + \Delta - hu)}{f(l_i + \Delta)^2}\, dl_i du,$$

which is, by assumption, of order $O(h^{-k}) = O[N(Nh^k)^{-1}] = o(N)$.

The final result needed in this step is to show that

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{\rho_{a,N}(v_i) - E[\rho_{a,N}(v_i)]\} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{\rho_a(v_i) - E[\rho_a(v_i)]\} \\ &\quad + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{\tau_{a,N}(v_i) - E[\tau_{a,N}(v_i)]\}, \end{aligned}$$

where

$$\begin{aligned}
\rho_1(v_i) &= [m(l_i+\Delta) - m(l_i-\Delta)]1\{L+\Delta \le l_i \le U-\Delta\} \\
&\quad + \frac{y_i f(l_i-\Delta)}{f(l_i)} 1\{l_i \ge L+2\Delta\} \\
&\quad - \frac{y_i f(l_i+\Delta)}{f(l_i)} 1\{l_i \le U-2\Delta\}, \\
\rho_2(v_i) &= [m(l_i+\Delta) - m(l_i-\Delta)]1\{L+\Delta \le l_i \le U-\Delta\} \\
&\quad + \frac{m(l_i-\Delta) f(l_i-\Delta)}{f(l_i)} 1\{l_i \ge L+2\Delta\} \\
&\quad - \frac{m(l_i+\Delta) f(l_i+\Delta)}{f(l_i)} 1\{l_i \le U-2\Delta\},
\end{aligned}$$

and $(1/\sqrt{N})\sum\{\tau_{a,N}(v_i) - E[\tau_{a,N}(v_i)]\} = o_p(1)$. To show this, we work off the terms $s_{a,N}(v_i) = E[q_{a,N}(v_i, v_j)|v_i]$, $a = 1, \ldots, 8$; that is $\sum_{a=1}^4 s_{a,N} = \rho_{1,N}$, $\sum_{a=5}^8 s_{a,N} = \rho_{2,N}$, and analogously define $\sum_{a=1}^4 s_a = \rho_1$, $\sum_{a=5}^8 s_a = \rho_2$, $\sum_{a=1}^4 t_{a,N} = \tau_{1,N}$, and $\sum_{a=5}^8 t_{a,N} = \tau_{2,N}$. Finally observe that $s(l, y)$ as defined in Theorem 1 is

$$s(l, y) = [m(l+\Delta) - m(l-\Delta)]I + \rho_1(v) - \rho_2(v).$$

We will only present the details for $s_{1,N}$ and $s_{3,N}$ because the analysis for $s_{2,N}$, $s_{a,N}$, $a = 4, \ldots, 8$, follows analogously.

To tackle $s_{1,N}(v_i)$, note that

$$\begin{aligned}
s_{1,N}(v_i) &= E\left[\frac{1}{h^k} K\left(\frac{l_i+\Delta-l_j}{h}\right) \frac{y_j}{f(l_i+\Delta)} I_i \,\middle|\, v_i\right] \\
&= \frac{I_i}{h^k f(l_i+\Delta)} \int K\left(\frac{l_i+\Delta-l}{h}\right) m(l) f(l)\, dl \\
&= \frac{I_i}{f(l_i+\Delta)} \int K(u) m(l_i+\Delta-hu) f(l_i+\Delta-hu)\, du \\
&= m(l_i+\Delta)I_i + \frac{I_i}{f(l_i+\Delta)} \int K(u) \\
&\qquad \times [m(l_i+\Delta-hu) f(l_i+\Delta-hu) - m(l_i+\Delta) f(l_i+\Delta)]\, du \\
&= s_1(v_i) + t_{1,N}(v_i).
\end{aligned}$$

The Lipschitz condition imposed on f and m imply that $t_{1,N}(v_i)$ is $o(1)$.

To tackle $s_{3,N}(v_i)$, note that

$$\begin{aligned}
s_{3,N}(v_i) &= E\left[\frac{1}{h^k}K\left(\frac{l_j+\Delta-l_i}{h}\right)\frac{y_i}{f(l_j+\Delta)}I_j \,\middle|\, v_i\right] \\
&= \frac{y_i}{h^k}\int_L^U K\left(\frac{l_j+\Delta-l_i}{h}\right)\frac{f(l)}{f(l+\Delta)}\,dl \\
&= y_i\int_{(L+2\Delta-l_i)/h}^{(U-l_i)/h} K(u)\frac{f(hu-\Delta+l_i)}{f(hu+l_i)}\,du \\
&= y_i\frac{f(l_i-\Delta)}{f(l_i)}1\{l_i \geq L+2\Delta\} \;+\; y_i\int_{(L+2\Delta-l_i)/h}^{(U-l_i)/h} K(u) \\
&\quad \times\left[\frac{f(hu-\Delta+l_i)f(l_i) \;-\; f(hu+l_i)f(-\Delta+l_i)}{f(hu+l_i)f(l_i)}\right]du \\
&= s_3(v_i) \;+\; t_{3,N}(v_i).
\end{aligned}$$

Observe that $t_{3,N}(v_i)$ has second moment that is is bounded above by

$$h^2E[y^2]\left[\int |u|\,|K(u)|\,du\right]^2 \;=\; o(1),$$

by definition of an integrable boundary kernel, implying that it converges to zero in probability. Finally note that $(1/\sqrt{N})\sum \rho_a(v_i)$ is a sum of i.i.d. random variables with $E[\rho_a]^2 < \infty$; hence the central limit theorem applies.

Step 3: Asymptotic Bias. The bias is controlled by using the standard manipulations well known in the kernel literature. For example, given l_1, the conditional expectation of the first term in $\tilde{\gamma}_{1,N}$ is

$$\begin{aligned}
&\frac{1}{h^k}\int K\left(\frac{l_1+\Delta-l_2}{h}\right)\frac{m(l_2)}{f(l_1+\Delta)}f(l_2)I_1\,dl_2 \;= \\
&\quad\int K(u)\frac{m(l_1+\Delta-uh)}{f(l_1+\Delta)}f(l_1+\Delta-uh)I_1\,du.
\end{aligned}$$

To determine the rate at which the above expectation tends to $m(l_1 + \Delta)$, take a mean value expansion of $m(l + \Delta - uh)f(l + \Delta - uh)$ of appropriate order to get a bias of order $h^p \int |u^p K(u)| \, du$. The expectation of this with respect to the distribution of l_1 is of order h^p by definition of the integrable boundary kernel.

Step 4: Trimming. We need to show that $(1/\sqrt{N})(\hat{\gamma}_N - \bar{\gamma}_N) = o_p(1)$, which follows because

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} \frac{\hat{g}_i(l_i+\Delta)}{\hat{f}_i(l_i+\Delta)}(\hat{I}_i - I_i) &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\hat{g}_i(l_i+\Delta)}{\hat{f}_i(l_i+\Delta)} - \frac{g_i(l_i+\Delta)}{f_i(l_i+\Delta)}\right](\hat{I}_i - I_i) \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\frac{g_i(l_i+\Delta)}{f_i(l_i+\Delta)}(\hat{I}_i - I_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}\{\frac{g_i(l_i+\Delta)[\hat{f}_i(l_i+\Delta) - f_i(l_i+\Delta)]}{\hat{f}_i(l_i+\Delta)f_i(l_i+\Delta)} \\
&\quad + \frac{f_i(l_i+\Delta)[g_i(l_i+\Delta) - \hat{g}_i(l_i+\Delta)]}{\hat{f}_i(l_i+\Delta)f_i(l_i+\Delta)}\}(\hat{I}_i - I_i) \\
&\quad + o_p(N^{-1/2}) \\
&= o_p(N^{-1/2}),
\end{aligned}
$$

by applications of Lemma 1.

Q.E.D.

B Boundary Kernels

As outlined in Section 2, we need to adjust the kernels so that the integrals in (2.13) and (2.14) are zero for points near the boundary of the support. Formally, for any $l \in [L, U]$, the kernel $K(\cdot)$ with support $[-c_1, c_2]$, $c_1 = \min\{(U - l)/h, 1\}$, $c_2 = \min\{(l - L)/h, 1\}$, is called a *boundary kernel* of order (p,q) if

$$\int_{-c_1}^{c_2} K(u)u^j\,du = \begin{cases} 0, & j=0,\dots,p-1,p+1,\dots,q-1, \\ (-1)^p p!, & j=p, \\ \neq 0, & j=q. \end{cases}$$

Note that in general for points near the boundary of the support, the functional form of the kernel K will vary, depending on the proximity of l to either L or U. When we want to explicitly denote this variation, we will write $K(u;c_1(l),c_2(l))$ for $K(u)$ (for notational simplicity, in the other sections of this paper, this dependency will be implicitly assumed as necessary). Making note of this variation is necessary for several bounds in step 2 in Appendix A, and it is also necessary for controlling the bias as in step 3. As a result, we require stronger conditions on the boundary kernel. That is, a boundary kernel of order (p,q) is called an *integrable boundary kernel* of order (p,q) if for any $x \in [L,U]$,

$$\int |u|^p |K(u;c_1(u+x),c_2(u+x))|\,du \quad < \quad \infty,$$

and $K(u;c_1(l),c_2(l))$ is uniformly bounded for all l and u.

In general, there are an infinite number of choices for boundary kernels given any p and q. We outline the $(0,4)$ and $(0,2)$ boundary kernels below using the methodology in Zhang and Karuhamuni (2000). There are two motivating factors for using boundary kernels of this form: they are based on the optimal kernel (e.g. see Gasser, Mueller, and Mammitzsch (1985)) and they can be written in closed form.

$(0,4)$ *Kernel*: Given a bandwidth parameter h, if $h+L \le l \le -h+U$, then

$$K(u) \quad = \quad \begin{cases} \frac{15}{32}(3-10u^2+7u^4), & \text{if } -1<u<1, \\ 0, & \text{otherwise;} \end{cases}$$

if $l < h + L$,

$$K(u) = \begin{cases} \frac{20}{(1+c)^8}[1 + 10u + 30u^2 + 35u^3 + 14u^4 \\ - 12c(1 + 10u + 30u^2 + 35u^3 + 14u^4) \\ + 6c^2(13 + 80u + 165u^2 + 140u^3 + 42u^4) \\ - 4c^3(41 + 185u + 270u^2 + 140u^3 + 14u^4) \\ + 15c^4(11 + 32u + 28u^2 + 7u^3) \\ - 60c^5(1+u)^2 \; + \; 10c^6(1+u)], & \text{if } -1 < u \leq c, \\ 0, & \text{otherwise,} \end{cases}$$

where $c = (l - L)/h$; and if $l > -h + U$,

$$K(u) = \begin{cases} \frac{-20}{(1+c)^8}[1 + 10u + 30u^2 + 35u^3 + 14u^4 \\ - 12c(-1 + 10u - 30u^2 + 35u^3 - 14u^4) \\ - 6c^2(13 - 80u + 165u^2 - 140u^3 + 42u^4) \\ + 4c^3(41 - 185u + 270u^2 - 140u^3 + 14u^4) \\ - 15c^4(11 - 32u + 28u^2 - 7u^3) \\ + 60c^5(-1+u)^2 \; + \; 10c^6(-1+u)], & \text{if } -c \leq u < 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $c = (-l + U)/h$. The form of the kernel when $l < h + L$ is taken from Zhang and Karuhamuni (2000, p. 206). To derive the kernel when $l > -h + U$, we use their methodology outlined on pp. 204–205 with, in their notation,

$$S^* = \begin{pmatrix} 1 + c & (1 - c^2)/2 & \cdots & 1 \\ (1 - c^2)/2 & (1 + c^3)/2 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ (1 - (-c)^{k+1})/(k+1) & (1 - (-c)^{k+2})/(k+2) & \cdots & 1 \end{pmatrix}.$$

It is interesting to note that when $c = 1$, the second and third representations for the kernel reduce to the first, which is the optimal fourth order kernel. In addition, $K'(\cdot)$ is a $(1,4)$ boundary kernel. Also, since the kernels are uniformly bounded and continuous in terms of c, they also are integrable boundary kernels.

$(0,2)$ *Kernel*: In the case when points are from the boundary, the kernel reduces to the Epanechnikov kernel, $K(u) = 0.75(1-u^2)$. When $l < h + L$, the kernel is

$$K(u) = \frac{6}{(1+c)^4}(1+u)(1+2u-2c(1+2u)+3c^2),$$

and

$$K(u) = \frac{-6}{(1+c)^4}(-1+u)(1-2u+2c(-1+2u)+3c^2)$$

when $l > -h + U$.

Finally, as in Zhang and Karuhamuni (2000, p. 210) , we used a varying bandwidth parameter, $(2-c)h$, when $l < h + L$ or $l > -h + U$.

References

Abowd, J. and Card, D. (1987). Intertemporal Labor Supply and Long Term Employment Contracts. *American Economic Review*, *77*, 50–68.

Altonji, J. (1986). Intertemporal Substitution in Labor Supply: Evidence from Micro Data. *Journal of Political Economy*, *94*, 176–215.

Altug, S. and Miller, R. (1990). Household Choices in Equilibrium. *Econometrica*, *58 (3)*, 543–570.

Arellano, M. and Meghir, C. (1992). Female Labor Supply and on the Job Search: An Empirical Analysis Using Contemporary Data Sources. *Review of Economic Studies*, *59*, 537–559.

Bhattacharya, P. K. (1967). Estimation of a Probability Density Function and its Derivative. *The Indian Journal of Statistics: Series A*, *29*, 373–383.

Bierens, H. (1987). Kernel Estimators of Regression Functions. In *Advances in Econometrics-Fifth World Congress*. Cambridge University Press.

Blundell, R., Duncan, A., and Meghir, C. (1998). Estimating Labor Supply Responses Using Tax Reforms. *Econometrica*, *66*, 827–862.

Blundell, R. and Walker, I. (1986). A Life-Cycle Consistent Empirical Model of Family Labor supply Using Cross-Section Data. *Review of Economic Studies*, *53*, 539–558.

Browning, M., Deaton, A., and Irish, M. (1985). A Profitable Approach to Labor Supply and Commodity Demands over The Life-Cycle. *Econometrica*, *53*, 503–543.

Card, D. (1990). Labor Supply with a Minimum Hours Threshold. In *Studies in Labor Economics in Honor of Walter Y. Oi*, Carnegie-Rochester Conference Series on Public Policy, pp. 137–168. North Holland.

Coppejans, M. (2000). Effective Nonparametric Estimation in the Case of Severely Discretized Data. Working Paper.

Ferrall, C. and Shearer, B. (1999). Incentives and Transaction Costs with the Firm: Estimating an Agency Model using Payroll Data. *Review of Economic Studies*, *66(2)*, 309–338.

Gasser, T., Mueller, H. G., and Mammitzsch, V. (1985). Kernels for Nonparametric Curve Estimation. *Journal of the Royal Statistical Society Series B*, *47*, 238–252.

Grossman, S. and Hart, O. (1983). An Analysis of the Principle-Agent Model. *Econometrica*, *51*, 7–46.

Härdle, W., Hart, J., Marron, J. S., and Tsybakov, A. B. (1992). Bandwidth Choice for Average Derivative Estimation. *Journal of the American Statistical Association*, *87*, 218–226.

Härdle, W. and Stoker, T. (1989). Investigation of Smooth Multiple Regressions by the Method of Average Derivatives. *Journal of the American Statistical Association*, *84*, 986–995.

Haubrick, J. (1994). Risk Aversion, Performance Pay, and the Principle Agent Problem. *Journal of Political Economy, 102*, 258–276.

Heckman, J. and MaCurdy, T. (1980). A Life Cycle Model of Female Labour Supply. *Review of Economic Studies, 47*, 47–74.

Holmstrom, B. (1979). Moral Hazard and Observability. *Bell Journal of Economics, 19*, 74–91.

Jensen, M. and Murphy, K. (1990). Performance Pay and Top Management Incentives. *Journal of Political Economy, 98*, 225–264.

Linton, O. (1997). Efficent Estimation of Additive Nonparametric Regression Models. *Biometrika, 84*, 469-473.

Loader, C. R. (1999). Bandwidth Selection: Classical or Plug-in. *Annals of Statistic, 27*, 415–438.

MaCurdy, T. (1981). An Empirical Model of Labor Supply in a Life-Cycle-Setting. *Journal of Political Economy, 89*, 1059–1085.

Margiotta, M. and Miller, R. (2000). Managerial Compensation and the Cost of Moral Hazard. *International Economic Review, 41 (3)*, 669–720.

Mroz, T. (1987). The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica, 55*, 765–800.

Newey, W., Powell, J., and Vella, F. (1999). Nonparametric Estimation of Triangular Simultaneous Equation Models. *Econometrica, 67*, 565–603.

Powell, J., Stock, J., and Stoker, T. (1989). Semiparametric Estimation of Index Coefficients. *Econometrica, 57(6)*, 1403–1430.

Powell, J., and Stoker, T. (1996). Optimal Bandwidth Choice for Density-Weighted Averages. *Journal of Econometrics, 75*, 291–316.

Robinson, P. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica, 56*, 931–954.

Roehrig, P. (1988). Conditions for Identification in Nonparametric and Parametric Models. *Econometrica, 56*, 433–447.

Rosen, S. (1985). Implicit Contracts: A Survey. *Journal of Economic Literature, 23*, 1144–1175.

Sieg, H. (2000). Estimating a Dynamic Model of Household Choices in the Presence of Income Taxation. *International Economic Review, 41 (3)*, 637–668.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

Zhang, S. and Karuhamuni, R. J. (2000). On Nonparametric Density Estimation at the Boundary. *Nonparametric Statistics, 12*, 197–221.