Essays on Semiparametric Estimation.

Kang-sun Lee
Louisiana State University and Agricultural & Mechanical College

ESSAYS ON SEMIPARAMETRIC ESTIMATION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Economics

by
Kang-Sun Lee
B.A., Dankook University, 1991
M.A., Dankook University, 1993
M.S., Louisiana State University, 1998
December 2001

UMI Number: 3042629

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my parents

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my major professor, Dr. R. Carter Hill, for his efforts and time to complete this dissertation. He was always willing to answer any questions and made writing the dissertation interesting and challenging to me. I extend man thanks to my committee members, Dr. Robert Newman, Dr. M. Dek Terrell, and Dr. James Geaghan for their encouragement and guidance. Without their help, I could not have finished the dissertation at all. Also, I would thank to the other faculty members of Dept. Economics for their encouragement and guidance throughout my stay at LSU.

Back in Korea, my deep gratitude is extended to my professors, Dong-Un Park, Hyung Chun, Seung-Joon Park, and Myung Hun Kang, who inspired me to study in the U.S. I also thank my friends at Dankook University in Seoul, Korea; Sung Ho Hong, Dong-Hyeon Kim, Moonseok Seo, Jong-Chan Noh and many.

I also would like to acknowledge the support I had received from my parents, Ki-eon Lee and Bum-Nam Choi, brothers and sister. I would like to thank my wife, Soonmi Joun, my daughter, Woojin Lee, and son, Jay Lee, for their patient support encouragement throughout these years. They have been a constant source of encouragement to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This dissertation is primary concerned on the study of semiparametric estimation approaches. In the respect of the usage of econometric analysis that is evaluating theoretical relationship, the semiparametric analysis is useful to get the flexibility of functional form. The kernel-type nonparametric methods are used for semiparametric approaches in this dissertation.

The first essay focuses upon performance of various bandwidth selector in the local linear regression method. The results indicate that the variable bandwidth selector is superior to constant bandwidth selector in the more skewed data set or complicated functional form. LSCV bandwidth selector fit well in the simple functional form. This essay also indicates that the variable bandwidth selector performs well in almost everywhere in general.

The second essay is the application of local linear regression method with variable bandwidth selector to the wage equation. The challenge for the quadratic or quartic relationship between log wage and experience recently make possible to apply the semiparametric estimation method to the wage equation. The comparison of semiparametric and parametric specifications indicates that semiparametric estimation methods capture nonlinearities in the earnings profiles. Also, the analysis of wage profile using semiparametric method confirms the stylized facts of U.S. earning profiles during the 1990's.

The semiparametric estimation method is applied to the qualitative response model in the third essay. The simultaneous two-stage probit model is studied using semiparametric method in the two-stage. Klein and Spady's semiparametric MLE is

applied in this essay. The Monte Carlo simulation results indicate that semiparametric method performs well in the both homoscedasticity and heteroscedasticity error terms. MSE of semiparametric estimation is smaller and steadier than two-stage probit estimation.

CHAPTER 1

INTRODUCTION

The main usage of econometric analysis is evaluating some theoretical relationship that is presumed to exist between the dependent and explanatory variables. For analysis some specific functional form is required, relating these variables and any others needed for control purposes, and the selection of a formula that will be provide the requisite information accurately. The linear relationship between response and explanatory variables is most prevalent. In the case of linear regression, a particular concern has been with the linearity of the functional form connecting the variables appearing in it. Sometimes there are scientific reasons for modeling response variable as a particular function of explanatory variables, while at other times the models is based on experience gained through analysis of previous data sets of the same type. However, there is a drawback to parametric modeling that needs to be considered. The restriction of functional form belonging to a parametric family means that functional form can sometimes be too rigid. For example, models may impose the functional form be a parabolic, periodic or monotone, each of which might be too restrictive for adequate estimation of the true regression function. If one chooses a parametric family that is not of appropriate form, at least approximately, then there is a danger of reaching incorrect conclusions from the regression analysis.

In general, if the error term in the regression models is well behaved, then the ordinary least squares (OLS) estimator retains the property of being best, linear, and unbiased estimators (BLUE) conditional upon explanatory variables. However, there is no obvious reason why linear relationship between dependent and independent variables

is an attractive property for an estimator once we deviate from the normality distributed data paradigm. It might be anticipated that the best estimator would be a nonlinear function, and so OLS is unlikely to be the preferred choice. Therefore choice of a best estimator of a parameter is of growing interests when the error term is not normal, particularly with the growth of cross-section data sets that seem to exhibit severe nonnormality. One way to proceed would be to endow the random variable with some density and to then devise an optimal estimator under this specification by maximum likelihood estimation (MLE) methods. However, if the density chosen for error term is incorrect, proceeding in this fashion might do more harms than is incurred by invalidity acting as if the disturbances were normally distributed.

The rigidity of parametric regression can be overcome by removing the restriction that functional forms belong to a parametric family. This approach leads to what is commonly referred to as nonparametric (NP) regression. The motivation for a nonparametric approach to regression is straightforward: when confronted with a scatterplot showing no discernible simple functional form then one would want to let the data decide which function fits them best without the restrictions impose by a parametric model. This is sometimes referred to as "letting the data speak for themselves".

There now exist many methods for obtaining a nonparametric regression estimate of functional form. However, the major complication in a nonparametric approach to estimation is the 'curse of dimensionality'.[1] Very large samples are needed

[1] The general smoothing parameterization of the kernel estimator in higher dimensions requires the specification of many more bandwidth parameters than in he univariate setting. This leads us to consider simpler smoothing parameterizations as well. Also, the sparseness of data in higher-dimensional space makes kernel smoothing difficult unless the sample size is very large. This phenomenon is called curse

if an accurate measurement of the function is to be made. The size of sample required increases rapidly with the number of variables involved in any relation. Such a feature leads to the proposition that one might well prefer to restrict some variables to have a linear impact while allowing a much smaller number to have a nonlinear one. For example, the wage is regarded as being influenced by the individual's personal characteristics as well as the number of years of job experience, whereas the impact of the personal characteristics is taken to be linear, that for experience is nonlinear. The estimation that involves a combination of parametric and nonparametric methods is called semiparametric (SP).

Semiparametric and nonparametric estimation methods have been employed in the estimation of many important econometric models.[2] A central feature of semiparametric and nonparametric estimation is the flexibility of functional form. Unlike parametric estimation methods, semiparametric and nonparametric estimation methods do not restrict the dimension of parameters for the some or all of the explanatory variables to easily capture the nonlinearity of the functional form.

Since semiparametric estimation is the hybrid of parametric and nonparametric methods, it has the advantage that parametric part will reduce the 'curse of dimensionality' and nonparametric part will capture the nonlinearity without any other functional form. Semiparametric estimation is done in two stages. The first step consists of the usual nonparametric regression of dependent variable (y) on nonlinear part explanatory variable (z), and linear part independent variables (x) on nonlinear part

---

dimensionality. It means that, with practical sample sizes, reasonable nonparametric density estimation is difficult in more than about five dimensions (Wand and Jones, 1995).

[2] Pagan and Ullah (1999) offer various theory and applications of non- and semi-parametric estimation methods. Fomby and Hill (2000) shows applications for the various economic phenomena.

explanatory variable (z), respectively. The second step is the ordinary least squares (OLS) regression of the residuals of the former regression on those of the latter to obtain the 'semiparametric estimates' of coefficient value of linear part variable ($\hat{\beta}$). To get the nonparametric conditional estimation we run the nonparametric regression of residuals of OLS ($y - x\hat{\beta}$) on z.

As noted in the above explanation, the semiparametric estimation method includes a nonparametric estimation procedure, and we need to understand nonparametric estimator properties. There are several approaches to the nonparametric method.[3] The general method of nonparametric estimation is based on a kernel-type smoothing estimator. We have three questions to decide when using nonparametric or semiparametric estimation methods.

First, which smoothing method do we use? One of the kernel type nonparametric estimators used broadly is the local polynomial regression estimator. The advantage of the local polynomial estimator is that can be analyzed with standard regression techniques. It also has the same first-order statistical properties irrespective of whether the explanatory variables are stochastic or nonstochastic. Further, there is an absence of boundary effects: the bias at the boundary stays automatically of the same order as in the interior, without use of specific boundary kernels. And, unlike most other methods, the local polynomial approximation method does not require knowledge of the location of the endpoints of the support.

[3] Hart (1997) displays some methods: Local averaging approach, Kernel smoothing, Fourier series estimators, Local polynomials estimator, Smoothing Splines, Rational Functions estimators, and Wavelets approximations.

Second, which kernel do we use? Usually, the kernel will be a symmetric probability density function. Silverman (1986) showed that there is very little difference to choose between the various kernels on the basis of mean integrated square error (MISE).[4]

The third fundamental question that remains in regards to any method of nonparametric estimation is the choice of the associated smoothing parameter (so called bandwidth, h). When the bandwidth is too small, the resulting curve is too wiggly, reflecting too much of the sampling variability. When the bandwidth is too large, the resulting estimate tends to smooth away important features of the underlying density. To determinate the bandwidth one might work with the mean integrated squared error (*MISE*). *Cross-validation* methods are frequently performed in this case by minimizing the estimated prediction error *EPE*, $n^{-1}\sum(y_i - \hat{m}(x_i))^2$, with respect to h, where $\hat{m}_i = \hat{m}(x_i)$ is computed as the 'leave-one-out' estimator deleting the ith observation in the sums.

Seather (1992) and Park and Turlach (1992) compared several constant bandwidth selectors using simulated and real data sets, separately.[5] They found that there is no best bandwidth selector that works in all cases. One, however, should expect that the window width should be larger when trying to estimate the tails of a density

[4] There are some different kernels: Epanechnikov kernel, Biweight kernel, Triangular kernel, Gaussian kernel, and rectangular kernel. The all above kernels are symmetric function satisfying $\int K(t)dt = 1$, $\int tK(t)dt = 0$, and $\int t^2K(t)dt = k \neq 0$, where k is constant.

[5] The six bandwidths selection methods are considered: 1. Least squares cross-validation, 2.Biased cross-validation, 3.The normal based rule of thumb, 4.The plug-in method of Park and Marron, 5.the plug-in method of Sheather and Jones, and 6.the root-n convergent method of Hall, Sheather, Jones and Marron. They found that *plug-in* methods performed well when the data has the several modes as well as one-mode and usually least squares *cross-validation* performed undersmoothed. But when the data has the

than in its center, since there will be fewer observations available in the former situation. Moreover, since the density will be small and flat in the tails, it will not matter much that observations distant from x are employed. In contrast, when it is varying rapidly, as in the central part of the density, incorrect estimates are likely to lead to undersmoothing in some part of the range and oversmoothing in another. A procedure that responds to this observation is the variable bandwidth density estimation. Although 'plug-in' estimators of bandwidth work well in the situation with density estimation, this 'plug-in' estimator has not been a great deal of merit for the nonparametric regression estimation (Pagan & Ullah (1999), and M.J. Lee (1996)).

Recently the usage of semiparametric estimation is broadening to limited dependent variable models as well as continuous variable models. For example, Ichimura (1993) and Lewbel (2000) considered semiparametric estimators based on the regression principles, and Klein and Spady (1993), and Gozalo and Linton (1994) used semiparametric estimators based on 'maximum likelihood' principles. Also, Klein and Spady's estimator attains the semiparametric efficiency bound as well as $\sqrt{n}$ consistentcy.

Given the above semiparametric estimation method, the purpose of this dissertation is to investigate the bandwidth selection methods and to apply the semiparametric estimation to the continuous and discrete data sets. In Chapter 2 of this dissertation briefly summarizes the theory of nonparametric and semiparametric estimation methods. We will focus about the kernel-type nonparametric estimations

skewed and long tail, none of them do quite fit the data well, since a global bandwidth fixed across the entire range of the data is not at all suited.

and their properties in this chapter. The general properties of kernel-type nonparametric and semiparametric estimation methods are considered.

In Chapter 3, we investigate the bandwidth selection rules. The choice of the bandwidth parameter is rather crucial and hence this should be done with a lot of care. There is little literature comparing constant and variable bandwidth directly. Zhang and Lee (2000) compared constant and variable bandwidth with normal distribution data sets only. We compare the variable bandwidth selection rules with a constant bandwidth selection rule in the local linear regression estimation through the three different situations: uniform, normal, and gamma distributed data sets. In the Monte Carlo experiment, the variables are generated from uniform, normal, and gamma distributions. The constant bandwidth selection rule is followed by Silverman's (1986) 'rule of thumb (ROT)' bandwidth, and least squares cross validation (LSCV). For simplicity, we use just two constant bandwidth rules among many of constant bandwidth selection methods since these two bandwidths represent that both are generally smallest bandwidth (LSCV) and largest bandwidth (ROT). Variable bandwidth selection rule come from Fan and Gijbel's (1992, 1995) global variable bandwidth selection rule.

In Chapter 4, we apply the local linear regression with variable bandwidth, which is investigated in Chapter 3, to the wage equation. In general, wages are a function of schooling of individual, individual's job experience, and other personal characteristics. The wage equation is known as a linear form of schooling and personal characteristics and a nonlinear form of job experience, which is quadratic or quartic functional form. Pudney (1993) and Ginther (2000), and Zheng (2000) applied the

nonparametric estimation to the wage equation. They found that nonparametric estimation methods capture nonlinearities in the earning profiles. Bound and Johnson (1992) investigated U.S. wage structure using parametric method during 1980's. In this chapter we will use semiparametric estimation method to study U.S. wage structure during 1990's. The first part of this chapter compares the parametric estimation and semiparametric estimation for the wage equation. The second part is the analysis of the reason of change of the U.S. wage structure during 1990's.

In Chapter 5, semiparametric estimation method is extended to the qualitative model analysis. Ichimura (1993), Klein and Spady (1993), and Lewbel (2000) investigated the single equation qualitative model. When there exist two or more equations and the variables are correlated, the model is complicated due to a coherence condition problem. Simultaneous probit model in which the binary endogenous variables appear among the explanatory variables is investigated. Lee (1995) investigated the simultaneous equation with limited dependent variable using Tobit model. In this chapter we investigate simultaneous probit model using the two-stage maximum likelihood (ML) estimation with semiparametric method.

CHAPTER 2

NONPARAMETRIC AND SEMIPARAMETRIC ESTIMATION

2.1. Nonparametric Regression Estimation

It is of common interest to explore the association between the covariate and the response in the case of bivariate observations. Nonparametric (NP thereafter) estimation is a flexible estimation method that does not make any assumption on the form of this function.

Consider the general statistical model:

(2.1.1) $$y = m(x) + u,$$

where the error term u has the properties $E(u|x) = 0$ and $E(u^2|x) = \sigma^2(x)$. In here, y is the dependent variable and x is a vector of regressors; these (x,y) variables are taken to be completely characterized by their unknown joint density $f(y,x)$, at the points (y,x). Suppose we have data (y_i, x_i) $(i = 1,\ldots,n)$ upon the random variables y and x, where y_i is a scalar and x_i is a vector of variables.

There are some assumptions:

A1. m and f are twice continuously differentiable in a neighborhood of the point x.

A2. The kernel K is a symmetric function satisfying:

$$\text{(i)} \int K(\psi)d\psi = 1, \quad \text{(ii)} \int \psi K(\psi)d\psi = 0, \quad \text{(iii)} \int \psi^2 K(\psi)d\psi = \mu_2 < \infty,$$

where $\psi = (\frac{x_i - x}{h})$, and h is bandwidth.

A3. $h = h_n \to 0 \quad nh \to \infty \quad \text{as} \quad n \to \infty$.

A4. x_i is i.i.d. and independent of the u_is.

A5. The second order derivatives of the marginal density of f of x_i are continuous and bounded in a neighborhood of x, and x is a point in the interior of the support of x_i.

Under the above assumptions, a general class of NP estimators of $m(x)$ can be written as

$$\tilde{m} = \tilde{m}(x) = \sum_{i=1}^{n} w_{ni}(x) y_i \tag{2.1.2}$$

where $w_{ni}(x) = w_n(x_i, x)$ represents the weight assigned to the i^{th} observation y_i, and it depends on the distance of x_i from the point x. Usually, the weight is high if the distance between x and x_i is small and low if the distance is large.

The weighted least squares criterion for the NP estimator of $m(x)$ is

$$\sum_{i=1}^{n} w^{*}_{ni}(x)(y_i - m(x))^2 \tag{2.1.3}$$

The resulting estimate, $\tilde{m}$, of $m(x)$ is precisely (2.1.2), after writing $w_{ni}(x)$ = $w^{*}_{n}(x)/\sum_{i=1}^{n} w^{*}_{ni}(x)$. Various NP estimators are special cases of (2.1.3), differing mainly with respect to the choice of w_{ni}.[1]

2.1.1. Local Linear Regression Estimation

One of the advanced kernel-type NP estimators used broadly is the local linear regression estimator. In recent years Fan (1992), Fan and Gijbels (1994) and Ruppert and Wand (1995) have extensively investigated this local linear regression estimator suggested by Stone (1977) and Cleveland (1979).

The logic of local linear regression smoothing can be seen by expanding $m(x_i)$ of equation (2.1.1) around x to get

---

[1] Nadaraya-Watson (NW thereafter) Kernel Estimator, Recursive Kernel Estimator, Fixed Design Estimators, k Nearest Neighbor Estimators, Spline smoothing, and the Local Linear and Polynomial Regression Estimators.

(2.1.4) $$m(x_i) = m(x) + \frac{\partial m}{\partial x}(x^*)(x_i - x) = m(x) + \beta(x^*)(x_i - x),$$

where x^* lies between x_i and x. The local linear regression estimator minimizes $\sum_{i=1}^{n}\{y_i - m - \beta(x_i - x)\}^2 K\left(\frac{x_i - x}{h}\right)$ with respect to m and β, where $K(\cdot)$ is a kernel and h is a bandwidth that determines how many of the x_is around x are to be used in forming the average. This form is the residual sum of squares from a regression using only observations close to $x_i = x$. This can be expended to the j^{th} order polynomial in $(x_i - x)$ under the assumption of existence of the derivatives $m^{(j)}$ which is local polynomial regression. This local linear regression estimator has the form of weighted least squares regression of y_i against $z_i' = (1, (x_i - x))$ with weights $K_i^{1/2}$,

(2.1.5) $$\tilde{m}(x) = \sum_{i=1}^{n} w_n(x) y_i,$$

with weights $w_n = e'_1 \left(\sum z_i K_i z'_i\right)^{-1} z_i K_i$, where e_1 is a column vector of dimension the same as z_i' with unity as first element and zero elsewhere (for example, $e'_1 = \{1,0\}$ for the univariate case).

The advantage of the local linear estimator is that the standard regression techniques can be used to analyze. It also has the same first-order statistical properties irrespective of whether the x_i's are stochastic or nonstochastic. Furthermore, the local linear estimator has automatic boundary modification. For estimating $m(x_0)$, with x_0 a point close to the boundary, the local neighborhood $x_0 \pm h$ can lie outside the design region. Hence, certain symmetric moment conditions which are valid for all interior points are no longer valid for x_0 in a boundary region by causing a large boundary bias

for most of the smoothing techniques. This problem is referred to as boundary effects. Since many of the smoothing techniques show this bias problem at the boundary, considerable efforts have been devoted to developing methods for correcting this boundary bias, such as boundary kernel methods and reflection methods. The local linear or polynomial estimators adapt automatically to estimation at the boundaries, and the use of local linear or polynomial fitting is more efficient than other methods. Fan and Gijbels (1992) proved that theoretically the local linear regression estimator adapts automatically to the boundary.

In comparison with the Nadaraya-Watson (NW) estimator of $m(x) = m$, which minimizes $\sum_{i=1}^{n}\{y_i - m\}^2 K\left(\frac{x_i - x}{h}\right)$ with respect to m and hence the solution of m as $\hat{m}(x) = K\left(\frac{x_i - x}{h}\right)y_i \Big/ \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$, local linear fit locally uses one parameter more without increasing the asymptotic variance. This extra parameter enables the local linear fit to reduce the bias. The NW estimator suffers from large bias at the boundary region and at the region where the derivative of the regression function or of the design density is large, but this estimator is not. Some proposals such as boundary kernel methods and reflection methods are proposed to handle this problem, but Fan and Gijbels (1992) shows that they are less efficient than the automatic boundary correction of the local linear fit. The reason of this comparison is that the broadly used estimator is NW estimator and NW estimator is one of the local polynomial regressions.

2.1.2. Properties of Local Linear Regression Estimation

From the rewriting (2.1.5), the estimator of $m(x)$ is

(2.1.6) $$\hat{m}(x) = e'_1 \left(\sum z_i w_i z'_i\right)^{-1} \sum z_i w_i y_i,$$

with $w_i = K\left(\frac{x_i - x}{h}\right)$. Since

$$(2.1.7)\qquad \begin{aligned} y_i &= m(x_i) + u_i \approx m(x) + (x_i - x)\beta(x) + (x_i - x)^2\gamma(x^*) + u_i \\ &= z'_i\,\delta + (x_i - x)^2\gamma(x^*) + u_i \end{aligned}$$

where $\beta(x) = m^{(1)}(x)$, $\gamma(x^*) = m^{(2)}(x^*), \delta' = (m(x)\ \beta(x))$, and $m^{(j)}(x)$ denotes j^{th} derivatives, we can find an alternative expression for $\hat{m}(x)$ by substituting for y_i from equation (2.1.6).

$$(2.1.8)\qquad \begin{aligned} \hat{m}(x) &= e'_1\,\delta(x) + e'_1\left(\sum z_i w_i z'_i\right)^{-1}\sum z_i w_i (x_i - x)^2\gamma(x^*) \\ &\quad + e'_1\left(\sum z_i w_i z'_i\right)^{-1}\sum z_i w_i u_i \end{aligned}$$

Because $e'_1\,\delta(x) = m(x)$, the conditional bias and variance are

$$(2.1.9)\qquad E_X\left(\hat{m}(x) - m(x)\right) = e'_1\left(\sum z_i w_i z'_i\right)^{-1}\sum z_i w_i (x_i - x)^2\gamma(x^*)$$

and

$$(2.1.10)\qquad V_X\left(\hat{m}(x)\right) = \sigma^2 e'_1\left(\sum z_i w_i z'_i\right)^{-1}\left(\sum z_i w_i z'_i\right)\left(\sum z_i w_i z'_i\right)^{-1} e_1 .$$

For large n, this expression is evaluated by using the asymptotic results to get[2]

$$\left((nh)^{-1}\sum z_i w_i z'_i\right)^{-1} \xrightarrow{p} \begin{bmatrix} f^{-1}(x) & -f^{(1)}(x)f(x)^{-2} \\ -f^{(1)}(x)f(x)^{-2} & \left\{\mu_2 f(x)h^2\right\}^{-1} \end{bmatrix},$$

$$\left((nh)^{-1}\sum z_i w_i^2 z'_i\right)^{-1} \xrightarrow{p} \begin{bmatrix} f(x)\int K^2(\psi)d\psi & hf(x)\int K^2(\psi)\psi d\psi \\ hf(x)\int K^2(\psi)\psi d\psi & h^2 f(x)\int K^2(\psi)\psi^2 d\psi \end{bmatrix}$$

where $\mu_2 = \int \psi^2 K(\psi)d\psi$ and we use $(nh)^{-1}\sum_{i=1}^{n} K^2(\psi_i)\psi_i^2 = f\int K^2(\psi)\psi^2 d\psi + o_p(1)$.[3]

[2] Ruppert and Wand (1994) proved theses properties.

[3] The small o and large O denote the order of magnitudes. The sequence $\{X_n\}$ of random variables (r.v.) is said to be at most of order n^k and denoted by $\{X_n\} = O_p(n^k)$, if $\frac{X_n}{n^k} - c_n \to 0$ as $n \to \infty$, where c_n is a nonstochastic sequence. Also $\{X_n\}$ is said to be smaller order than n^k and denoted by $\{X_n\} = o_p(n^k)$, if $\frac{X_n}{n^k} \to 0$.

Using these results give the asymptotic bias and variance of $\hat{m}$. The asymptotic bias variance of the local linear regression estimator of $m(x)$ are

(2.1.11) $$\text{Bias}(\hat{m}(x)) = \frac{1}{2}\mu_2 h^2 m^{(2)}(x),$$

(2.1.12) $$V(\hat{m}(x)) = \sigma^2 \frac{(nh)^{-1}}{f(x)} \int K^2(\psi)d\psi.$$

Also, under the assumptions A3 and A4, and two more additional assumption[4], $(nh)^{1/2}(\hat{m} - E_X(\hat{m}))$ which is the kernel-type NP estimator is asymptotically distributed as

(2.1.13) $$N\left(0, f^{-1}(x)\sigma^2 \int K(\psi)^2 d\psi\right)$$

These moments of local linear regression estimator have some interesting aspects compared to other estimators. First, the variance is the same but the bias is different with the NW kernel estimator. Second, Fan and Gijbels (1992) show that its bias and variance are of the same order of magnitude in both the interior and near the boundary of the support of f. Third, the fact that the bias is of order h^2 does not derive from the symmetry of the kernel. Fourth, the bias of the NW kernel estimator is large if either $|m^{(1)}|$ or $|f^{(1)}/f|$ is large, but neither term appears in (2.1.11). Fifth, when $m(x)$ is linear, the bias in the local linear estimator vanishes. Finally, Fan (1992) suggests that it is generally the case that the local linear regression estimator has smaller MSE than the other kernel estimators, and this turns out to be particularly true around the boundary points.

---

[4] (A6). K be the any Borel measurable, bounded, real-valued functions $K(\psi)$ such that (i) $\int K(\psi)d\psi = 1$, (ii) $\int |K(\psi)|\, d\psi < \infty$, (iii) $|\psi||K(\psi)| \to 0$ as $|\psi| \to \infty$, (iv) $\sup|K(\psi)| < \infty$, and (v) $\int K^2(\psi)d\psi < \infty$. (A7). $E|u_i|^{2+\delta} < \infty$ and $\int |K(\psi)|^{2+\delta} d\psi$ for some $\delta > 0$ hold.

2.2. Semiparametric Regression Estimation

2.2.1. General Estimation Method

To describe the semiparametric (SP hereafter) procedure, suppose that the i^{th} observation is given by a $(1+p+k)\times 1$ vector (y_i, x'_i, z'_i), $i = 1, .., n$, which is generated by the model

(2.2.1) $$y_i = f(x_i) + g(z_i, \beta) + u_i,$$

where $f(x)$ is an arbitrary function of x, while $g(z,\beta)$ is a known parametric function of z and a vector of unknown parameters β. The disturbance term u_i is assumed to satisfy

(2.2.2) $$E(u_i \mid x_i, z_i) = 0.$$

The most popular functional form of $g(\bullet)$ is linear, i.e.

(2.2.3) $$g(z,\beta) = z'\beta.$$

The parameter of interest is β so that the issue is how to estimate it in the presence of the unknown function f. Taking the conditional expectation of (2.2.1) with (2.2.3) gives the results $E(y_i \mid x_i) = E(z_i \mid x_i)'\beta + f(x_i)$.

Therefore,

(2.2.4) $$y_i - E(y_i \mid x_i) = (z_i - E(z_i \mid x_i))'\beta + u_i,$$

and

(2.2.5) $$f(x_i) = E(y_i \mid x_i) - E(z_i \mid x_i)'\beta.$$

Since (2.2.4) has the properties of a linear regression model with dependent variable $y_i - E(y_i \mid x_i)$ and independent variables $(z_i - E(z_i \mid x_i))$, an estimator of β is

(2.2.6) $$\hat{\beta} = \left[\sum_{i=1}^{n}(z_i - \hat{m}_{zi})(z_i - \hat{m}_{zi})'\right]^{-1}\left[\sum_{i=1}^{n}(z_i - \hat{m}_{zi})(y_i - \hat{m}_{yi})\right]$$

where $\hat{m}_{zi}$ and $\hat{m}_{yi}$ are the kernel-type nonparametric estimators of $m_{zi} = E(z_i \mid x_i)$ and $m_{yi} = E(y_i \mid x_i)$ respectively. Once $\hat{\beta}$ is found, then $f(x_i)$ can be estimated from (2.2.5) as $\hat{f}(x_i) = \hat{m}_{yi} - \hat{m}'_{zi}\hat{\beta}$.

2.2.2. Properties of the Linear Part in the Semiparametric Estimation

Two important questions arise over many of the SP estimators. First, what is the asymptotic distribution of this estimator? Second, how efficient is it relative to an estimator that used $g(z,\beta)$? In the (2.2.4), the estimator of β is OLS on that equation, and the SP estimator $\hat{\beta}$ could be viewed as an analogous estimator in which the unknown quantities m_{zi} and m_{yi} were replaced by their nonparametric estimators $\hat{m}_{zi}$ and $\hat{m}_{yi}$.

Rewriting the equation (2.2.4) with $\eta_i = z_i - E(z_i \mid x_i) = z_i - m_{zi}$;

(2.2.7) $$y_i = m_{yi} + \eta_i\beta + u_i = m_{yi} + \hat{\eta}_i\beta + (\eta_i - \hat{\eta}_i)\beta + u_i.$$

Substituting (2.2.7) into (2.2.6) and simplifying gives

(2.2.8) $$n^{1/2}(\hat{\beta} - \beta) = \left(n^{-1}\sum \hat{\eta}_i^2\right)^{-1}\left[n^{-1/2}\left(\sum \hat{\eta}_i((m_{yi} - \hat{m}_{yi}) + (\eta_i - \hat{\eta}_i)\beta + u_i\right)\right],$$

where $\hat{\eta}_i$ can be represented by $\hat{\eta}_i = \eta_i + m_{zi} - \hat{m}_{zi}$. Assume that the moments of η_i and u_i up to fourth order are bounded, $E(\eta_i u_i) = 0$, and a CLT[5] and LLN[6] applies to $n^{-1/2}\sum \eta_i u_i$ and $n^{-1}\sum \eta_i \eta'_i$, denominator of (2.2.8) converges to $V_{\eta\eta}$. By the CLT,

[5] Lindberg-Levy Theorem: Let $\{X_i\}$ be a sequence of i.i.d. r.v.s such that $EX_i = \mu$, $V(X_i) = \sigma^2 < \infty$. Then $\sqrt{n}\dfrac{(X_i - \mu)}{\sigma} \xrightarrow{d} N(0,1)$.

[6] Chebychev's Theorem: Let the r.v. be s.t. $EX_i = \mu_i, V(X_i) = \sigma_i^2 < \infty$, and cov($X_i, X_j$), $i \neq j$, then $\bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$ if $\bar{\sigma}^2 \to 0$ as $n \to \infty$, where $\bar{X}_n = n^{-1}\sum X_i$, $\bar{\mu}_n = n^{-1}\sum \mu_i$, and $\bar{\sigma}^2 = n^{-1}\sum \sigma_i^2$

the numerator converges to the respectively. By the LLN, $n^{-1}\sum E(\hat{\eta}_i^2) = V_{\eta\eta}$, where $V_{\eta\eta} = E(\eta_i \eta'_i)$, which means same limiting distribution as $n^{-1/2}\sum \eta_i u_i$. Consequently, the asymptotic distribution of $n^{1/2}(\hat{\beta} - \beta)$ is $N\left(0, \sigma^2 V_{\eta\eta}^{-1}\right)$.

Not just any consistent estimator of the conditional means will work to give $n^{1/2}$ consistency for $(\hat{\beta} - \beta)$. If the conditional moments could be estimated parametrically then $n^{1/2}(\hat{m}_{yn} - m_{yn})$ would be $O_p(1)$ and consistency is straightforward, but the convergence rate of nonparametric estimators is slower than $n^{1/2}$ and depends on the bandwidth and the number of conditioning variables. Even though the kernel estimate of the conditional means converge to their true values quite slowly, the fact that these are used in a regression indicates that the values are effectively being averaged, and it is this feature that makes it possible for the estimator of β to exhibit the same convergence rate as in a parametric model.

2.3. Kernel Choice

In the nonparametric estimation the choice of kernel density is not a crucial problem. Almost always K is taken to be a symmetric function around zero satisfying,

$$
\begin{aligned}
&(i) \int K(t) = 1, \\
(2.3.1) \qquad &(ii) \int t^2 K(t)dt = \mu_2 \neq 0, \\
&(iii) \int K^2(t)dt \rightarrow \infty,
\end{aligned}
$$

which is based on simplicity of interpretation. But there is theoretical reason for insisting that K be symmetric density estimator. Cline (1988) said a kernel K is *admissible* if the only kernel K_1 satisfying $MISE(\hat{f}_{n.K_1}) \leq MISE(\hat{f}_{n.K})$, for all n and f,

is $K_1 = K$. The admissible kernel is characterized as those have nonnegative Fourier transforms bounded by 1, and thus admissibility implies symmetry.

To make a choice bandwidth or kernel it is necessary to have some criterion. By far the most popular strategies have been to either minimize $\int[\hat{f}(x)-f(x)]^2dx$, the Integrated Squared Error (ISE), or minimize $\int E[\hat{f}(x)-f(x)]^2dx$, the Integrated Mean Squared Error (MISE). These correspond to loss and risk respectively. The first criterion depends on the data whereas the second should not. Since the expression of MISE is difficult to obtain, generally researchers use an approximation to MISE, AMISE, is expressed:

$$\begin{aligned} &AMISE\{\hat{f}(\cdot;h)\} = C(K_{\delta_0})\{(nh)^{-1} + \tfrac{1}{4}h^4R(f'')\} \\ &where \\ &C(K) = \{R(K)^4\mu_2(K)^2\}^{1/5}, R(K) = \int K(t)^2dt \end{aligned} \tag{2.3.2}$$

where h is bandwidth and this $C(K)$ is invariant to rescaling of K. Wand and Jones (1995) show the derivation of this form.

The problem of determining the optimal kernel shape is to choose K to minimizes $C(K)$. Because of the scale invariance of $C(K)$, the optimal K is the one that minimize $C(K)$ subject to

$$\int K(t)dt = 1,\ \int tK(t)dt = 0,\ \int t^2K(t)dt = \mu_2 < \infty \text{ and } K(t) \geq 0 \text{ for all } t.$$

The solution can be shown to be

$$K^a(t) = \tfrac{3}{4}\{1 - t^2/(5a^2)\}/(5^{1/2}a)1_{\{|t|<5^{1/2}a\}}, \tag{2.3.3}$$

where a is an arbitrary scale parameter. The simplest version of $K^a(t)$ corresponds to $a^2 = 1/5$, which yields the Epanechnikov Kernel

(2.3.4) $$K^*(t) = \tfrac{3}{4}\left(1-t^2\right)1_{\{|t|<1\}}.$$

Consider the efficiency of any symmetric kernel K by comparing it with the Epanechnikov kernel. Define the efficiency of K to be

(2.3.5) $$eff(k) = \{C(K_e)/C(K)\}^{5/4} = \frac{3}{5\sqrt{5}}\left\{\int t^2K(t)dt\right\}^{-1/2}\left\{\int K(t)^2dt\right\}^{-1},$$

where $K_e(t)$ is Epanechnikov kernel. The reason this ratio is called the efficiency of K relative to K^* is that it represents the ratio of sample sizes necessary to obtain the same minimum AMISE (for a given f) when using K^* as when using K. Table 2.1 shows that there is very little to choose among the various kernels on the basis of mean integrated square error (Silverman (1986) and Wand and Jones (1995)).

2.4. Bandwidth Choice

The fundamental question that remains in regards to any method of nonparametric estimation is the choice of the associated smoothing parameter (the so called bandwidth, h). When h is too small, the resulting curve is too wiggly, reflecting too much of the sampling variability. When h is too large, the resulting estimate tends to smooth away important features of the underlying density. In practice, h must be set to achieve the best possible trade-off between bias and variance. Like optimal kernel choice, to determinate the bandwidth one might work with the integrated mean squared error (MISE). The optimal bandwidth is defined as the value of h that minimizes MISE. The problem is how to compute an estimator of this quantity, since the expression for exact MISE is difficult to obtain. This generally leads to some approximation to MISE, AMISE. To get this, we use the formulas for asymptotic bias and variance given in (2.1.11) and (2.1.12).

Table 2.1. Some Kernels and Their Efficiencies

Kernel	$K(t)$	*Efficiency*
Epanechnikov	$\frac{3}{4}(1-\frac{1}{5}t^2)/\sqrt{5}$ for $\lvert t\rvert < \sqrt{5}$ 0 otherwise	1.00
Biweight	$\frac{15}{16}(1-t^2)^2$ for $\lvert t\rvert < 1$ 0 otherwise	.994
Triweight	$1 - \lvert t\rvert$ for $\lvert t\rvert < 1$ 0 otherwise	.987
Normal	$\frac{1}{\sqrt{2\pi}}e^{-(1/2)t^2}$	.951
Triangular	$(1-\lvert t\rvert)1_{\{\lvert t\rvert<1\}}$	.936
Uniform	$\frac{1}{2}1_{\{\lvert t\rvert<1\}}$	.930

The integrated mean squared error is

(2.4.1) $$MISE = \int[(bias(m(x))^2 + V(m(x))]dx.$$

The AMISE for the local linear regression is

(2.4.2)
$$\begin{aligned} AMISE &= \frac{1}{4}\mu_2^2 h^4 \int (m^{(2)}(x))^2 dx + (nh)^{-1}\sigma^2 \frac{1}{\int f(x)dx}\int K^2(\psi)d\psi \\ &= \lambda_0 (nh)^{-1} + \frac{1}{4}\lambda_1 h^4 \end{aligned}$$

where $\lambda_0 = \sigma^2 \int K^2(\psi)d\psi$, and $\lambda_1 = \mu_2^2 \int (m^{(2)}(x))^2 dx$.

Differentiate this AMISE with respect to h, and equate the outcome to zero, and we obtain

(2.4.3) $$h^3\lambda_1 - \lambda_0 n^{-1} h^{-2} = 0, \text{ or } h = \left(\frac{\lambda_0}{\lambda_1}\right)^{1/5} n^{-1/5} = cn^{-1/5}.$$

Thus the value of h for which the bias and variance are of the same order of magnitude is $h \propto n^{-1/5}$. Silverman (1986) uses the direct calculation of density based on a specific density to plug in the magnitude of c. For example, if Gaussian kernel is being used

and $f(x) \sim N(\mu, \sigma^2)$, the bandwidth is given by $h = 1.06\sigma n^{-1/5}$, which is the Rule-of-Thumb bandwidth.

Alternatively, Cross-validation (CV) methods are frequently performed in this case by minimizing the estimated prediction error (EPE), $n^{-1}\sum(y_i - \hat{m}(x_i))^2$, with respect to *h*, where $\hat{m}_i = \hat{m}(x_i)$ is computed as the 'leave-one-out' estimator deleting the i^{th} observation in the sums. Denote $\hat{m}_h$ as any estimate of the regression function $m(.)$. The least squares cross-validation technique uses the weighted average of squared errors

$$(2.4.4) \qquad n^{-1}\sum_{i=1}^{n}\{y_i - \hat{m}_{h,-i}(x_i)\}^2 K\left(\frac{x_i - x}{h_n}\right),$$

where $\hat{m}_{h,-i}(x_i)$ denote a regression function which is built from the data $\{(x_i, y_i), j \neq i\}$ and h_n is a sequence of positive bandwidths. The least squares criterion estimates

$$(2.4.5) \qquad n^{-1}\sum_{i=1}^{n}\{m(x_i) - \hat{m}_h(x_i)\}^2 K\left(\frac{x_i - x}{h_n}\right) + n^{-1}\sum_{i=1}^{n}\sigma^2(x_i).$$

The first term is a discrete approximation of the weighted integrated squared error, with weight function $K(\cdot)$, and the second term is independent of h_n. The least squares cross-validation selector is the one that minimizes (2.4.5).

The constant bandwidth depends neither on the location of x nor on that of the data x_i. Such an estimator does not fully incorporate the information provided by the density of the data points. Furthermore, a constant bandwidth is not flexible enough for estimating curves with a complicated shape. All these considerations lead to introducing a variable bandwidth $h_n / a(x_i)$, where $a(\cdot)$ is some nonnegative function

reflecting the variable amount of smoothing at each data point. The weighted least squares criterion for the nonparametric estimation of $m(x)$ is

$$(2.4.6) \qquad \sum_{i=1}^{n}\{y_i - m(x_i)\}^2 K\left(\frac{x_i - x}{h_n} a(x_i)\right),$$

which is same with equation (2.1.4) except variable term $a(\cdot)$.

This variable bandwidth has some advantages. It gives a certain flexibility in smoothing various types of regression functions. Optimization over all possible variable bandwidths leads to an optimal bandwidth and hence improves this performance. It will be seen that for an optimal choice of $a(\cdot)$ this function is proportional to $f_x^{1/5}$, where f is marginal density of x, and this is precisely how an ideal variable kernel smoother should behave. With a particular choice of the variable bandwidth, the estimator will have a homogeneous variance, i.e., independent of the location point x.

For the local linear regression with variable bandwidth, Fan and Gijbels (1992) studied the theoretical choice and the asymptotic properties. The form of the estimator for the local linear smoother is

$$(2.4.7) \qquad \sum_{i=1}^{n}\{y_i - m(x) - \beta(x_i - x)\}^2 a(x_i) K\left(\frac{x_i - x}{h_n} a(x_i)\right) .$$

Then the regression estimator is defined as

$$\hat{m}(x) = \sum_{i=1}^{n} w_i y_i \Big/ \sum_{i=1}^{n} w_i,$$

$$(2.4.8) \qquad w_i = a(x_i) K\left(\frac{x_i - x}{h_n} a(x_i)\right)\{S_{n,2} - (x_i - x) S_{n,1}\},$$

$$\text{where } S_{n,j} = \sum_{i=1}^{n} a(x_i) K\left(\frac{x_i - x}{h_n} a(x_i)\right)(x_i - x)^j .$$

The conditional mean squared error (MSE) of the estimator is given by

$$(2.4.9) \qquad E[(\hat{m}(x)-m(x))^2 \mid X_1,\dots,X_n] = \left(b_n^2(x)+v_n^2(x)\right)\left(1+o_p(1)\right),$$

where $b_n(x)=\dfrac{1}{2}m''(x)\dfrac{s_2^2-s_1s_3}{s_2s_0-s_1^2}\left(\dfrac{h_n}{a(x)}\right)^2$, and

$$v_n^2(x)=\left(\frac{\int_{-\infty}^{+\infty}[s_2-\psi s_1]^2K^2(\psi)d\psi}{[s_2s_0-s_1^2]^2}\right)\frac{a(x)\sigma^2(x)}{f_x(x)nh_n},$$

with $s_l=\int_{-\infty}^{+\infty}K(\psi)\psi^l d\psi$, $l=0,1,2,3$ and $\psi=\left(\dfrac{x_i-x}{h_n}a(x_i)\right)$.

If $a(\cdot)$ = 1, the preceding result generalizes the known result for the estimator with a constant bandwidth. In case the kernel function K is a density with mean zero, and asymptotic expression for the conditional MISE is defined by

$$(2.4.10) \qquad AMISE(\hat{m},m)=\int_{-\infty}^{+\infty}\left[\frac{1}{4}\left(m''(x)s_2\left(\frac{h_n}{a(x)}\right)^2\right)^2+\frac{a(x)\sigma^2(x)}{f_x(x)nh_n}\int_{-\infty}^{+\infty}K^2(\psi)d\psi\right]W(x)dx,$$

where $W(x)$ denotes a weight function.

In order to find an optimal choice of the function $a(\cdot)$ we minimize the AMISE with respect to h. This yields the optimal constant bandwidth

$$(2.4.11) \qquad h_{n,a}=\left(\frac{\int_{-\infty}^{+\infty}a(x)\sigma^2(x)W(x)/f_x(x)dx\int_{-\infty}^{+\infty}K^2(\psi)d\psi}{s_2^2\int_{-\infty}^{+\infty}[m''(x)]^2W(x)/a^4(x)dx}\right)^{1/5}n^{-1/5}.$$

Substituting this optimal choice into (2.4.8) leads to

$$(2.4.12) \qquad AMISE(\hat{m},m)=\frac{5C_K}{4n^{4/5}}\left(\int_{-\infty}^{+\infty}[m''(x)]^2\frac{W(x)}{a^4(x)}dx\left[\int_{-\infty}^{+\infty}a(x)\sigma^2(x)\frac{W(x)}{f_x(x)}dx\right]^4\right)^{1/5}.$$

where $C_K=s_2^{2/5}\left[\int_{-\infty}^{+\infty}K^2(\psi)d\psi\right]^{4/5}$.

Now minimize (2.4.12) with respect to $a(\cdot)$. The solution to this optimization problem is established by (2.4.13),

$$\text{(2.4.13)} \qquad a_{opt}(x) = \begin{cases} b\left(\dfrac{f_x(x)[m''(x)]^2}{\sigma^2(x)}\right)^{1/5}, & \text{if } W(x) > 0, \\ a^*(x), & \text{if } W(x) = 0, \end{cases}$$

where b is any arbitrarily positive constant and $a^*(x)$ can be taken to be any positive value. With the preceding optimal choice of $a(\cdot)$, the optimal constant bandwidth $h_{n,a}$ in (2.4.11) is equal to

$$\text{(2.4.14)} \qquad h_{n,opt} = b\left(\frac{\int_{-\infty}^{+\infty} K^2(\psi)d\psi}{s_2^2}\right)^{1/5} n^{-1/5}.$$

With these optimal choices of the constant and the variable bandwidth, the AMISE is given by

$$\text{(2.4.15)} \qquad AMISE_{v,opt} = \frac{5C_K}{4n^{4/5}}\left(\int_{-\infty}^{+\infty}[m''(x)]^{2/5}\left[\frac{\sigma^2(x)}{f_x(x)}\right]^{4/5} W(x)dx\right).$$

On the other hand, the expression for the AMISE with $a(\cdot) = 1$ and an optimal choice of the constant bandwidth is

$$\text{(2.4.16)} \qquad AMISE_{c,opt} = \frac{5C_K}{4n^{4/5}}\left(\int_{-\infty}^{+\infty}[m''(x)]^2 W(x)dx \int_{-\infty}^{+\infty}\left[\sigma^2(x)\frac{W(x)}{f_x(x)}dx\right]^4\right)^{1/5}.$$

It is easy to see that $AMISE_{v,opt} \le AMISE_{c,opt}$ and this fact reflects one of the advantages of using a variable bandwidth. The concept of variable bandwidth is intuitively appealing: A different amount of smoothing is used at different data locations. The optimal variable bandwidth a_{opt} depends on $f_x(\cdot)$, σ^2 and m'' only through a 1/5 power

function. The intuitive choice for the variable bandwidth is $a(x) = f(x)/\sigma^2(x)$. This choice implies that a large bandwidth is used at low-density design points and also at locations with large conditional variance. With such a variable bandwidth, the regression smoother has a homogeneous variance.

2.5. Summary

The non- and semi-parametric estimation is widely used in the applied econometrics. Since the semiparametric estimation is one of the hybrids from parametric and nonparametric, it is important to study the nonparametric estimation procedure. In this chapter we have investigated the non- and semi-parametric estimators, their properties, choice of optimal kernel, and some bandwidth selection rules. Unlike most of kernel-type nonparametric estimators, local linear regression estimator has some advantages, i.e., it does not have the problem of "boundary effects" and it has the standard regression format. Furthermore, variable bandwidth will give certain flexibility in smoothing various types of regression functions. From the selection of kernel type selection Epanechnikov kernel will gives the optimal kernel shape. We will focus on the performance of the local linear smoother with variable bandwidth in this paper.

CHAPTER 3

PERFORMANCE OF BANDWIDTH SELECTION RULES FOR THE LOCAL LINEAR REGRESSION

3.1. Introduction

Local linear regression estimation uses a random sample $\{x_i, y_i\}$ $i = 1, \ldots, n$ to estimate the curve $\hat{y}(x) = \hat{m}(x)$ by minimizing

(3.1.1) $$\sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1(x_i - x)\}^2 K_h(x_i - x),$$

where $K_h(x_i - x) = K[(x_i - x)/h]$, K is called the kernel function and h is called the bandwidth. If the regression function $m(x)$ is approximated locally by a linear Taylor's expansion in a neighborhood of x, then the local linear regression estimator performs a weighted regression of y_i against $z'_i = (1, (x_i - x))$ using $w_i^{1/2} = \{K[(x_i - x)/h]\}^{1/2}$ weight. The local linear regression estimator is obtained by fitting local straight lines. An interesting collection of effective data analysis carried out by this simple and intuitive estimator is given in Fan and Gijbels (1996). Like every kernel-type estimator, the bandwidth selection in the local linear regression estimation is important. When h is too small, the resulting curve is too wiggly, reflecting too much of the sampling variability. When h is too large, the resulting estimate tends to smooth away important features. For this reason, data-driven choice of h has been a key issue of the kernel type nonparametric estimation. The general criterion of the bandwidth selection is Mean Integrated Squared Error defined by;

$$MISE = \int MSE(x)dx,$$

where $MSE = E[\{\hat{m}(x) - m(x)\}^2 \mid x]$ and $x = (x_1, \ldots, x_n)$.

Seather (1992) and Park and Turlach (1992) compared several constant bandwidth selectors using simulated and real data sets for density estimation, separately. They found that *plug-in* methods performed well when the data has the several modes as well as one-mode and usually least squares *cross-validation* undersmoothed. But when the data has the skewed and long tail, none of them fit the data well, since a global bandwidth fixed across the entire range of the data is not at all suited. They said that there is no best bandwidth selector that works in all cases. Although 'plug-in' estimators of h work well in the situation with density estimation, this 'plug-in' estimator does not have a great deal of merit for the conditional moment estimation (Hardle (1990), Pagan & Ullah (1999), and M.J. Lee (1996)). A procedure that responds to this observation is variable bandwidth estimation.

A variable bandwidth is introduced to allow for different degrees of smoothing by Brieman, Meisel and Prucell (1977), resulting in a possible reduction of estimation bias at peaked regions and a reduction of the variance at flat regions. This enhances the flexibility of the local polynomial fitting, so that it can adapt to spatially non-homogeneous curves. Fan and Gijbels (1992, 1995) used the variable bandwidth for the local linear smoothers and they argued that the variable bandwidth has theoretical advantages. Zhang and Lee (2000) showed that the Mean Integrated Squared Error (MISE) of variable bandwidth is much smaller than the cross-validation method and the theoretical optimal constant bandwidth. Lee and Solo (1999) studied bandwidth selection for the local linear regression with constant bandwidth selectors. Although they suggested the two new simple selectors, the least-squares cross-validation method performed better than other selectors generally.

The one thing we consider here is that the empirical performance has been judged using only the uniform or normal distributions for covariates. Sheather (1992) and Park and Turlach (1992) used the mixture of normal densities for kernel density estimation, Fan and Gijbels (1992, 1995) and Zhang and Lee (2000) used a normal or an uniform distribution for the covariates. Such choices do not represent all real situations formed with real data. For example, in economics, where estimating an equation measuring wages, the worker's experience seems to be approximated well by the log-normal distribution, i.e. highly skewed to the right-hand side.

Our goal of this chapter is the comparison of some bandwidth selection rules using different distributions of covariates. The selected bandwidth methods are rule-of-thumb method, least squares cross-validation constant bandwidth and variable bandwidth estimator. This chapter is organized as follows. In the next section, we briefly introduced three bandwidth methods. The simulation study is provided in section 3 and section 4 will gives some summary.

3.2. Some Bandwidth Selection Rules

3.2.1. Rule-of-Thumb Bandwidth

In many data analyses, one would like to get a quick idea about how large the amount of smoothing should be. A "rule-of-thumb (ROT)" bandwidth selection is very suitable in such a case. Such a rule is meant to be somewhat crude, but possesses simplicity and requires little programming effort that other methods are hard to compete with. Pudney (1993) and Ginther (1999) used the ROT bandwidth selection method for their empirical studies.

With the local polynomial regression method such a crude bandwidth selector can easily be obtained as follows. Consider the asymptotically optimal constant bandwidth, which come from minimizes the asymptotic weighted Mean Integrated square error (WMISE)

$$(3.2.1)\qquad \begin{aligned} WMISE &= E[\int (\hat{m}(x) - m(x))^2 w(x)dx] \\ &= \int ([Bias\{\hat{m}(x)\}]^2 + Var\{\hat{m}(x)\})w(x)dx \end{aligned}$$

where

$$Bias\{\hat{m}(x_0)\} = \{\int x^2 K(x)dx\}\frac{1}{(2)!}m''h + o(h)$$

$$Var\{\hat{m}(x_0)\} = \int K^2(x)dx \frac{\sigma^2(x_0)}{f(x_0)nh} + o(\frac{1}{nh})$$

with $w \geq 0$ some weight function, leads to a theoretical optimal constant bandwidth. Using the asymptotic expression of conditional bias and variance of local linear regression estimator, an asymptotically optimal constant bandwidth is

$$(3.2.2)\qquad h_{opt} = n^{-1/5} C(K)\left[\frac{\int \sigma^2(x)w(x)/f(x)dx}{\int \{m''(x)\}^2 w(x)dx}\right],$$

where $C(K)$ is some constant values, m'' is the second derivative function estimation, and $f(x)$ is density function of x. Fan and Gijbels (1995) give $C(K) = 2.719$ when the function $m(\cdot)$itself is estimated with local linear regression. It contains the unknown quantities $\sigma^2(\cdot)$, $m''(\cdot)$ and $f(\cdot)$, which need to be estimated. The "Rule of Thumb" bandwidth fits a polynomial of order 4 globally to $m(x)$ by the parametric fit

$$(3.2.3)\qquad \hat{m}(x) = \hat{\alpha}_0 + \ldots + \hat{\alpha}_4 x^4.$$

The standardized residual sum of squares from this parametric fit is denoted by $\hat{\sigma}^2$. Substitute the estimated value for the equation (3.2.2), and then we can obtain the rule of thumb bandwidth selector

$$\text{(3.2.4)} \qquad \hat{h}_{ROT} = C(K)\left[\frac{\hat{\sigma}^2 \int w(x)dx}{\sum_{i=1}^{n}\{\hat{m}''(x_i)\}^2 w(x_i)}\right]^{1/5}.$$

3.2.2. Cross-Validation Bandwidth Selection Rule

The most widely studied bandwidth selector is least squares cross-validation (LSCV), proposed by Rudemo (1982) and Bowman (1984). There are some of applications of this method; e.g. Stock (1989), McMillen and Thorsnes (1999), Iwata et al. (1999), and Zheng (1999). The basic idea behind this cross-validation (CV) procedure is to choose h by minimizing the Integrated Squared Error (ISE) defined by $ISE = \int \{\hat{m}(x) - m(x)\}^2 dx$. Let $\hat{m}_h(\cdot)$ denote any estimate, involving a smoothing parameter h, of the regression function $m(\cdot)$. For each given i, we use data $\{(x_j, y_j),\ j \neq i\}$ to build a regression function $\hat{m}_{h,-i}(\cdot)$ and then validate the model by examining the prediction error $y_i - \hat{m}_{h,-i}(x_i)$. The least squares cross-validation technique uses the weighted average of squared errors

$$\text{(3.2.5)} \qquad CV(h) = n^{-1}\sum_{i=1}^{n}\{y_i - \hat{m}_{h,-i}(x_i)\}^2 w(x_i),$$

as an overall measure of the effectiveness of the estimation scheme $\hat{m}_{h,-i}(\cdot)$ where $w(x_i)$ is some positive function. From (3.2.5), the expression $\hat{m}_{h,-i}(x_i)$ is the 'leave-one-out' estimator of (3.2.1) omitting the i^{th} observation. The least squares cross-validation

bandwidth selector is the one that minimizes (3.2.5). The method to find the minimum of (3.2.5) is the grid search method. Find the all of CV for the grid sets of h values. The least squares cross-validation bandwidth is

$$\hat{h}_{cvls} = \arg\min_{h}[CV(h)]. \quad (3.2.6)$$

3.2.3. Variable Bandwidth Selection Rule

The concept of the variable bandwidth was introduced by Breiman, Meisel and Prucell (1977) in the density estimation context. Instead of (3.2.1), the local linear regression estimator is obtained by minimizing

$$\sum_{i=1}^{n}\{y_i - \beta_0 - \beta_1(x_i - x)\}^2 \alpha(x_i) K_h(\frac{x_i - x}{h}\alpha(x_i)), \quad (3.2.7)$$

with respect to β_0 and β_1, where $\alpha(\cdot)$ is some nonnegative function reflecting the variable amount of smoothing at each data point. The optimal variable bandwidth is the same method that minimizes WMISE with respect to h, except that the variable bandwidth has the varying term $\alpha(\cdot)$ to be chosen. Fan and Gijbels (1992) suggested that an optimal choice of $\alpha(\cdot)$ is proportional to $f_x^{1/5}(\cdot)$, where f_x is marginal distribution of x, and this is precisely how an ideal variable kernel smoother should behave. The optimal variable bandwidth is defined by

$$\hat{h}_v = h_{opt} / \alpha(x_i) = h_{opt} / f_x^{1/5}, \quad (3.2.8)$$

where h_{opt} is the optimal constant bandwidth.

3.3. A Simulation Study

A simulation study is conducted to evaluate the practical performance of the proposed bandwidth schemes; Rule-of-Thumb bandwidth (ROT), Cross-Validation bandwidth and Variable bandwidth. Four test functions are used:

1: $m(x) = 0.4x + 1$ $\qquad x \in [0,1]$,

2: $m(x) = 0.3 + 4x - 3x^2$ $\qquad x \in [0,1]$,

3: $m(x) = x + 2\exp(-16x^2)$ $\qquad x \in [-2,2]$,

4: $m(x) = 1 + 48x - 180x^2 + 145x^4$ $\qquad x \in [0,1]$,

Let x and y be the two random variables whose relationship can be modeled as

(3.2.9) $$y = m(x) + \sigma(x)\varepsilon, \quad E(\varepsilon) = 0, \quad \mathrm{var}(\varepsilon) = 1,$$

where x and ε are independent.

The test functions 1 and 3 are used by Fan and Gijbels (1995), their covariates are generated from a normal distribution for test function 1 and from a uniform distribution for test function 3. The test functions 2 and 4 are quadratic and quartic functions that are chosen arbitrarily. The reason of choosing of the test functions 2 and 4 is that the quadratic and quartic functions are often used in econometric modeling, for example, the estimation of wage equation in labor economics. These four test functions are plotted in Figure 3.1.

Three signal-to-noise ratios (s/n) and three design densities were used. Here signal-to-noise is defined to be the variance of the function divided by the variance of the noise: $s/n = \mathrm{var}(m)/\sigma^2$. The three s/n were: low=2, medium=4 and high=8, and the three design densities were the uniform density, normal density, and gamma density. Normal random errors were used for all test function. For each test function, the distribution of error terms follow: $\sigma_1 = 0.15$, $\sigma_3 = 0.75$, $\sigma_4 = 0.25(\max m - \min m)$, and $\sigma_1 = 0.15$, where subscription denotes each of the test functions. We use sample sizes n = 200, and 400 and number of replications in the simulation is 1000. This

formatting is similar to the previous researchers' setup. In each of the examples we use the Epanechnikov kernel $K(u) = 0.75(|1-u^2|_+)$.

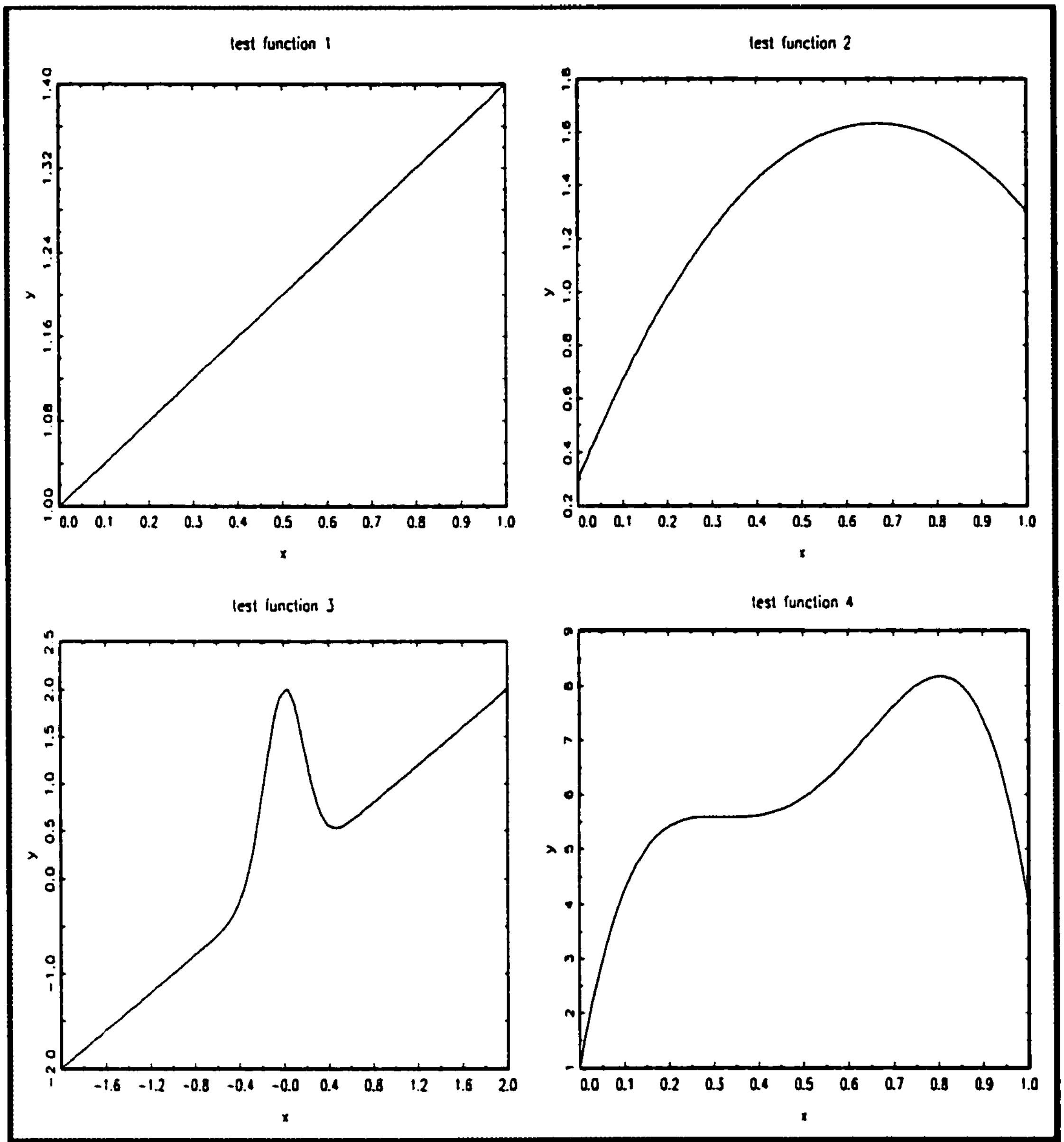


Figure 3.1. Plots of Test functions

For the least square cross-validation (LSCV) procedure, the estimated curves are evaluated in grid points x_j, $j = 1, \ldots, n_{grid}$. So the integral involved in the methodology are implemented as averages over appropriate grid points. The grid points

are used in arithmetic type, i.e. $h_i = C * h_{\min}$, where $h_{\min}$ denotes the first grid point and C is the grid span. We start from $h = h_{\min}$, keeping h by factor C and compute $M(h)$ at these geometric grid points. We stop when the function values $M(h)$ increase consecutively a certain number of times or when $h > h_{\max}$. Then we choose the minimizer of $M(h)$ as the grid point having the smallest computed $M(h)$ value. In our implementation we took $h_{\min} = (x_{(n)} - x_{(1)})/n$, $h_{\max} = (x_{(n)} - x_{(1)})/2$, and C = .1 where $h_{\min} = (x_{(n)} - x_{(1)})/n$, and $x_{(n)} = \max_i x_i$. For the variable bandwidth, we use the LSCV bandwidth for the pilot bandwidth, h_{opt} in equation (3.3.9), and the density function f_x based on this pilot bandwidth.

3.4. Results of the Simulation Study

We conduct a simulation study to evaluate and compare each of the bandwidth selectors. Tables below show WMISE of variable bandwidth and relative of efficiency of estimator of LSCV and ROT bandwidth selector for each test function for the different distribution covariates. The efficient ratio is computed similar to Fan's method (Fan, 1992):

$$rf_1 = \left(\frac{WMISE \text{ of the estimator with LSCV bandwidth}}{WMISE \text{ of the estimator with Variable bandwidth}} \right),$$

and

$$rf_2 = \left(\frac{WMISE \text{ of the estimator with ROT bandwidth}}{WMISE \text{ of the estimator with Variable bandwidth}} \right).$$

The weighted mean integrated squared error is defined in our simulation by:

$$WMISE = E[\int (\hat{m}(x) - m(x))^2 w(x) dx],$$

where $w(x)$ is sample density function from each bandwidth selectors. To get *WMISE*, we calculate weighted mean squared error, $E[(\hat{m}(x)-m(x))^2 w(x)]$, for each iteration, and sum them by number of iteration. The relative efficiency shows that if rf_i is greater than one, then the variable bandwidth is efficiency than other constant bandwidth rules, vice versa.

3.4.1. Uniform Random Design

Table 3.1 represents the uniform density design for three signal-to-noise ratios (s/n), low = 2, medium = 4 and high = 8. For the test function 1, the variable bandwidth estimator slightly dominates the LSCV bandwidth estimator in the efficiency respect. The relative efficiency ratios of LSCV (rf_1) are not significantly different from one. Also, variable bandwidth estimator dominates the ROT bandwidth estimator with an exception. ROT bandwidth estimator is significantly more efficient than variable bandwidth estimator, when $n = 400$ and s/n = 4.

For the test function 2, the relative efficiencies of LSCV and ROT bandwidth estimators for the variable bandwidth estimator range from about 1.15 to 2.75, and about 1.16 to 4.10, respectively. The variable bandwidth estimator strictly dominates the LSCV and ROT bandwidth estimators for test function 2 in the uniform design.

For the test function 3, variable bandwidth estimator and LSCV bandwidth estimator are not different statistically, although LSCV bandwidth estimator dominates variable bandwidth estimator. ROT bandwidth estimator has large of relative efficiency ratio to variable bandwidth (rf_1) for all designs, indicating the danger of using a fixed bandwidth when the function exhibits regions of rapid change. ROT bandwidth

estimator is not a good estimator for a 'humped shape' functional function, since the ROT bandwidth estimator fits usually over-smooth for the 'humped' part.

Table 3.1. *Relative Efficiency Ratios* of Bandwidth Selection for the Uniform Design

B/W	s/n	WMISE of Variable bandwidth		Relative Efficiency of LSCV (rf_1)		Relative Efficiency of ROT (rf_2)	
		N=200	N=400	N=200	N=400	N=200	N=400
Fn 1	2	5.86E-05	4.21E-04	1.22 (0.85)	1.38 (2.04)	1.48 (1.65)	2.71 (7.98)
	4	2.92E-05	1.33E-04	1.17 (0.85)	0.87 (-0.71)	1.54 (2.43)	0.66 (-2.26)
	8	1.77E-05	4.70E-05	1.35 (1.05)	1.32 (1.82)	5.16 (5.38)	2.98 (7.72)
Fn 2	2	0.01	0.04	1.49 (2.16)	2.29 (8.81)	1.46 (2.07)	4.10 (21.16)
	4	3.17E-03	0.04	1.15 (0.55)	2.34 (8.75)	1.16 (0.59)	3.11 (12.85)
	8	3.64E-03	0.02	2.10 (3.29)	2.75 (9.70)	1.58 (1.75)	1.32 (1.78)
Fn 3	2	0.03	0.06	0.90 (-0.31)	0.96 (-0.29)	476.70 (3.74)	1686.38 (6.74)
	4	0.01	0.03	0.91 (-0.33)	0.92 (-0.42)	1809.98 (4.02)	3141.65 (7.54)
	8	0.01	0.02	0.90 (-0.26)	0.93 (-0.52)	253.00 (3.49)	4122.80 (7.23)
Fn 4	2	1.40	0.26	1.46 (1.98)	1.23 (1.29)	0.96 (-0.12)	139.06 (787.45)
	4	0.80	13.71	1.31 (0.90)	1.19 (4.80)	0.87 (-0.39)	1.19 (0.83)
	8	0.63	7.64	1.92 (2.69)	2.38 (5.00)	0.70 (-0.88)	0.57 (-1.55)

NOTE: The numbers of parenthesis are t-value of $H_0: rf_i = 1, t = (rf_i - 1)/s.e.(rf_i)$, where $s.e(rf_i) = \sqrt{\mathrm{var}(WMISE_{var}) + \mathrm{var}(WMISE_j)}$, with i=1,2 and j = LSCV and ROT.

For the test function 4, variable bandwidth estimator is more efficient than LSCV bandwidth estimator. The relative efficiency gain of variable bandwidth estimator is statistically significant when s/n ratio is larger. The ROT bandwidth estimator acts very well on one exceptional case, when n = 400 and s/n = 2. When n = 200 ROT has greater efficiency relative to the variance bandwidth, but these differences

are not significant. For n = 400 and s/n = 2, the variable bandwidth estimator has a large relative efficiency. Usually the variable bandwidth and LSCV bandwidths under smooth. But in this low s/n cases the variable bandwidth rule does not yield a under smoothed fit. ROT bandwidth estimator depends on variance of error term $\sigma_4 = 0.25(\max m - \min m)$ which represents a wider bandwidth.

Table 3.2. *Relative Efficiency Ratios* of Bandwidth Selection for the Normal Design

B/W	S/n	WMISE of Variable bandwidth		Relative Efficiency of LSCV (rf_1)		Relative Efficiency of ROT (rf_2)	
		N=200	N=400	N=200	N=400	N=200	N=400
Fn 1	2	6.31E-05	1.68E-05	1.09 (0.30)	1.03 (0.35)	1.42 (1.64)	1.03 (0.37)
	4	5.01E-05	9.79E-05	1.14 (0.44)	1.03 (0.27)	3.66 (12.26)	1.28 (4.90)
	8	2.56E-05	8.56E-05	1.24 (0.60)	1.10 (0.51)	1.81 (1.85)	2.55 (5.00)
Fn 2	2	3.19E-04	0.03	0.73 (-1.38)	2.43 (6.65)	0.74 (-1.35)	2.93 (8.98)
	4	0.01	0.04	1.56 (1.80)	3.68 (13.36)	1.84 (2.70)	2.73 (8.61)
	8	2.38E-04	0.02	1.16 (0.40)	2.99 (12.06)	1.01 (0.02)	1.06 (0.35)
Fn 3	2	0.02	0.04	1.30 (0.64)	1.38 (1.77)	1961.44 (4.32)	6119.27 (9.63)
	4	0.15	0.02	2.37 (1.49)	1.21 (1.23)	52.66 (4.24)	14500.97 (8.59)
	8	0.09	1.87	2.99 (2.88)	3.32 (4.72)	150.75 (4.90)	97.56 (9.16)
Fn 4	2	1.03	0.26	1.33 (1.10)	1.25 (1.15)	1.00 (0.01)	29.08 (128.54)
	4	2.06	11.17	1.50 (1.15)	2.68 (7.78)	1.07 (0.17)	0.86 (-0.63)
	8	0.47	7.75	1.15 (0.48)	2.50 (5.87)	0.88 (-0.38)	0.43 (-2.23)

NOTE: The numbers of parenthesis are t-value of $H_0: rf_i = 1, t = (rf_i - 1)/s.e.(rf_i)$, where $s.e(rf_i) = \sqrt{\mathrm{var}(WMISE_{var}) + \mathrm{var}(WMISE_j)}$, with i=1,2 and j = LSCV and ROT.

The relative efficiency of ROT bandwidth estimator to variable bandwidth estimator is increased when signal-to-noise ratio (s/n) is increasing but there are not significantly different for $n = 200$ case.

For the uniform design, we do not have a uniformly dominating bandwidth selection rule. Variable bandwidth estimator has more efficiency gain in the high signal-to-noise ratio design (s/n=8) than low signal-to-noise ratio (s/n=2), and larger number of observation in most cases of our simulation.

3.4.2. Normal Random Design

Table 3.2 shows the results of each bandwidth selection for the normal density design. For the test function 1, the relative efficiency ratio of LSCV bandwidth estimator for variable bandwidth (rf_1) is near 1 and there are not significantly different. The variable bandwidth estimator weakly dominates ROT bandwidth estimator. The relative efficiency of variable bandwidth estimator is increased when n is increasing and s/n is larger.

For the test function 2, variable bandwidth estimator is not significantly different from LSCV bandwidth estimator when $n = 200$. For larger sample size, the variable bandwidth estimator has a large relative efficiency compared to the constant bandwidth estimator.

For test function 3, the variable bandwidth estimator is more efficient than the LSCV bandwidth estimator and it is statistically significant when s/n = 8. The large amounts of observation are in the middle part and small amounts of observation are in both tail sides in normal design. Variable bandwidth changes with different observation

sizes, and it fits better at both tail sides where there are small observations. ROT bandwidth estimator acts the same as in the uniform design.

Table 3.3. *Relative Efficiency Ratios* of Bandwidth Selection for the Gamma Design

B/W	s/n	WMISE of Variable bandwidth		Relative Efficiency of LSCV (rf_1)		Relative Efficiency of ROT (rf_2)	
		N=200	N=400	N=200	N=400	N=200	N=400
Fn 1	2	2.29E-05	5.54E-05	0.97 (-0.15)	0.93 (-0.42)	0.84 (-0.82)	0.29 (-5.71)
	4	9.92E-05	1.14E-04	1.31 (1.37)	0.53 (-6.77)	2.01 (3.75)	0.47 (-8.69)
	8	4.35E-05	3.79E-05	0.57 (-1.21)	1.74 (2.84)	1.07 (1.37)	1.10 (0.64)
Fn 2	2	0.07	0.02	1.05 (0.19)	0.95 (-1.14)	1.05 (0.68)	0.75 (-1.14)
	4	0.02	0.04	1.21 (1.13)	0.86 (-1.19)	3.26 (4.38)	0.84 (-1.31)
	8	0.01	0.01	0.35 (-2.02)	0.82 (-1.59)	1.11 (1.13)	1.09 (0.82)
Fn 3	2	0.03	0.07	1.49 (1.36)	1.66 (3.26)	285.42 (2.96)	512.71 (6.32)
	4	0.02	0.03	1.52 (1.43)	1.42 (2.38)	670.94 (5.05)	860.57 (7.47)
	8	0.01	0.01	1.42 (1.28)	1.65 (3.58)	808.94 (4.31)	3197.55 (8.08)
Fn 4	2	22.43	8.24	1.17 (0.57)	2.01 (2.66)	1.17 (0.58)	2.01 (2.74)
	4	5.07	16.56	1.43 (0.68)	1.52 (2.56)	1.10 (0.69)	1.09 (0.44)
	8	2.22	1.08	0.19 (-2.91)	2.66 (8.89)	0.92 (-0.28)	0.98 (-0.12)

NOTE: The numbers of parenthesis are t-value of $H_0: rf_i = 1, t = (rf_i - 1)/s.e.(rf_i)$, where $s.e(rf_i) = \sqrt{\mathrm{var}(WMISE_{vw}) + \mathrm{var}(WMISE_j)}$, with i=1,2 and j = LSCV and ROT.

For the test function 4, variable bandwidth estimator and LSCV bandwidth estimator are not different statistically when n = 200. When n = 400, the relative efficiency of variable bandwidth estimator for LSCV bandwidth estimator is statistically significant when s/n is large. The relative efficiency of ROT is increasing when s/n ratio is larger. It represents that variable and LSCV bandwidth estimators have large

bias in more widely scattered data design since they have smaller bandwidth than ROT bandwidth estimator relatively and this small bandwidth gives wiggly fitting for the both tail parts.

Like the uniform density design, there is no uniformly dominating bandwidth selector for the normal density design. The variable bandwidth estimator is more relative efficient when s/n is larger than LSCV bandwidth estimator.

From the results of the uniform and normal density designs, there is no absolute dominating bandwidth selector. A different efficiency selector is selected for the different situations. However, for those two density designs, variable bandwidth estimator performs well.

3.4.3. Gamma Random Design

Table 3.3 shows the results of WMISE of variable bandwidth estimator and relative efficiency ratios when the covariates are generated from skewed distribution function. For the test functions 1 and 2, there is no dominating bandwidth selection rule. In our simulation, the constant bandwidth estimator has the greatest relative efficiency in several cases. However, variable bandwidth estimator has improved relative efficiency when s/n is larger, and variable and constant bandwidth estimators are not significantly different in efficiency respect in most of cases. For small s/n ratio, the reason that the relative efficiency gain of ROT bandwidth estimator is large is that variable and LSCV bandwidth estimators undersmooth for the skewed tail part.

For the test function 3, variable bandwidth estimator is strictly dominate constant bandwidth estimator, although it is not significantly different from LSCV bandwidth estimator when $n = 200$. The skewed observation and humped middle part

in the functional form has a relatively small number of observations in the humped part, and the variable bandwidth estimator fits this feature better than constant bandwidth estimator. ROT bandwidth estimator fits over-smoothly as in uniform and normal designs.

For the test function 4, the efficiency gain of variable bandwidth relative to the LSCV bandwidth estimator is much larger when s/n is larger and $n = 400$. For n =200 and larger s/n ratio, the variable bandwidth is less efficient than constant bandwidth estimator, which means that variable bandwidth estimator under-smooths.

For all test functions, relative efficiency ratios are larger when the signal-to-noise ratio is higher and $n = 400$. Also, like other designs, the efficiency gain of ROT bandwidth estimator is increasing when s/n ratio is larger. For $n = 200$, the relative efficiency gain of variable bandwidth estimator is reducing when s/n is larger because skewed design with small observation has too small and broadly scattered observations in the long tail part.

3.5. Summary

In this paper we surveyed three existing bandwidth selectors for local linear regression with different density designs. All selectors were empirically assessed by means of a simulation study. Numerical results demonstrate that the variable bandwidth estimator compare favorably to the other selectors.

Numerical results suggest that the LSCV selector performed well in the simple functional form and uniform or normal density design. This observation agrees with the study reported in Lee and Solo (1999).

There is no bandwidth selector that is uniformly the best in our uniform and normal design simulation. There are a few important empirical results:

1. LSCV bandwidth selector superior to ROT bandwidth selector in most designs.
2. ROT bandwidth estimator fits over-smoothly for the 'humped' part over all cases.
3. The constant bandwidth estimator seems enough to fit the simple linear functional form regardless of random design.
4. for the more complicated functional forms, the variable bandwidth estimator performs better than other bandwidth selector in our simulation.
5. For large samples, the relative efficiency gain of variable bandwidth relative to LSCV bandwidth estimator increases in the skewed data.
6. When the functional form has a 'humped' part, variable bandwidth fits better than any other constant bandwidth estimators in skewed data design, in general.

The variable bandwidth selector performs well in almost everywhere in our simulation with some exceptions. When the data are highly skewed or the functional form is very complicated in larger data set, the variable bandwidth selector is superior to the other bandwidth selectors.

CHAPTER 4

SEMIPARAMETRIC ESTIMATION OF WAGE EQUATION

4.1. Introduction and Literature Review

The usual human capital earnings function expresses the logarithm of earnings per hour, week, or year as a linear function of the number of years of school completed and as a quadratic function of experience (Mincer, 1974). The use of a quadratic in experience has been widely accepted. The quadratic parametric relationship between earnings and experience was challenged by Murphy and Welch (1990). They showed that the quadratic specification does not fit the data well, resulting in severely biased estimates of the earnings profile. They conclude that the quartic specification has a sufficiently small bias and could be used as the standard specification. There are some studies that use nonparametric or semiparametric methods to explore the specifications of the mean and variance of earnings with age. Basu and Ullah (1992) found that nonparametric specification indicates a 'dip' around prime-age. They said that a possible explanation for this dip is the generation effect, because the cross section data represent the earnings of people at a point of time who essentially belong to different generations. There is little argument that income-schooling relationships are stable. The dip indicates the lack of global concavity and that the quartic relationship may be more appropriate than the quadratic specification. The question is whether the difference between semiparametric and quadratic or quartic specification is significant and we will address this question using the nonparametric specification testing.

Buchinsky (1994) applied a semiparametric estimator to study shifts in the conditional distribution of earnings as a function of various characteristics over time

using the quantile regression. Quantile estimators seek to find the relationship between y_i and x_i at different quantiles of the conditional density of y_i given x_i. DiNardo et al. (1996) used the kernel-based density estimation to analyze the effects of institutional and labor market factors for the whole wage distribution itself. Although quantile estimation gives a clearer view of how the conditional density changes with x_i than that provided by the conditional mean, it has not been used extensively in the linear model framework (Fan and Gijbels, 1996). Ginther (1999) used the nonparametric mean regression and quantile regression to study the U.S. earning distribution.

In this paper, we first will estimate the human capital earning function by using the semiparametric method for 4 demographic groups; white-male, white-female, nonwhite-male, and nonwhite-female. A semiparametric model has two parts: one linear and one nonlinear part. The linear part contains a constant term, schooling, and dummy variables (part-time, SMSA, region, married, and major industries). The nonlinear part is years of experience. Semiparametric estimation is done in two stages. The first step consists of the usual nonparametric regression of dependent variable (y) and linear independent variables (x) on nonlinear part explanatory variable (z) to get the error terms, $y - E(y \mid z)$ and $x - E(x \mid z)$, respectively. The second step is to regress $y - E(y \mid z)$ on $x - E(x \mid z)$ to obtain the 'semiparametric estimates' of the coefficient value of linear part variable ($\hat{\beta}$). Additionally, to get the nonparametric conditional estimation for nonlinear part we run the nonparametric regression of residuals from OLS ($y - x\hat{\beta}$) on z.

Second, we will construct a confidence interval for the parameter of interest by bootstrapping.[1] Semiparametric estimation methods are evaluated by comparing semiparametric estimates to estimates from two widely used parametric specifications in wage equation: the quadratic model and quartic model (Juhn, Murphy, and Pierce, 1993).[2] If parametric and semiparametric estimates are significantly different from one another, we can conclude that parametric models are biased and semiparametric methods are warranted. Bootstrapping confidence interval will be compared with the parametric regression methods used in previous research in order to assess the potential contribution of semiparametric methods. Ginther (1999) analyzed the U.S. male earnings distribution using nonparametric methods with the rule of thumb bandwidth.[3] From the comparison of nonparametric and parametric specification, she concludes that nonparametric methods reveal nonlinearities in earnings profile. In this paper, we will apply semiparametric regression methods and use a data driven variable bandwidth selection rule.[4]

Third, the purpose of estimating a wage-earning profile is to analyze the wage structure. Bound and Johnson (1992) have used parametric estimation methods to analyze changes in the structure of wages in the 1980's. They examined between and within-group earnings inequality using data from 1973, 1979, and 1988. They attribute

[1]Bootstrapping is more accurate than first-order asymptotic distribution theory for estimating sampling distribution of asymptotically pivotal statistics (Horowitz, 2000).
[2] The quadratic model regresses log wages on years of schooling, years of experience, and years of experience squared. The quartic model regresses log wage on four schooling dummies for less than 12 years, exactly 12 years, between 13 and 15 years, and 16 or more years of schooling, a linear term in schooling and a quartic in experience fully interacted with all schooling terms.
[3] Rule of Thumb is defined by $h = A(K)\Sigma n^{-1/(d+4)}$, where $A(K)$ is constant, Σ is a diagonal matrix of standard deviation of covariates, d is the demension of the regression.
[4] She used the rule of thumb to choosing the optimal bandwidth. But rule of thumb bandwidths tend to oversmooth the mean and quantile estimates, while cross-validation methods have undersmoothed estimates.

changes in the earnings distribution to skill-biased technological change approximately measured by the residuals from a mean wage regression. In this chapter we examine 1990's wage structure using semiparametric estimation methods. There exists substantial agreement on the 'facts' that need to be explained. Recent changes in the U.S. wage structure can be summarized as follows (Katz, 1999):

- From the 1970s to the mid-1990s wage dispersion increased dramatically for both men and women. Earnings inequality increased even more dramatically if one includes the very top end of the distribution.
- Wage differentials by education and occupation increased. The labor market returns to years of formal schooling, work-place training, and computer skills appear to have increased in the 1980s and early 1990s. Wage differentials by age (experience) have expanded for non-college workers. But gender wage differentials have narrowed sharply since 1979.
- Wage dispersion expanded within demographic and skill groups. The wages of individuals of the same age, education, sex, and those working in the same industry and occupation, are increasingly unequal.
- Increased cross-sectional earnings inequality has not been offset by increased earnings mobility. Permanent and transitory components of earnings variation have risen by similar amounts. This implies that year-to-year earnings instability has also increased.
- Since these wage structure changes have occurred in a period of sluggish mean real wage growth (deflating wages by official consumer price indices), the real earnings of less-educated and lower-paid workers appear to have declined

relative to those of analogous workers two decades ago. The employment rates of less-educated and minority males fell substantially from the early 1970s to the early 1990s. The real wages and employment rates for disadvantaged workers have started to improve over the past few years.

The rise in U.S. wage dispersion has involved both large increases in educational wage differentials and a sharp growth in within-group (or residual) wage inequality.

This paper will be organized as follows: section two discusses the data and key variables. Section three describes the semiparametric estimation methods and specification tests of the assumptions of the linear models. Section four reports the nonparametric estimation results. Section five summarizes this study.

4.2. Data

Current Population Surveys (CPS) data from the merged outgoing rotation group file (fourth edition) 1979, 1988, and 1997 are used for this paper. The merged outgoing rotation group file contains information for a worker's weekly and hourly wages, weeks and hours worked, age, years of schooling, gender, race, industry specification, and residence. In order to examine changes in the wage distribution for workers most attached to the labor force and to compare results to previous studies, the sample criteria are quite similar to that used in Bound and Johnson (1992). Each sample eliminates all workers in agriculture, forestry, and fisheries, as well as household service and individuals with imputed hourly wages less than $1.00 or greater than $100. Also, self-employed and the governments workers are eliminated. Each sample has only individuals between the ages of 18 and 64 who reported employment as their normal

weekly activity. The resultant samples have 156,075 observations for 1979, 159,879 for 1988, and 126,593 for 1997, respectively. Potential work experience is calculated as age, minus years of schooling, minus six. The negative values for experience are reassigned by age. For example, if experience is less than zero and age is 18 then experience is one, if age is 19 then experience is two, and so on. Hourly earnings are calculated as edited weekly earning divide by work hours per week and deflated using the Consumer Price Index (CPI) with 1988 as the base year.

The CPS changed its coding of the schooling variable between the 1988 and 1997 surveys. Instead of measuring schooling by years completed, the 1997 survey records schooling as degrees completed. To accommodate this change, the 1997 variable is recoded to reflect the number of years it usually takes to complete a certain degree. This method is similar to Jaeger's (1997) approach.[5]

[5] Beginning with the 1992 survey, the CPS changed its coding of the schooling variable. Before the 1992 CPS was coded as the highest grade attended. The schooling variable took on values ranging between 0 and 18 years where 18 was the top code. The 1992 survey coded the schooling variable as educational attainment. Below the new codes for the schooling variable, the recode used to match it to previous years of the CPS, and Jaeger's (1997) recoding of the CPS schooling variable are listed.

1992 CPS Code	Description:	Record (years)	Jaeger's Recode
31	Less than first grade	1	0
32	1st, 2nd, 3rd, or 4th grade	4	2.5
33	5th or 6th grade	6	5.5
34	7th or 8th grade	8	7.5
35	9th grade	9	9
36	10th grade	10	10
37	11th grade	11	11
38	12th grade or no diploma	12	12
39	high school graduate (diploma or GED)	12	12
40	some college, no degree	13	13
41	associates degree: occupational/vocational	14	14
42	associates degree: Academic	15	14
43	Bachelors degree (BA, BS, AB)	16	16
44	Master's degree (MA, MS, MEng, MEd, MBA, JD)	17	18
45	Professional degree (MD, DDS, DVM, LLB, JD)	18	18
46	Doctorate degree	18	18

4.3. Semiparametric Estimation Method

Pudney (1993) and Ginther (1999) used the nonparametric estimation methods to examine conditional distributions of wealth and income of China and the U.S. male earning profile, respectively. Pudney used local constant nonparametric regression methods, and Ginther used local linear regression methods with 'Rule of Thumb' bandwidth selection to estimate wealth and income conditional on age. In contrast, our study uses semiparametric estimation with local linear methods to estimate conditional mean with binning data[6] for the nonparametric estimation part. The local linear methods provide an improved fit over local constant nonparametric methods because they reduce the bias of the nonparametric estimate, especially at the boundaries of the data.

To describe the semiparametric procedure, suppose that the ith observation is given by a $(1+p+k)\times 1$ vector (y_i, x_i', z_i'), $i = 1,\dots,n$, which is generated by the model

$$(4.1) \qquad y_i = f(z_i) + g(x_i, \beta) + u_i$$

where $f(\mathbf{z})$ is an arbitrary function of $\mathbf{z}$, while $g(\mathbf{x},\beta)$ is a known parametric function of $\mathbf{x}$ and a vector of unknown parameters β where the bold characteristic represents vector matix. The disturbance term u_i is assumed to satisfy

$$(4.2) \qquad E(u_i \mid x_i, z_i) = 0$$

and

$$(4.3) \qquad E(u_i u_j \mid x_i, x_j, z_i, z_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ 0 & \textit{otherwse} \end{cases}$$

[6] Data binning method is one of the fast algorithms. The kernel type nonparametric estimations have been considered computationally slow in comparison to other methods, although these methods are widely popular choices because of their simplicity and interpretability. Here we use the linear binning method, discussed by Silverman (1986).

The most popular functional form of $g(.,.)$ is linear, i.e.

$$g(x,\beta) = x'\beta. \tag{4.4}$$

Assuming (4.2), (4.3), and (4.4), the estimates of f and g are given by

$$\begin{aligned} \hat{f} &= S(y-\hat{g}) = S(y-X\hat{\beta}) \\ \hat{g} &= P(y-\hat{f}) = X(X'X)^{-1}X'(y-\hat{f}) = X\hat{\beta} \end{aligned} \tag{4.5}$$

where S be a linear smoothing operator described below and $P = X(X'X)^{-1}X'$ is a projection matrix. Combining the above two equations, we obtain

$$\hat{\beta} = [X'(I-S)X]^{-1}X'(I-S)y \tag{4.6}$$

$$\hat{f} = S(y-X\hat{\beta}), \tag{4.7}$$

For example, for the Nadaraya-Watson kernel regression estimator, the (i,j) element of S is given by $S_{ij} = \dfrac{K(z_i - z_j)}{\sum_j K(z_i - z_j)}$ for $i,j = 1,\ldots,n$, where $K(\cdot)$ is a kernel function, which generates the weights with a maximum at zero and satisfies certain moment conditions. Under some regularity conditions[7] on the bandwidth parameter h, defined below, $\hat{\beta}$ and $\hat{f}$ are consistent estimators of β and $f = E(y - X\beta \mid Z)$, respectively. In particular, it is well known that the convergence rate of $\hat{\beta}$ to β as $n \to \infty$ is the same as in the parametric case (Robinson, 1988).

We will apply the nonparametric local linear method to estimate conditional distributions of the nonlinear part of the wage equation.[8] Nonparametric local linear mean estimates find the local parameters, $\hat{\beta}_j$, conditional on the local observations of

---

[7] (1) $h = h_n \to 0$ as $n \to \infty$, (2) $nh_n \to \infty$ as $n \to \infty$. These imply that, as n increases, h should be decreased at a slower speed than n^{-1}.

[8] By Fan and Gijbels, 1996, the local linear methods provide an improved fit over local constant nonparametric methods because they reduce the bias of nonparametric estimate, especially at the boundaries of data.

x_j (experience), that minimize the weighted least squares equation (4.8) where nonparametric kernel, $K(.)$, acts as the weight. The form of local linear regression is:

$$(4.8) \qquad \sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1(x - x_i)\}^2 K\left(\frac{x_i - x}{h}\right),$$

where x be the vector of conditioning points, n be the number of observation, and h be the bandwidth. The mean estimator of y given x is defined as $\hat{\beta}_0 = \hat{m}(x)$ with $m(x) = E(y \mid x)$.

It is important to choose a kernel and bandwidth in order to implement local linear nonparametric regressions estimates like other kernel-type nonparametric estimation. The Epanechnikov kernel, $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ is used for nonparametric mean and variance estimates in equation (4.8). The bandwidth is selected by adapting the variable bandwidth to the univariate regression problem instead of a constant bandwidth. The constant bandwidth depends neither on the location of x nor on that of the data x_i. Such an estimator does not fully incorporate the information provided by the density of the data points. Furthermore, constant bandwidth is not flexible enough for estimating curves with a complicated shape. All these considerations lead to introducing a variable bandwidth $h / a(x_i)$, where $a(\cdot)$ is some nonnegative function reflecting the variable amount of smoothing at each data point. The nonparametric regression with variable bandwidth minimizes

$$(4.9) \qquad \sum_{i=1}^{n} \{y_i - m(x_i)\}^2 a(x_i) K\left(\frac{x_i - x}{h} a(x_i)\right)$$

where $m(x_i) = E(y_i \mid x_i = x)$.

This variable bandwidth has an important advantage. It gives certain flexibility in smoothing various types of regression functions. Optimization over all possible variable bandwidths leads to an optimal bandwidth and hence improves performance. It will be seen that for an optimal choice of $a(\cdot)$ in this function is proportional to $f_x^{1/5}$, and this is precisely how an ideal variable kernel smoother should behave.

For the local linear regression with variable bandwidth, the theoretical choice and the asymptotic properties were studied by Fan and Gijbels (1992). Instead of (4.8), The form of the estimator for the local linear smoother is the minimization of (4.10) with respect to β_0 and β_1.

$$\sum_{i=1}^{n}\{y_i-\beta_0-\beta_1(x_i-x)\}^2 a(x_i)K\left(\frac{x_i-x}{h}a(x_i)\right) , \tag{4.10}$$

where

$$\begin{aligned} &\hat{m}(x)=\sum_{i=1}^{n}w_i y_i\Big/\sum_{i=1}^{n}w_i, \\ &w_i=a(x_i)K\left(\frac{x_i-x}{h_n}a(x_i)\right)\{S_{n,2}-(x_i-x)S_{n,1}\} \\ &\text{where } S_{n,j}=\sum_{i=1}^{n}a(x_i)K\left(\frac{x_i-x}{h_n}a(x_i)\right)(x_i-x)^j \end{aligned} \tag{4.11}$$

Non- and semi-parametric methods reduce the explicit statistical assumptions needed to estimate a model. However, these estimation methods are computationally expensive. We will evaluate whether semiparametric estimation methods are warranted by comparing semiparametric estimates to parametric specifications: the quadratic and quartic models. If parametric and semiparametric estimates are significantly different from one another, this would provide evidence that parametric models are biased and semiparametric methods are warranted.

We will use a Hausman (1978) test to determine whether nonparametric part is statistically different from the null parametric specification. Let the semiparametric estimator of the conditional mean of estimates by $\hat{m}_{sp}$ and parametric one by $\hat{m}_p$. Although the semiparametric estimates are inefficient when the parametric specification is correct, we find that $\hat{m}_{sp} \approx \hat{m}_p$ since both $\hat{m}_{sp}$ and $\hat{m}_p$ are consistent. When the model is more nonlinear than the parametric specification, the semiparametric estimates are consistent but the OLS estimates are not consistent, and we find $\hat{m}_{sp} \neq \hat{m}_p$. Denote the asymptotic covariance matrix estimates for $\hat{m}_{sp}$ and $\hat{m}_p$ as $\hat{V}_{sp}$ and $\hat{V}_p$, the test statistic is $(\hat{m}_{sp} - \hat{m}_p)'(\hat{V}_{sp} - \hat{V}_p)^{-1}(\hat{m}_{sp} - \hat{m}_p)$, which is distributed χ^2 with K–1 degrees of freedom where K denotes number of covariates.

Bootstrapping is applied to construct confidence intervals for semiparametric estimates, and of critical values for consistent semiparametric test statistics. The bootstrap approximations to the small sample properties are usually far superior to those provided by the first-order asymptotic approximations (Horowitz, 1997).

The general wage equation will be denoted as:

$$y_i = z'_i \beta + m(x_i) + u_i, \tag{4.12}$$

where z is education and some dummy variables, x is experience. In (4.12) $E(u_i \mid z_i, x_i) = 0$. Taking the conditional expectation (4.12) leads to

$$E(y_i \mid x_i) = m_{2i} = E(z_i \mid x_i)'\beta + m(x_i). \tag{4.13}$$

Consequently,

$$\begin{aligned} y_i - E(y_i \mid x_i) &= (z_i - E(z_i \mid x_i))'\beta + u_i, \\ m(x_i) &= E(y_i \mid x_i) - E(z_i \mid x_i)'\beta \end{aligned} \tag{4.14}$$

Since (4.14) has the properties of a linear regression model with dependent variable [y_i – $E(y_i|x_i)$] and independent variables [$z_i - E(z_i|x_i)$], an estimator of β is

$$\hat{\beta} = \left[\sum_{i=1}^{n}(z_i - \hat{m}_{1i})(z_i - \hat{m}_{1i})'\right]^{-1}\left[\sum_{i=1}^{n}(z_i - \hat{m}_{1i})(y_i - \hat{m}_{2i})\right] \tag{4.15}$$

where $\hat{m}_{1i}$ and $\hat{m}_{2i}$ are the kernel-based estimators of $m_{2i} = E(y_i \mid x_i)$ and $m_{1i} = E(z_i \mid x_i)$, respectively. Once we found $\hat{\beta}$, $m(x_i)$ can be estimated from (4.14) as $\hat{m} = \hat{m}_{2i} - \hat{m}_{1i}\hat{\beta}$.

4.4. Estimation Results

4.4.1. Comparison between OLS and Semiparametric Estimation

Nonparametric or semiparametric methods reduce the explicit statistical assumptions needed to estimate a model. However, the shortcoming of these methods is that they are computationally expensive. To overcome this problem, we use the binning method. This section evaluates whether comparing semiparametric estimates to two widely used parametric specifications warrants semiparametric estimation method: the quadratic and quartic wage equation. If parametric and semiparametric estimates are significantly different from one another, this would provide evidence that parametric models are biased and a semiparametric method is warranted.

In order to compare the parametric models with semiparametric estimates, all specifications estimate trimmed means of wage. Unlike other researchers who have assumed a Pareto tail for the top-coded portion of wage distribution to identify in CPS data, we will use the Ginther's (1999) method which used trimmed means and variances after trimming both tails of earnings distribution by 1.5 percent. These trimmed estimates are compared with 95% bootstrapping confidence intervals from the

semiparametric mean using 1,000 subsamples. A complete table of these results appears in Appendix Table A.1, Figure 4.1, and Figure A.1.1 - A.1.2.

Figure 4.1 represents the graphical comparison semiparametric and OLS specifications. Panels A and B show mean regression conditioning on experience for white-males and white-females in 1997, respectively. Semiparametric estimations show a hump at the middle experience group and that quadratic OLS estimation lies out of 95% confidence interval, and this hump is significantly different from quadratic specification. Quadratic and Quartic specifications are over estimated relative to semiparametric estimation at lower experience and under estimated at higher semiparametric estimation at the lower experience. The positions of highest real hourly wages in parametric specification are reached earlier than semiparametric specification. At this highest position semiparametric specification is significantly different from parametric specification. Panels C and D graph mean regressions for nonwhite workers. Semiparametric specification is wiggly at larger than 20 years of experience workers where there are few observation. Appendix Figure A.1.1 through Figure A.1.2 show the graphical comparison between semiparametric and OLS specification for white-female, nonwhite-male, and nonwhite-female in 1988 and 1979, respectively. From the Panel A of Appendix A.1.1 and A.1.2, semiparametric and OLS specifications are not much different from each other for the white-male workers in the graphical analysis. The semiparametric specification is significantly different from OLS specification for the other demographic groups in 1979 and 1988.

From the graphical analysis of comparison between semiparametric and OLS specification, conditional mean estimates of quadratic specification lie outside of the

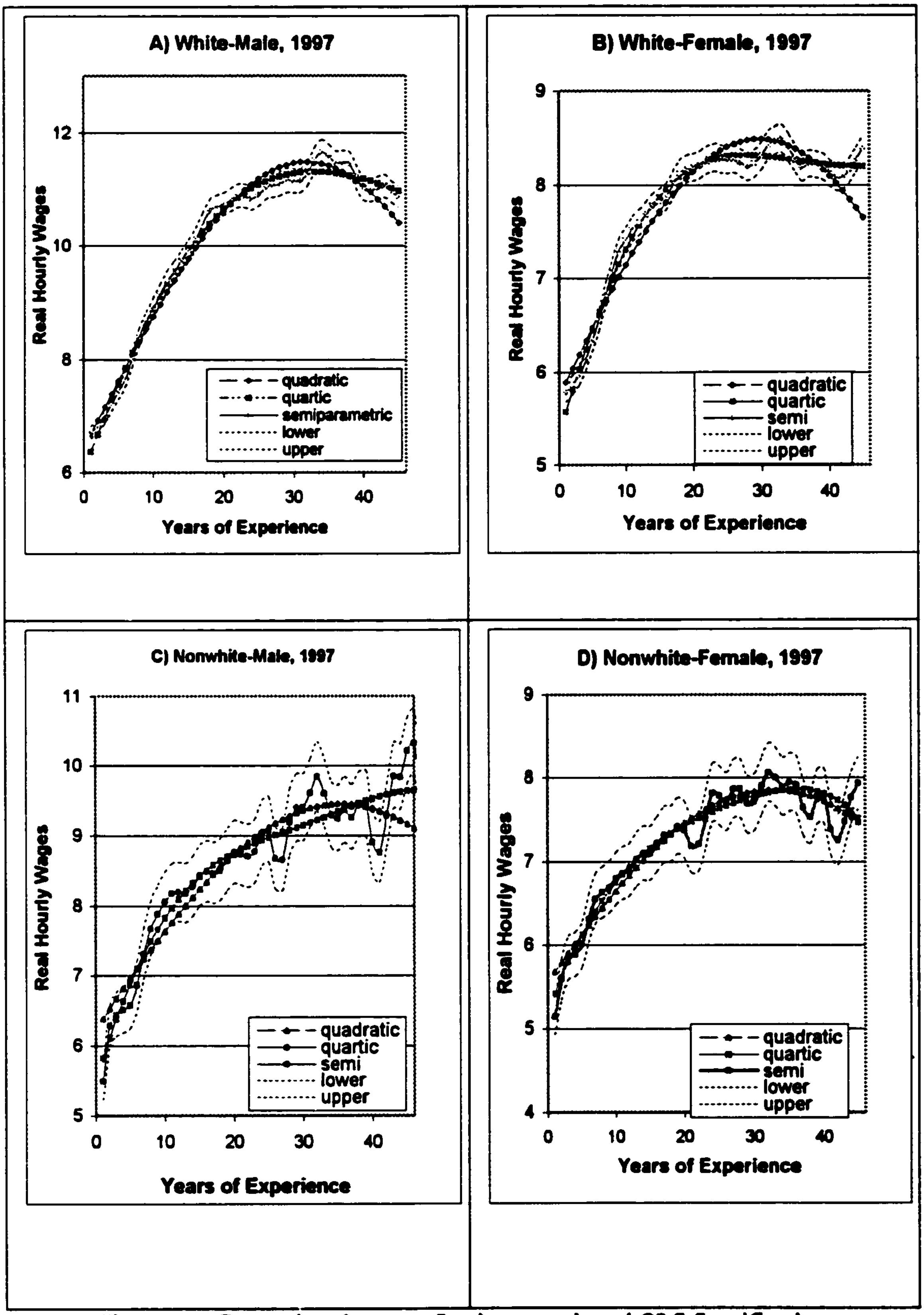


Figure 4.1. Comparison between Semiparametric and OLS Specification

semiparametric bootstrapping confidence intervals from 15.2 % of white-female 1997 for quartic model to 77.1% of white-female 1988 for quadratic model. The largest deviation of quartic model from semiparametric confidence interval is 31.9% and the others are less than 23%. The quartic model fits data better than the quadratic model.

To test the linear specification (H_0) against the semiparametric specification, we conduct a Hausman test. The test is based on the contrast between the OLS estimate $\hat{\beta}$ and the semiparametric estimate $\tilde{\beta}$. If H_0 is true, that is, the linear specification is correct, then $\hat{\beta}$ is consistent and more efficient than $\tilde{\beta}$, while $\tilde{\beta}$ is consistent whether or not H_0 is true. This test provides a simple and convenient way to examine the validity of the semiparametric specification. Since $\tilde{\beta}$ is $\sqrt{n}$ consistent, the test statistic has the usual limit distribution and the test may be conducted in the same manner as in the parametric case. The test statistic is given by:

$$m = (\hat{\beta} - \tilde{\beta})'[\mathrm{var}(\hat{\beta}) - \mathrm{var}(\tilde{\beta})]^{-1}(\hat{\beta} - \tilde{\beta}) \tag{4.16}$$

which, under H_0, has chi-squared distribution with degrees of freedom equal to the dimension of β. The semiparametric estimates of the parametric portion of the model and the results of two alternative OLS models are presented in Table 4.1 and Table A.2. The first two columns of Table 4.1 and Appendix Table A.2 report the estimates based on the OLS of quadratic and quartic wage equation model with dummy variables. The Hausman statistic computed from 1997, white-male data is 304.008 and 50.339 for quadratic and quartic model, respectively. Since the 99% quantile of the chi-squared distribution with 24 degrees of freedoms is 42.980, the consistency of OLS regression estimates is rejected, providing support for our semiparametric model.

These results indicate that semiparametric estimation methods capture the nonlinearities in the earnings profile, and quadratic and quartic models exhibit some asymptotic bias since semiparametric estimation is consistent. Semiparametric estimation methods are warranted given the significant differences in parametric and semiparametric specification.

4.4.2. Semiparametric Mean Estimates

The estimated values of the real hourly wages and 95% confidence interval of conditioning on experience and schooling are reported in Appendix Table A.3.

Figure 4.2 graphs the mean of real hourly wages and 95% confidence interval for 1979, 1988, and 1997 conditioning on experience for 10, 12, 16, and 18 years of schooling for white-male workers. Panels A and B of Figure 4.2 show significant differences in the mean wage at all levels of experience holding schooling constant at 10 and 12 years among three estimated years. The means of wages are steadily decreasing from 1979 to 1997, while the gap between the estimations become smaller from 1979-1988 to 1988-1997. The slopes of the 1997 estimate in panel A and B are somewhat steeper than in 1988 for lower experience workers, indicating an increasing return to experience for workers with 10 and 12 years of schooling. The confidence interval is increasing from small years of experience to large years of experience, indicating wage dispersion increase. Holding schooling constant at 16 in Panel C, there is no significant difference in the level or the slope of the mean wage in 1988 and 1997, while there is significant different mean wage between 1997 and 1979. The slope of estimates in 1997 is steeper than in 1988 for the workers with less than 25 years of experience. Although real hourly wages are significantly decreasing for workers with a college degree between

1979 and 1988, it does not happen between 1988 and 1997. Panel D shows a significant increase in 1997 in the level of the mean wage for workers with 18 years of schooling and all levels of experience workers. The real hourly wages in 1997 are larger than those in 1988 for workers with higher education level. The slope of mean wage is also steeper in 1997 than in 1988, indicating an increasing return to experience for those workers. From Panels A through D of Figure 4.2, the decline of real hourly wages is much larger with lower educated and lower experience workers.

Appendix Figure A.2 represents that the mean of real hourly wages and 95% confidence interval for 1979, 1988, and 1997 conditioning on experience for 10, 12, 16, and 18 years of schooling for white-female, nonwhite-male, and nonwhite-female workers. For white-female, the real hourly wages in 1997 are greater than those in 1988 for higher education than high school almost everywhere of experience with steeper slopes, and the increasing ratio is larger with higher education. The increase in real hourly wages are fast for the first 10 years experience workers and after 10 years experience real hourly wages do not much increase for under high school graduate level workers. For the nonwhite workers in the Panels C and D of Appendix Figure A.2 the real hourly wages have similar pattern with white workers with larger variance. The mean estimates in 1997 represent a hump around 25 years of experience for every schooling and demographic group workers.

The panels in Figure 4.3 indicate dissimilar rates of return to schooling across experience levels. Panel A and B shows the results for workers with 5 and 10 years of experience. Mean of real hourly wages conditioning on schooling for workers with less than 16 years of schooling in 1997 are less than those in 1988.

Table 4.1. Coefficient Estimates 1997, Male-White

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.310(0.017)	0.270(0.018)	
Region2	-0.024(0.006)	-0.024(0.006)	-0.025(0.006)
Region3	-0.071(0.006)	-0.070(0.006)	-0.071(0.006)
Region4	-0.027(0.006)	-0.026(0.006)	-0.027(0.006)
FTPT	0.211(0.009)	0.200(0.010)	0.161(0.005)
SMSA	0.137(0.005)	0.136(0.005)	0.122(0.005)
Marry	0.112(0.004)	0.108(0.005)	0.116(0.004)
Dind1	0.056(0.010)	0.056(0.010)	0.056(0.010)
Dind2	0.057(0.009)	0.058(0.009)	0.057(0.009)
Dind3	0.011(0.010)	0.011(0.010)	0.009(0.010)
Dind4	-0.005(0.011)	-0.005(0.011)	-0.006(0.011)
Dind5	0.155(0.015)	0.157(0.015)	0.154(0.014)
Dind6	-0.052(0.011)	-0.053(0.011)	-0.053(0.012)
Dind7	-0.188(0.009)	-0.186(0.009)	-0.189(0.009)
Dind8	0.118(0.015)	0.118(0.015)	0.119(0.015)
Dind9	-0.048(0.012)	-0.048(0.012)	-0.047(0.013)
Dind10	-0.221(0.017)	-0.222(0.017)	-0.221(0.018)
Dind11	-0.149(0.017)	-0.147(0.017)	-0.151(0.018)
Dind12	-0.089(0.018)	-0.091(0.018)	-0.093(0.020)
Dind13	-0.074(0.018)	-0.075(0.018)	-0.075(0.018)
Dind14	-0.274(0.030)	-0.276(0.030)	-0.281(0.037)
Dind15	-0.152(0.012)	-0.150(0.011)	-0.155(0.012)
Dind16	0.081(0.013)	0.080(0.013)	0.080(0.014)
Dind17	0.046(0.011)	0.047(0.011)	0.046(0.011)
Schooling	0.095(0.001)	0.095(0.001)	0.095(0.001)
Experience	0.035(0.001)	0.049(0.003)	
Experience^2	-0.001(1.56E-06)	-0.001(2.91E-05)	
Experience^3		1.89E-06(1.01E-06)	
Experience^4		-1.0E-08(1.16E-08)	
R-square	0.365	0.366	0.366
Hausman test: OLS vs. Semiparametric	304.008	50.339	

Note. Standard errors constructed by bootstrapping are in parentheses.

After 16 years of schooling mean wages in 1997 are higher than those in 1988. The levels of real hourly wages in 1997 are higher than those in 1988 for higher schooling and lower experience workers. Also, the slopes of estimates of mean wages in 1997 are steeper than those in 1988. The mean wages between in 1988 and in 1997 are not significantly different for higher schooling although it is significantly different from 1979. Panel C and D show that mean wages conditioning on schooling for higher experience workers with less educated than college graduate in 1997 are less than those in 1988. For the higher experience workers mean wages conditioning on schooling are similar pattern with 5 and 10 years experience workers with the decreasing the gap of levels of mean real hourly wage from 1979 to 1997. The slopes of the mean wages in 1997 are steeper than other periods for all experience workers, indicating return to higher education for workers. Appendix Figure A.3 shows mean wages conditioning on schooling for others demographic groups. Figure A.3.1 graph mean wages for white-female group. Mean wages conditioning on schooling of each years are not significantly different for less education than college graduate level. Mean wages conditioning on schooling in 1997 are dramatically increase for workers with higher education level than college graduate. Figure A.3.2 and A.3.3 represent mean wages conditioning on schooling for nonwhite demographic groups. Although the patterns of mean wages conditioning on schooling for nonwhite group are similar to those of white group, the levels of real hourly wages are not significantly different among every years.

Between 1988 and 1997 the relative return to schooling and experience, measured by the slope of the mean wage function, increased, while the level of mean real hourly wages decreased or stayed for almost all workers.

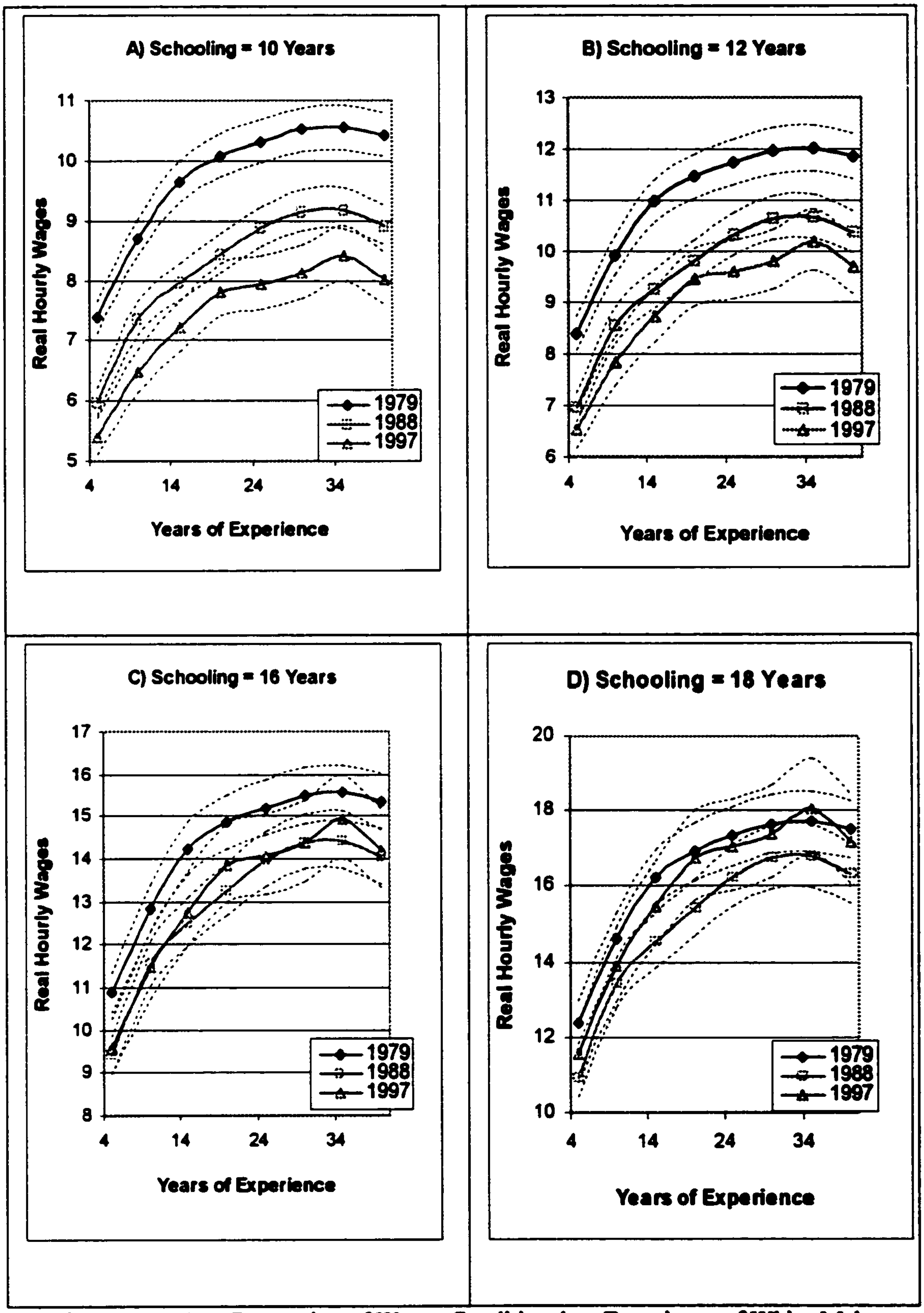


Figure 4.2. Mean Regression of Wages Conditional on Experience of White-Male

4.4.3. Changes in Between-Group Earning Inequality

The first stylized fact of between groups is wage differentials by education increased. The labor market returns to years of schooling, and work-place training appear to have increased in the 1980s and early 1990s. The second stylized fact is wage differentials by experience have expanded for non-college workers. The third fact is gender wage differentials have narrowed sharply since 1979.

Table 4.2 shows the wage difference of college-high school for all demographic groups is consistently increasing. The real hourly wage ratio between college and high school for white-male group is increasing by amount of 0.08 between 1988 and 1997, while those between in 1979 and 1988 increased only 0.043, for example. This increasing in the wage ratio is similar for all groups. Also, the wage ratio between college graduate and more educated than college workers increased. The first stylized fact is clearly seen from Table 4.2 and Figure 4.4. The increase in the wage ratio between high school and college is much higher than those between college and more than college. Figure 4.4 represents that the ratio between college and high school is larger when the lower experience workers.

For the second facts, Table 4.3 represents the wage differentials for non-college workers of all groups. The wage differentials are consistently expanded with one exception; the wage differential in 1997 is 3.529, while that in 1988 is 3.816 for nonwhite-male group.

The third fact that gender wage differentials are decreasing is represented in Figure 4.5. In Panel A, male/female differentials for white workers are reduced very clearly. The difference ratio is more reduced in higher experience than lower

experience. Panel B shows gender differentials for non-white workers. The gender differentials in 1988 and 1997 are reduced from that in 1979, but the pattern of gender differentials between 1988 and 1997 is different. Gender differential of nonwhite workers with less than 20 experiences is increased from 1988 to 1997.

Table 4.2. College (CO)/High School (HS) Wage Ratio and College/more than College Wage Ratio

CO/HS	1979	1988	1997
White-Male	0.259	0.301	0.381
White-Female	0.254	0.319	0.421
Nonwhite-Male	0.218	0.283	0.367
Nonwhite-Female	0.244	0.297	0.390
CO/more than CO			
W-M	0.129	0.151	0.191
W-F	0.127	0.159	0.211
Nw-M	0.109	0.142	0.183
Nw-F	0.122	0.149	0.195

4.4.4. Changes in Within-Group Earning Inequality

There are some stylized facts for the changes in within-group earning inequality. First, wage dispersion increased dramatically for both male and female from the 1970s to the mid-1990s. Second, wage dispersion expanded within demographic and skill groups.

The coefficient of variation is used for analysis of within group earning inequality. The coefficient variation (CV) measures the scaled variance of the earnings distribution. Let $\sigma(x)$ be the standard deviation of earnings conditional on schooling

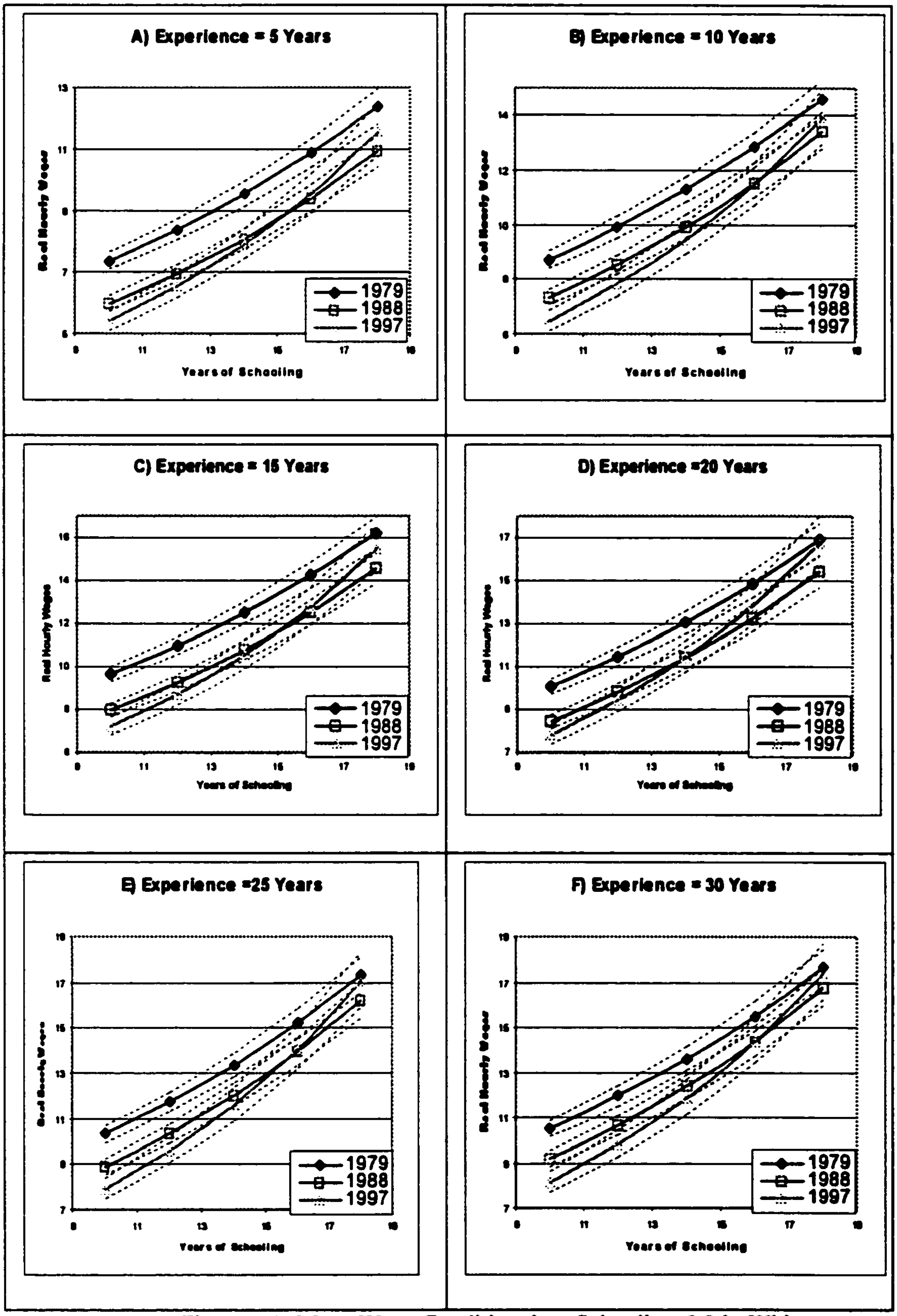


Figure 4.3. Mean Wage Conditional on Schooling, Male-White

and experience and $\mu(x)$ be the conditional mean, then the conditional coefficient of variation is given by;

$$CV = \frac{\sigma(x)}{\mu(x)}. \tag{4.17}$$

Estimates of within-group inequality are presented in Figure 4.6 and Figure A.4. Panel A through D of Figure 4.6 graph estimates of the CV of white-male workers holding schooling constant at 10, 12, 16, and 18 years and conditioning on experience for 1979, 1988, and 1997. In all panels, conditional earnings inequality increased between 1988 and 1997, while magnitude of these changes differ given the level of schooling. Estimates of the CV in 1997 are significantly higher than those in 1988 in all experience years. The conditional earnings inequality between 1979 and 1988 are different from those between 1988 and 1997. In panel A and B of Figure 4.6, estimates of the CV in 1988 are not significantly different from those in 1979 for all years of experience. The CV in 1988 is lower than those in 1979 between 5 and 20 years experience for 10 years schooling workers, and higher between 20 and 40 years experience for 12 years schooling workers. In panel C and D, the coefficient of variance in 1988 is higher than those in 1979 for all years of experience for higher level of schooling workers. Figure A.4 shows the estimates of CV for white-female, nonwhite-male, nonwhite-female workers. Panel A in A.4.1, the estimates of CV for white-female workers in 1988 are lower than those in 1979 for all years of experience and 10 years of schooling workers. Panel B shows the estimates of CV in both 1979 and 1988 are not significantly different for 12 years of schooling workers. In Panels C and D in A.4.1, the estimates of CV in 1988 are higher than those in 1979 for all years of experience and higher schooling workers. Figure A.4.2 and A.4.3 represent the

estimates of CV for nonwhite-male and nonwhite-female workers, respectively. For all nonwhite-male workers, showed in A.4.2, the estimates of CV in 1988 are lower than those in 1979, while the gap between 1988 and 1979 is decreasing with larger years of experience. In Panel A and B of A.4.3 represent the estimates of CV in 1988 are lower than those in 1979 when the years of experience are less than 10 years, while it is higher with larger years of experience than 10 years. Panel C and D of A.4.3 show the estimates of CV for higher grade schooling workers in 1988 are higher than those in 1979.

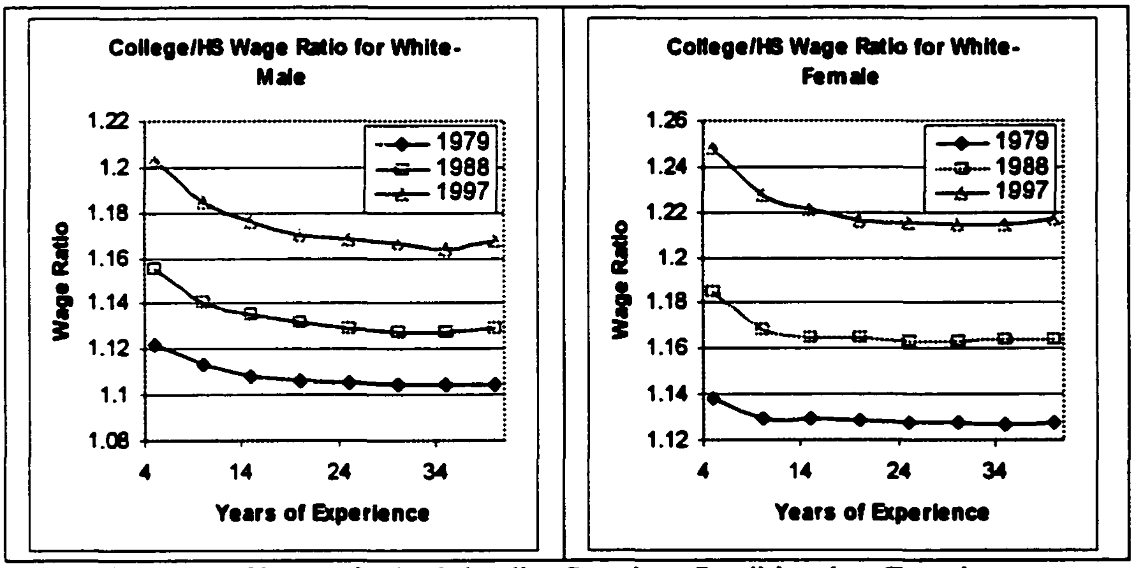


Figure 4.4. Changes in the Schooling Premium Conditional on Experience

The graphs presented in Figure 4.6 and Figure A.4 suggest that within-group earnings inequality changed by different amounts for different schooling and experience groups between 1979 and 1988, while those between 1988 and 1997 are significantly different for all workers. Semiparametric estimation methods reveal heterogeneous changes in within-group earnings inequality at various levels of schooling and experience. By focusing on two time periods, between 1988 and 1997, this study shows

the location of changes in within-group inequality. Within-group earnings inequality increased the all workers with all experience and all schooling group workers.

Table 4.3. Wage Differentials for Non-College Workers

CO/HS	1979	1988	1997
White-Male	4.639	4.685	4.791
White-Female	1.831	2.215	2.698
Nonwhite-Male	3.162	3.816	3.529
Nonwhite-Female	1.895	2.358	2.655

4.5. Evidence on Alternative Explanations

Our next step is to perform an empirical analysis of the reason for the observed wage change of the 1980's and 1990's periods. Bound and Johnson (1992) set up the theoretical model for empirical analysis of the reasons for the observed relative changes. They divided the aggregate work force into I=32 demographic groups (by four experience, four education, and gender). The wage rate of group-i workers in industry j is W_{ij}, and this is defined as the product of the 'competitive' wage, W_{ic}, and a relative rent, μ_{ij}.

We can observe the geometric mean of the wage rate for group-i workers by defining Y_{ij} and M_{ij} as the logarithms of W_{ic} and μ_{ij}, that is,

$$Y_i = Y_{ic} + \sum_j M_{ij}\phi_{ij} \tag{4.18}$$

where $Y_{ic} = \ln(W_{ic})$ and $\phi_{ic} = N_{ij} / N_i$ is the proportion of group-i workers who are employed in industry j. The change in the relative average log wage of each group i is

$$\begin{aligned} dY_i &= dY_{ic} + \sum_j (\phi_{ij} dM_{ij} + M_{ij} d\phi_{ij}) \\ &= (1 - 1/\sigma)d(\ln b_i) - (1/\sigma)d(\ln N_i) + (1/\sigma)d(\ln D_i) \\ &\quad + \sum_j (\phi_{ij} dM_{ij} + M_{ij} d\phi_{ij}), \end{aligned} \tag{4.19}$$

where b_i is an index of the technical efficiency of group-i workers, σ is intrafactor elasticity of substitution assumed constant and equal across industries. The detailed description of this formula is in the Appendix of Bound and Johnson (1992). The change in the relative average wage for group-i workers is (4.18) minus its weighted average across all I groups. Bound and Johnson (1992) utilized a conventional model of the determination of competitive wages for each of the I demographic groups and the

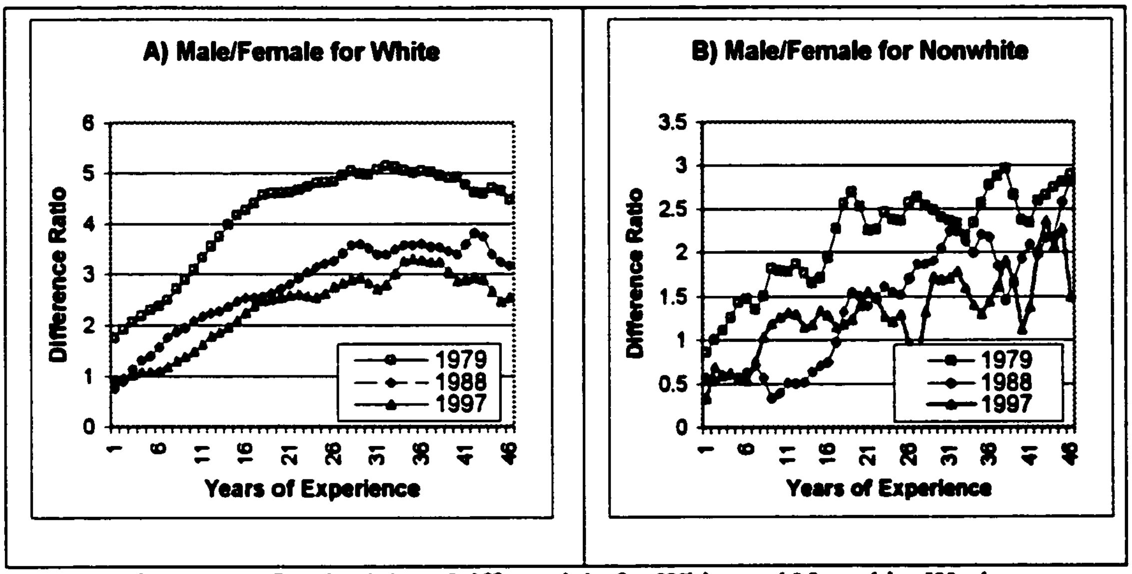


Figure 4.5. Gender Wage Differentials for White and Nonwhite Workers

employment level for each group in each of J industries (N_{ij}) to check the alternative explanations of changes in competitive wage levels. They used five assumptions for purpose of checking[9].

The model leads to the conclusion that the change in the competitive wage of group-i workers depends positively on their average rate of technical change $d(\ln b_i)$,

[9] (1) output in each industry is a function of efficiency units of employment, $b_{ij}N_{ij}$, of each of the demographic groups, where b_{ij} is an index of the technical efficiency of group-i workers in industry j. (2) the demand for the output of each industry is a function of its relative price and an exogenous shift parameter. (3) the employment levels of all groups in each industry are determined by equations setting the marginal revenue products of the I labor inputs equal to their competitive wage rates. (4) the economy is at full employment in the sense that the total effective aggregate labor supply of each labor group (N_i) is employed in the J industries in the economy. (5) the N_i's are exogenous.

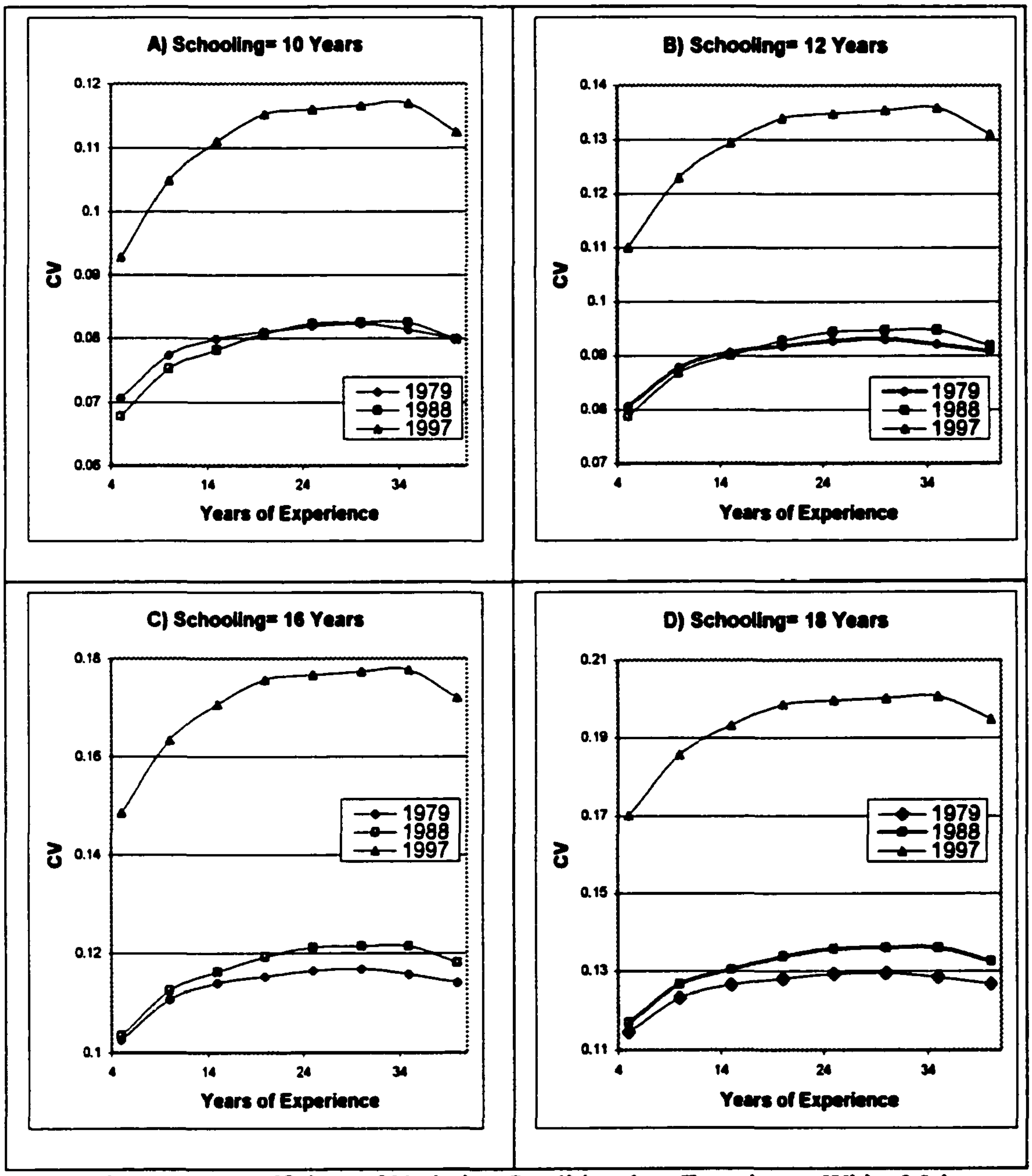


Figure 4.6. Coefficient of Variation Conditional on Experience, White-Male

negatively on their relative supply change $d(\ln N_i)$, and positively on the change in their relative product-demand-shift index $d(\ln D_j)$.

We now apply the model developed in Bound and Johnson (1992) to the question of the reasons of the wage structure that were described in Section 4.4.4,

examining in turn explanations that focused on changes in average rents, changes in the structure of product demand, and technical change.

4.5.1. Changes in the Industrial Wage Structure

It is necessary to estimate group-average wage rates by industry, Y_{ij}'s, to explain how much of the observed wage-structure changes was due to changes in the demographic composition of employment between high- and low-wage industries and how much was due to changes in industry wage differentials. These can be obtained from the estimated parameters of our original semiparametric estimation for each of the three years on which log CPS wages. The average log wage for group-i workers in industry j is the predicted value for the particular educational level (10,12,14,16) and experience level (5, 15, 25, 35) for full-time workers residing in SMSA's in the average region for that group with the relevant industry coefficient in effect.

To estimate average industry wage effects across all groups in each period, we regressed weighted Y_{ij} (by $[k_i \phi_{ij}]^{0.5}$) on dummy variables for the 32 groups and 17 industries which is shown in Table 4.4 where k_i is group-i's share of total employment in particular year. Table A.6 shows the estimated relative wage changes and employment distributions. This decomposition represents a group effect and a common-industry effect of each Y_{ij}, and the deviations of the estimated value of the common-industry effect from its mean are reported as M_j (the estimated value of the log of μ_j) in Table 4.4. The first three columns of Table 4.4 represent the fraction of total employment ($\phi_j = \sum_i \phi_{ij} k_i$) from industries, and we can see a significant shift of fraction of total employment during 1980's and 1990's from industries that were the

traditional employers of male blue-collar labor (manufacturing industries) toward industries that employ larger fractions of women and highly educated labor (like finance and professional services). For example, the employment fraction of Durables/mining industry is reduced from 0.178 in 1979 to 0.128 in 1997 and that of professional services is increased from 0.025 in 1979 to 0.038 in 1997.

Table 4.4. Estimated Aggregate Weights for 17 CPS Industries in 1979, 1988, and 1997[10]

	Weight (ϕ_j)			Wage Effect (M_j)		
Industry	1979	1988	1997	1979	1988	1997
Construction	0.064	0.061	0.061	0.002	0.000	0.000
Durables/mining	0.178	0.140	0.129	0.069	0.050	0.051
Nondurables	0.108	0.092	0.081	0.039	0.028	0.025
Transport	0.045	0.050	0.052	-0.004	-0.001	0.001
Utilities	0.018	0.018	0.016	-0.031	-0.031	-0.038
Wholesale trade	0.041	0.043	0.043	-0.004	-0.005	-0.005
Retail trade	0.160	0.166	0.178	0.055	0.052	0.067
Finance	0.031	0.037	0.037	-0.019	-0.015	-0.016
Business services	0.022	0.044	0.051	-0.022	-0.003	0.003
Personal services	0.023	0.027	0.027	-0.026	-0.021	-0.023
Entertainment	0.010	0.011	0.017	-0.041	-0.039	-0.034
Medical	0.033	0.042	0.053	-0.018	-0.011	-0.004
Hospitals	0.050	0.051	0.036	-0.001	-0.002	-0.015
Welfare	0.019	0.019	0.017	-0.027	-0.028	-0.037
Education	0.105	0.100	0.098	0.031	0.026	0.026
Professional services	0.025	0.037	0.038	-0.022	-0.011	-0.014
Public administration	0.069	0.063	0.065	0.018	0.011	0.012

The estimated contribution of changes in average industry wage effects on the real hourly wage change of each demographic group can be separated as two parts; the part due to changes in the industry weight ($\sum_j M_j \Delta\phi_{ij}$), and other part due to changes

[10] To compare with the results of Bound and Johnson (1992), there are some different in the wage effects of 1979 and 1988. For example, the direction of wage effects have opposite sign at the 7 industries (transport, utilities, retail trade, finance, hospitals, education, and professional services). This difference may cause a different results from Bound and Johnson's results.

in industry wage effects ($\sum_j(\phi_{ij} + \Delta\phi_{ij})\Delta M_j$). The values of the estimated contribution of the two sources of the changes in average industry wage effects to the relative wage changes for 1979-1988 and 1988-1997 are reported in Table 4.5. Columns (i) and (iv) give the relative wage changes for 1979-1988 and 1988-1997, (ii) and (v) give the effects of changes in industry weights, and (iii) and (vi) give the effects of industry wage effects. For example, the 0.151 proportional increases in the relative wage of male-white college graduates relative to high school graduates during the 1980's, can be attributed to differential movements between high- and low-wage industries, 0.001, and can be attributed to changes in industry wage effects, 0.001. All estimated changes do not explain a very large part of any the relative wage changes.

Table 4.5. Estimated Effects of Changes in Industry Wage Effects

Comparison Groups	Sex	Relative Wage Change	Industry Effects	
			Weights	Wages
A. 1979-1988:		(i)	(ii)	(iii)
College/high school	Male	0.151	0.001	0.002
	Female	0.125	-0.002	0.000
High school/dropout (X<30)	Male	0.048	0.001	0.000
	Female	0.043	0.001	0.000
Old/young (noncollege)	Male	0.036	0.001	-0.002
	Female	0.021	-0.001	-0.001
Women/men		0.097	0.000	-0.002
B. 1988-1997:		(iv)	(v)	(vi)
College/high school	Male	0.038	-0.001	-0.002
	Female	0.051	-0.001	-0.004
High school/dropout(X<30)	Male	0.035	-0.002	-0.001
	Female	0.089	0.000	-0.002
Old/young(noncollege)	Male	-0.024	-0.001	-0.003
	Female	0.023	-0.001	-0.003
Women/men		-0.020	-0.001	-0.001

For the 1990's the all net average industry effects are negative. From Table 4.5 we can see that the wage change of 1980's and 1990's are not much explained by changes in the industries wage structure.

4.5.2. Changes in the Structure of Product Demand

Changes in structure of product can shift the relative labor-demand function or different groups. The basic assumption is that the production function is CES (constant elasticity of substitution) function and the economy is at full employment. The CES assumption statement is that output of each J industries (Q_j) depends on employment if each of the I demographic groups (N_{ij}) according to the CES (constant elasticity of substitution) function

(4.20) $$Q_j = a_j \left[\sum_i \delta_{ij} (b_{ij} N_{ij})^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}$$

where b_{ij} is an index of the technological efficiency of group-i workers in industry j, a_j is a parameter representing the (neutral) technological efficiency of the industry and the effect of capital intensity, and σ is the elasticity of intrafactor substitution, which is assumed to be equal across industries. The full employment economy is that the effective labor force of each group is allocated among the J industries, $N_i = \sum_j N_{ij}$.

The relative demand for the output of industry j relative to some reference industry r is assumed to be

(4.21) $$Q_j / Q_r = \theta_j P_j^{-\varepsilon} \qquad j \neq r$$

where P_j is the price of Q_j relative Q_r, θ_j is an exogenous parameter reflecting consumer tastes and other factors (such as foreign competition) relative to good r, and ε

is the absolute price elasticity of product demand for each industry. The marginal conditions for each industry are given by

(4.22) $$P_j \partial Q_j / \partial N_{ij} = P_j a_j \delta_{ij} b_{ij}^{1-1/\sigma} (Q_j / N_{ij})^{1/\sigma} = W_{ic}.$$

The share of total group-i employment in industry j is

(4.23) $$\phi_{ij} = \delta_{ij}^{\sigma} (b_{ij} / b_i)^{\sigma-1} x_j / D_i$$

where b_i is the average value of the technological-efficiency parameter for group-i workers across industries and

(4.24) $$\begin{aligned} &(a) \quad D_i = \sum_j \delta_{ij}^{\sigma} (b_{ij} / b_i)^{\sigma-1} x_j \\ &(b) \quad x_j = a_j^{\sigma-1} \theta_j^{\sigma/\varepsilon} Q_j^{1-\sigma/\varepsilon}. \end{aligned}$$

The ratio of the competitive wage for group-i workers to that of some other group s is

(4.25) $$W_{ic} / W_{sc} = (b_i / b_s)^{1-1/\sigma} (D_i / D_s)^{1/\sigma} (N_i / N_s)^{-1/\sigma}$$

where D_i is an index of the effects of the θ_j's, a_j's, and Q_j's, and proportional changes in its values are referred to as a "product-demand-shift index". Holding constant the variables that affect W_{sc}, the total logarithmic derivative of (4.25) is

(4.26) $$d(\ln W_{ic}) = (1 - 1/\sigma) d(\ln b_i) + (1/\sigma) d[\ln(D_i / N_i)].$$

The estimation in the demand-shift variable in (4.26), $d(\ln D_i)$, which reflects changes in the θ_j's, and a_j's. Total differentiation of (4.24.a) yields the product-demand-shift index

(4.27) $$d(\ln D_i) = \sum_j \phi_{ij} d(\ln x_j)$$

for $\sum_j \phi_{ij} d[\ln(b_{ij} / b_i)] = 0$. The $d(\ln x_j)$'s are not directly observed, but the total derivative of (4.23) is

$$d(\ln \phi_{ij}) = (1-\phi_{ij})d(\ln x_j) - \sum_{k \neq j} \phi_{ik} d(\ln x_k) + (\sigma - 1)[d(\ln(b_{ij}/b_i))] \tag{4.28}$$

which may be rewritten in matrix form as equation (4.29)

$$\begin{bmatrix} d(\ln \phi_{11}) \\ d(\ln \phi_{12}) \\ \vdots \\ d(\ln \phi_{IJ}) \end{bmatrix} = \begin{bmatrix} 1-\phi_{11} & -\phi_{12} & \cdots & -\phi_{1J} \\ -\phi_{11} & 1-\phi_{12} & \cdots & -\phi_{1J} \\ \vdots & \vdots & \ddots & \vdots \\ -\phi_{I1} & -\phi_{I2} & \cdots & 1-\phi_{IJ} \end{bmatrix} \begin{bmatrix} d(\ln x_1) \\ d(\ln x_2) \\ \vdots \\ d(\ln x_J) \end{bmatrix} + \begin{bmatrix} d[\ln(b_{11}/b_1)] \\ d[\ln(b_{12}/b_1)] \\ \vdots \\ d[\ln(b_{IJ}/b_I)] \end{bmatrix} \tag{4.29}$$

In the absence of any information about the pattern of industry/group-specific technical change, the $d[\ln(b_{ij}/b_i)]$'s are treated as an error term, and the $d(\ln x_j)$'s may be estimated by OLS and then substituted back into (4.27) to obtain estimates of the product of the product-demand-shift index.

One factor of the influence of product demand shifts on relative labor-demand functions is the average employment growth by industry weighted by the initial employment distribution of each demographic group (Bound and Johnson, 1992). That is

$$EMP_i = \sum_j \Delta(\ln \phi_j)\phi_{ij} \tag{4.30}$$

where $\Delta(\ln \phi_j)$ is the proportionate change in the logarithm of industry *j*'s share of aggregate employment in each period. The calculated value of $\Delta(\ln \phi_j)$ for each of 17 major industries for 1979-1988 and 1988-1997 are reported under $\Delta\phi$ in columns (i) and (iv) of Table 6. EMP_i can be considered as a rough proxy for the discrete version of $d(\ln D_i)$ in equation (4.19). The assumption of the demand-shift explanation of the wage-structure changes of the 1980's is that the EMP_i's were in the right direction (i.e., toward more-educated, older, and female labor) whereas it is not clear for the 1990's.

An alternative approach that gets around this possible bias is to estimate the demand-shift indexes by industry for the two periods. To estimate demand-shift factors by industry for the two periods, there is a useful equation,

$$d(\ln D_i) = \sum_j \phi_{ij} d(\ln x_j), \tag{4.31}$$

and the results of estimated demand-shift indexes by industry are reported under Δx_1 in columns (ii) and (iv) of Table 4.6. Because the estimated growth in each industry as a deviation from the weighted rates of growth across demographic groups of its initial employment distribution, $\Delta(\ln x_j)$ is greater than $\Delta(\ln \phi_j)$ in industries (like durables/mining, nondurables, and transfort) that tended to hire low-educated and male labor and smaller in industries (like finance) that have the opposite demographic composition.

Table 4.6. Proportionate Employment Changes ($\Delta(\ln \phi_j)$) and Derived Demand Indexes ($\Delta(\ln x_j)$)

Industry	1979-1988		1988-1997	
	$\Delta\phi$	Δx_1	$\Delta\phi$	Δx_1
Construction	0.108	0.340	0.024	0.157
Durables/mining	-0.184	0.079	-0.079	0.074
Nondurables	-0.114	0.163	-0.101	0.031
Transport	0.152	0.391	0.004	0.126
Utilities	-0.007	0.173	-0.155	-0.127
Wholesale trade	0.039	0.299	0.028	0.117
Retail trade	0.107	0.335	0.159	0.259
Finance	0.108	0.229	0.024	-0.009
Business services	0.687	0.750	0.172	0.262
Personal services	0.259	0.481	0.017	0.167
Entertainment	0.160	0.354	0.506	0.460
Medical	0.143	0.386	0.212	0.318
Hospitals	-0.085	0.057	-0.383	-0.380
Welfare	-0.081	0.017	-0.177	-0.259
Education	-0.201	0.019	-0.116	-0.050
Professional services	0.354	0.512	-0.039	-0.358
Public administration	-0.143	-0.018	-0.038	-0.028

The calculation of a derived demand-shift index, $DEM_i = \sum_j \Delta(\ln x_j)\phi_{ij}$, which is analogous to the calculation of the EMP_i index in equation (4.31) is the next step. The average values of the alternative demand-shift indexes for the demographic groups are reported in columns (ii) and (vi) of Table 4.7 under *EMP* for the index based on industry employment changes and in columns (iii) and (vii) under *DEM* for the index based on estimated derived demand changes. For example, the proportional change in the relative supply of male college graduates relative to high school graduates for male-white [from column (vi)] was 0.011, which implies that their relative wages should have decreased by $(1/\sigma)\times0.011$. From Table 4.7, what is happening in the two decades overall is that the traditional employments of males with lower education were declining and employments of female with lower education were increasing, while higher educated workers both male and female were increasing during 1980's but decreasing during 1990's. Use of the *EMP* index of relative-demand changes suggests that $(1/\sigma)\times0.001$ of the failure of this relative wage to fall could be accounted for by product demand shifts. However, use of the *DEM* index only deepens the ambiguousness, for that index suggests that demand shifts were on balance slightly unfavorable to highly educated labor.

4.5.3. Intra-industry Employment Shifts

The major wage-change phenomena of the 1980's and 1990's are not adequately accounted for by explanations based on institutional factors or changes in the structure of product demand from the results thus far. Another possibility of the wage change in the 1980's and 1990's were characterized by major changes in technology that were nonneutral with respect to different types of labor. Variations across demographic

groups in the ceteris paribus effects on wages of technical change or changes in average group quality are reflected in the $(1-1/\sigma)d(\ln b_i)$ term in (4.19). Given the maintained assumption that $\sigma > 1$, the wage-structure facts are attributable to this set of explanations if the relative values of the b_i's for the more educated, older, and female demographic groups increased during the 1980's and mid-1990's.

Table 4.7. Proportionate Supply Changes, Alternative Product-Demand-Shift Indexes, and Specific-Industry Technical Change by Aggregated Groups

Comparison Groups	Sex	Supply	Demand-Change indexes		SPEC	GEN
			EMP	DEM		
A. 1979-1988:		(i)	(ii)	(iii)	(iv)	(v)
College/high school	Male	0.007	-0.008	0.022	0.046	0.098
	Female	0.008	-0.083	0.097	-0.056	0.094
High school/dropout	Male	0.005	-0.012	-0.016	0.092	0.005
	Female	0.004	0.037	0.027	0.009	0.064
Old/young	Male	-0.002	-0.038	-0.036	-0.140	0.016
	Female	0.000	-0.041	-0.041	-0.004	-0.101
Women/men		0.005	0.008	-0.007	-0.010	0.018
B. 1988-1997:		(vi)	(vii)	(viii)	(ix)	(x)
College/high school	Male	0.004	-0.026	-0.062	-0.083	0.097
	Female	0.010	-0.079	-0.113	-0.016	0.092
High school/dropout	Male	-0.005	-0.018	-0.018	0.092	0.041
	Female	-0.008	0.024	0.016	0.036	0.065
Old/young	Male	0.008	-0.041	-0.039	-0.021	0.015
	Female	0.009	-0.050	-0.055	0.048	-0.102
Women/men		0.000	0.005	-0.017	0.047	0.017

It has been argued that the relative demand for highly educated workers in periods of rapid technical change may increase because of their superior ability to adapt to and refine new methods of production. The 1980's and 1990's, as well as the 1970's to some extent, have been characterized popularly as a period in which computer technology was adopted throughout most of the U.S. economy, and changes in production methods have been favorable to professional and technical workers relative

to blue-color workers. To the extent that this technical change was common across most industries in the economy, we would expect that the $d(\ln b_i)$'s for certain demographic groups would have risen relative to others.

If it were true that the rate of growth of the technical-efficiency parameter for group-*i* workers in industry *j*, $d(\ln b_{ij})$, were equal to the weighted mean for that group, $d(\ln b_i)$, plus a random error, all that is predicted by the model is that the relative wages of those groups most favored by the technical change would rise. Usually changes in the b_i's are not observed, so cannot test this. However, if the effects of innovation in the relative demand for skilled labor may vary across industries, some of the variation in the values of $d(\ln b_i)$ can be identified.

Suppose that there is a subset of industries, J', in which the rate of growth of the efficiency parameters for a subset of the demographic groups, I', differs from their average growth in other industries. Specifically,

$$d(\ln b_{ij}) = \begin{cases} c_{0i} + c_{1i} & i \text{ in } I' \text{ and } j \text{ in } J' \\ c_{0i} & \text{otherwise.} \end{cases} \tag{4.32}$$

This implies that the average change across all industries in the efficiency parameters for a group in I' is $d(\ln b_i) = c_{0i} + c_{1i}T_i$, where T_i is the proportion of group *i*'s employment that is in the J's industries. To estimate the extent of this group/industry specific technical change, note that

$$d[\ln(b_{ij} / b_i)] = c_{1i} D_{I'} (D_{J'} - T_i) \tag{4.33}$$

where $D_{I'}$ and $D_{J'}$ are dummy variables for the relevant groups and industries. The equation (4.33) should be substituted for $d[\ln(b_{ij} / b_i)]$ in the regression equation (4.29)

to estimate industry demand shifts, and the coefficient on this variable is an estimate of $(\sigma-1)c_{1i}$ for each characteristic group-*i* and it is in Appendix Table A.4. The resultant proportional change in the average value of b_i for each group can then be calculated as $d(\ln b_i) = c_{0i} + c_{1i}T_i$. In other words, the change in the average-efficiency parameter for group-*i* workers equals a general component c_{0i}, which applies to all industries, plus a specific component $c_{1i}T_i$, which applies only to certain industries.

Bound and Johnson (1992) aggregated four of the five traditional blue-collar industries (durables/mining, nondurables, transportation, and public utilities) into *J'* sector. The group characteristics that were selected to be included as dummy variables included the four educational groups and the four experience groups separately for male and female, and the four experience groups for those who had not complete college separately for male and female, and gender itself. The summary values of $(\sigma-1)c_{1i}T_i$ for the aggregated comparison groups are reported under SPEC in columns (iv) and (ix) of Table 4.7. The intra-industry shifts represented by the variable were an important factor of the relative demand shifts for high school/dropouts but they were unfavorable for women relative to men in the aggregate. The negative effect of experience on the employment shifts of workers who had not finished college is opposite sign from the previous research.

4.5.4. General Technical Changes

Variation in the other component of the proportionate change in the average-efficiency parameter for each group, c_{0i}, is not directly observable. The addition of relative specific technical change to the other explanations does not add to outweigh the

perverse effects of relative supply changes for the 1980's. Now, we estimate c_{0i} for the remaining explanation.

Following (4.19), the per annum growth of the relative wage of group-i workers over each period (t=1 for 1979-1988 and t=2 for 1988-1997) may be written as

$$(4.33) \qquad dY_{ai}(t) = -(1/\sigma)dN_{ai}(t) + (1-1/\sigma)c_{0i}(t)' + u_i(t)$$

where $dY_{ai}(t)$ is the annualized proportionate change in the relative wage of group-i workers adjusted for the change in total average industry wage effects, $dN_{ai}(t)$ is the per annum proportionate change in relative supply adjusted for product demand shifts and industruy-specific technical change, $c_{0i}(t)'$ is the per annum value of general technical change, and $u_i(t)$ is a random error term. It then follows that the difference between the rates of growth of adjusted relative wages in the two periods is

$$(4.34) \qquad d^2Y_{ai} = -(1/\sigma)d^2N_a + (1-1/\sigma)[c_{0i}(2)' - c_{0i}(1)'] + u_i,$$

where $d^2Y_{ai} = dY_{ai}(2) - dY_{ai}(1)$, $d^2N_{ai} = dN_{ai}(2) - dN_{ai}(1)$, and $u'_i = u_i(2) - u_i(1)$.

We specify that the growth in each group's efficiency parameter relating to all industries in the 1988-1997 period equals its value in the 1979-1988 period plus a difference A_i, that is,

$$(4.35) \qquad c_{0i}(2)' = c_{0i}(1)' + A_i\,.$$

It is assumed initially that A_i is uncorrelated with either $c_{0i}(1)'$ or d^2N_{ai}, which is equivalent to assuming that the pattern of general technical change in the two periods was identical. If these assumptions are correct, the reciprocal of the elasticity of intrafactor substitution can be estimated by regressing d^2Y_{ai} on d^2N_{ai}, for the influence of general technical change disappears as a fixed effect. The estimated slope

coefficient of a regression (weighted by $[k_i(1988)]^{0.5}$) of d^2Y_{ai} on d^2N_{ai} is -0.558 (SE = 0.350), which implies a value of σ = 1.793. It is then possible to obtain an estimate of the effect of general technical change on the per annum growth rate of group-i workers, $(1-1/\sigma)c'_{0i}$ by computing the average of the residuals, $dY_{ai}(t)+(1/\sigma)dN_{ai}(t)$, over the two periods.

There is a possibility that the pace of general technical change for some groups may have raised or fallen from 1980's to the 1990's (i.e., that certain A_i's were not zero). We added five young, low education groups dummy variables (that is, both men and women dropouts and high school graduates in the lowest experience interval and male dropouts with X=10-19) for equation (4.34). Inclusion of this dummy variable yielded an estimated coefficient on d^2N_{ai} of –0.499 (SE = 0.427). The results of estimation are in the Table A.5. Since all dummy variables are not statistically significant, the test for the possibility that the pace of general technical change for some groups may change from the 1980's to 1990's are rejected. The results are in the Table A.5. With the estimated common values of $(1-1/\sigma)c'_{0i}$, the estimated effects of general technical change on relative wage changes are $GEN_i(1)=9\times(1-1/\sigma)c'_{0i}$ for 1979-1988 period and $GEN_i(2)=9\times(1-1/\sigma)c'_{0i}$ for the 1988-1990 period. The average relative values of these estimates for the summary comparison groups are reported in Table 4.7.

It is apparent from inspection of the estimates of the estimated values of GEN for the 1990's that our major conclusion is that the principal cause of the significant wage-structure changes of the past two decade was a shift in the structure of the b_i's that

were extremely favorable to certain groups. The average value of GEN of women relative to men during the 1980's and 1990's were 0.018 and 0.017, which means that the other things held constant, women's wages grew faster than men's wages by about 1.7 percent. The large negative values of GEN for female-younger workers with low levels of education during two decades. The possible explanation of this is that the workers with low levels of education who entered the labor market had a much lower level of innate ability than their older counterparts who entered the labor market in the 1970's or is that they were the most susceptible to competition from undocumented immigrants (Bound and Johnson, 1992).

Table 4.8. Decomposition of Estimated Sources of 1979-1988 and 1988-1997 Relative Wage Changes

Comparison groups	sex	Relative wage change	Source of relative wage change				
			Rents	Supply	Demand	Technical change	
						Specific	General
A.1979-88:		(i)	(ii)	(iii)	(iv)	(v)	(vi)
CO/HS (X<30)	M	0.151	0.003	0.007	0.022	0.046	0.098
	W	0.125	-0.002	0.008	0.097	-0.056	0.094
HS/DO (X < 30)	M	0.048	0.001	0.005	-0.016	0.092	0.005
	W	0.043	0.001	0.004	0.027	0.009	0.064
Old/young (noncollege)	M	0.036	-0.001	-0.002	-0.036	-0.140	0.016
	W	0.021	-0.002	0.000	-0.041	-0.004	-0.101
Women/men		0.097	-0.002	0.005	-0.007	-0.010	0.018
B. 1988-97:		(viii)	(ix)	(x)	(xi)	(xii)	(xii)
CO/HS	M	0.038	-0.003	0.004	-0.062	-0.083	0.097
	W	0.051	-0.005	0.010	-0.113	-0.016	0.092
HS/DO (X<30)	M	0.035	-0.003	-0.005	-0.018	0.092	0.041
	W	0.089	-0.002	-0.008	0.016	0.036	0.065
Old/young (noncollege)	M	-0.024	-0.004	0.008	-0.039	-0.021	0.015
	W	0.023	-0.004	0.009	-0.055	0.048	-0.102
Women/men		-0.02	-0.002	0.000	-0.017	0.047	0.017

To summarize the results, the proportionate relative wage change of each demographic group equals the change in total industry wage effects plus the effects of

relative supply changes, product demand shifts, and average technical change, which separated into that arising in specific industries and in general by equation (4.19). Estimates of the contribution of each of these effects for the comparison groups are reported for 1979-1988 in columns (ii)-(vi) of Table 4.8 and for 1988-1997 in columns (ix)-(xiii). The decompositions for the 1980's and 1990's suggest an easy explanation. First, total changes in average industry wage effects were in the right direction but accounted for a small fraction of relative wage changes. Second, for the comparisons involving education and gender, relative supply changes were large and in the wrong direction. Third, the estimates of the effects of product demand shifts on relative wages are small and of uneven direction. Fourth, the two forms of technical change, SPEC and GEN, comprise the principle source of the increase in educational differentials and the decrease in the gender differential, and the large positive values of SPEC for older noncollege workers account for a large amount of the increase in their relative wages.

4.6. Conclusions

We use semiparametric estimation methods to examine changes in the earnings distribution among 1979, 1988, and 1997. The semiparametric regression method used in this research substantiates results reported in the stylized facts of earning profiles.

There are some conclusions:

- A comparison of semiparametric and parametric specifications indicates that semiparametric estimation methods capture nonlinearities of the experiences in the earnings profiles.
- Earnings inequality continued to increase through 1997 between and within groups defined by schooling and experience.

- The mean wage conditional on schooling and experience has declined since 1979 among almost all groups defined by schooling and experience with a few exceptions.
- Workers with more than 16 years of schooling were better off in real wage terms in 1997 than similar workers in 1988 for all demographic group workers.
- Between 1988 and 1997 the relative return to schooling and experience is increasing.
- Increases in between-group inequality result mostly from a decrease in wages for workers with 12 years of schooling. Real wages of workers with 16 years of schooling are not significantly different between 1988 and 1997, while real wages for workers with more than 16 years of schooling increased. In addition, the increase in between-group inequality is concentrated among less-experienced workers.
- The estimates of residual inequality have a heterogeneous increase in within-group earnings inequality.
- Earnings inequality within schooling and experience groups changed at differing rates among different groups. Inequality increased the most among less experienced workers with 14 years and more of schooling and more experienced workers with 12 years or less of schooling.

Bound and Johnson (1992) explains the increase in inequality of earning profile as skill-biased technological change that increases the demand for workers with more schooling and experience. The between-group results are consistent with an increase in demand for skilled workers. Our study shows that the increase in demand for skilled

workers, measured by the increase in between-group inequality, occurred for less-experienced workers. We have attempted to evaluate the evidence concerning several alternative explanations of the dramatic wage-structure developments in the U.S. during the 1980's and 1990's. Our analysis points to the conclusion that the major reason for the increases in wage differentials by educational attainment and the decrease in the gender differential is a combination of skilled-labor-biased technical change and changes in unmeasured labor quality.

Our estimations have some different results from Bound and Johnson's results that are not explained and it is beyond of this paper. It may come from different data sets and different estimation method.

CHAPTER 5

SEMIPARAMETRIC TWO-STAGE ESTIMATION OF SIMULTANEOUS PROBIT MODEL

5.1. Introduction

A general model of simultaneous equations model in terms of latent variables is

$$(5.1) \qquad By_t^* + \Gamma X_t = u_t,$$

where u_t has zero mean and covariance matrix Σ, y_t^* is the $G\times 1$ vector of 'latent' endogenous variables, X_t is a $K\times 1$ vector of exogenous variables, B is $G\times G$ nonsingular matrix , and Γ is $G\times K$ matrix. Also, define $E(u_t u'_t) = \Sigma$. The identification problems in this model are the same as those in the usual simultaneous equations model, except for the fact that the parameters (B, Γ, Σ) are estimable only up to certain scale factors because some (or all) elements of y_t^* are observed as qualitative variables. Consider the simultaneous discrete choice model; with G = 2,

$$(5.2) \qquad \begin{aligned} y_{1i} &= 1, \quad \text{if } y_{1i}^* > 0; \quad y_{1i} = 0, \quad \text{if } y_{1i}^* \le 0 \\ y_{2i} &= 1, \quad \text{if } y_{2i}^* > 0; \quad y_{2i} = 0, \quad \text{if } y_{2i}^* \le 0 \end{aligned}$$

where

$$(5.3) \qquad \begin{aligned} y_{1i}^* &= \beta_{12} y_{2i}^* + x'_{1i}\gamma_1 + u_{1i}; \\ y_{2i}^* &= \beta_{21} y_{1i}^* + x'_{2i}\gamma_2 + u_{2i}, \quad i = 1,\cdots,N. \end{aligned}$$

This model is considered by Mallar (1977). For the estimation of this model, the Maximum Likelihood estimation (MLE) method can be used. If u_{1i} and u_{2i} are independent, then one can estimate both equations separately by the probit ML method. If u_{1i} and u_{2i} are not independent, this method does not give consistent estimates of the parameters. The two-stage estimation method can be applied for this model.

There are many applications of the simultaneous discrete choice model which involves two binary dependent variables associated with two latent variables, each in turn depending on two set of exogenous variables.

In general, maximum likelihood estimation method is the most popular estimation method in microeconometrics. If the model is specified correctly, the ML method yields consistent, asymptotically normal and efficient estimators. However, the correct specification may not be known beforehand. There are two major sources of misspecifications: incorrect specification of the functional form of the relationship under study (for example, omitting exogenous variables or misspecification of the functional form) and misspecification of the stochastic structure of the model (for example, neglecting heteroscedasticity or misspecification of the distribution of the random variables). The maximum likelihood estimator is generally inconsistent in both cases. Recently, Carrasco (2001) investigated the impact of controlling for unobserved heterogeneity in the binary choice model in panel data. Manski (1975) suggested the semiparametric methods for specific microeconometric models. These models do not require complete distributional assumptions or less restrictive distributional assumptions than the assumption of normality. These methods yield consistent estimator of the parameters of interest without a complete specification of the distribution of the stochastic variables in the model. A recent survey of methods available is Pagan and Ullah (1999). Amemiya (1978) suggested applying minimum distance estimation methods (MDE) to limited dependent variable simultaneous models where the reduced form (RF) is estimated with MLE. Newey (1985) applied semiparametric methods to a RF model. Lee (1995) extended Newey's approach.

Our approach for simultaneous probit model is to use the two-stage maximum likelihood (ML) estimation with semiparametric method. There are many semiparametric procedures for the univariate probit model. Klein and Spady's (1993) method uses the ML type estimation. Klein and Spady's estimator attains the semiparametric efficiency bound. These semiparametric estimations for index functions permit multiplicative heteroscedasticity of a general but known form and heteroscedasticity of an unknown form if it depends only on the index. Section 2 describes the semiparametric simultaneous probit model. Section 3 presents a numerical illustration and section 4 is conclusion.

5.2. Semiparametric Estimation of Simultaneous Probit Model

In the parametric simultaneous probit model, the reduced forms of model (5.3) is

$$\begin{aligned} y_1^* &= \Pi_1 X + \varepsilon_1 \\ y_2^* &= \Pi_2 X + \varepsilon_2. \end{aligned} \tag{5.4}$$

If $Var(\varepsilon_1) = Var(\varepsilon_2) = 1$, without loss of generality, we can estimate Π_1 and Π_2 by probit ML method. Then substitute the predicted values of $E(y_2^*)$ and $E(y_2^*)$, and estimate the structural equations. Define $\alpha'_1 = (\beta_{12}, \gamma'_1)$, and $\alpha'_2 = (\beta_{21}, \gamma'_2)$. Let $a_1 = \frac{\phi_1}{\Phi_1(1-\Phi_1)}$, $a_2 = \frac{\phi_2}{\Phi_2(1-\Phi_2)}$, where ϕ_i, Φ_i are the standard normal pdf and cdf, respectively. Also define

$$A_1 = \phi_1 a_1,\ A_2 = \phi_2 a_2,\ Z = \begin{bmatrix} \Pi_2^* X \\ X \end{bmatrix},$$

$$W_1 = \frac{1}{N}\sum_1^N A_1 ZZ',\ W_2 = \frac{1}{N}\sum_1^N A_2 XX',\ W_3 = \frac{1}{N}\sum_1^N A_1(\beta_{12}) ZX',$$

$$W_4 = \frac{1}{N}\sum_1^N a_1 a_2 E[(y_1 - \Phi_1)(y_2 - \Phi_2)]XZ'.$$

Then the covariance matrix of $N^{1/2}(\hat{\alpha}_1 - \alpha_{01})$, where α_{01} is the true value of α_1 and $\hat{\alpha}_1$ is the two-stage estimator is $W_1^{-1}[W_1 - W_3W_2^{-1}W_4 - W_4'W_2^{-1}W_3' + W_3W_2^{-1}W_3']W_1^{-1}$ (Maddala, 1983). The covariance matrix of $\hat{\alpha}_2$ will be similar expression, with the subscripts 1 and 2 interchanged in the definitions of Z, W_1, W_2, W_3, and W_4.

To get our two-stage semiparametric estimator, we will use the consistent parametric estimator in the first stage. Applying the probit MLE to get the reduced form estimates $\hat{y}_{1i}^*$ and $\hat{y}_{2i}^*$. The second stage estimator of equation (5.4) is

$$
\begin{aligned}
y_{1i}^* &= \beta_{12}\hat{y}_{2i}^* + x'_{1i}\gamma_1 + u_{1i}; \\
y_{2i}^* &= \beta_{21}\hat{y}_{1i}^* + x'_{2i}\gamma_2 + u_{2i},
\end{aligned}
\tag{5.5}
$$

The second stage of each equation is to apply the semiparametric maximum likelihood estimation of Klein and Spady (1993).

Klein and Spady suggested semiparametric estimation for the single binary choice dependent variable using maximum likelihood estimation. Consider the general form of binary choice model given by

$$
y = \begin{cases} 1 & \text{if } v(x;\theta_0) \geq u_0 \\ 0 & \text{otherwise} \end{cases}, \tag{5.6}
$$

where $v(\cdot;\cdot)$ is a known function, x is a vector of exogenous variables, θ_0 an unknown parameter vector, and u_0 a random disturbance. It is assumed that observations $\{x_i, y_i\}$ are i.i.d. and that the model satisfies $E(y \mid x) = E[y \mid v(x;\theta_0)]$, where E is the indicated conditional expectation and $v(x;\theta_0)$ is an aggregator or index. The most common way to estimate θ_0 in (5.6) is to apply the method of maximum likelihood. When the disturbances are independently distributed according to a known conditional parametric

distribution, the log-likelihood is

$$
(5.7) \qquad \begin{aligned} L &= \sum_{i=1}^{n}\left[y_i \ln(P_i^*(\theta)) + (1-y_i)\ln(1-P_i^*(\theta))\right] \\ P_i^*(\theta) &\equiv P^*[v(x_i;\theta);\theta] = \Pr[u < v(x;\theta) \mid v(x;\theta)] = F_{u|x}[v(x;\theta)] \end{aligned}.
$$

Once the probability function is replaced with its maximum likelihood estimator, the resulting concentrated likelihood is not a smooth function of θ and it is difficult to establish the asymptotic distribution for the estimator of θ. To circumvent this problem, a semiparametric likelihood that is a smooth function of θ and that locally approximates the corresponding parametric likelihood was proposed by Klein and Spady. To construct $P_i(\theta)$, with C as the event $[u < v(x;\theta)]$, P^* in (5.7) is equivalent to $\Pr[C \mid v(x;\theta)]$, the probability of the event C conditioned on $v(x;\theta)$. This probability function is characterized by

$$
(5.8) \qquad P^*[v(x;\theta);\theta] \equiv \Pr[C \mid v(x;\theta)] = \Pr[C] g_{v|C}(v;\theta) / g_v(v;\theta),
$$

where Pr(C) is the unconditional probability of C, and $g_{v|C}$ is the density for $v = v(x;\theta)$ conditioned on C, and g_v is the unconditional density for v. Let C_0 be the event C at θ_0: $[u < v(x;\theta)]$, which is observable and is equivalent to the event $y = 1$. If C is replaced with C_0 in equation (5.8), we can obtain the probability function $P(v;\theta)$, i.e. $P(v;\theta)$ is the probability of the event C_0 conditioned on $v = v(x;\theta)$:

$$
(5.9) \qquad \begin{aligned} P[v(x;\theta);\theta] &\equiv \Pr[C_0 \mid v(x;\theta] = \Pr[C_0] g_{v|C_0}(v;\theta) / g_v(v;\theta) \\ &= \Pr[y=1] g_{v|y=1}(v;\theta), \quad P_i \equiv P(v_i;\theta) \end{aligned},
$$

where $g_{v|y=1}$ is the density for v conditioned on $y = 1$. At $\theta = \theta_0$ and under the index restriction, (5.9) is equivalent to the true choice probability. With $v_0 \equiv v(x;\theta_0)$,

$P(v_0,\theta_0) = \Pr[C_0 \mid v(x;\theta_0) \equiv \Pr[y=1 \mid v(x;\theta_0)] = \Pr[y=1 \mid x]$. Therefore, the function P may be viewed as a local approximation to the corresponding parametric probability function. Although the probability function in (5.9) is unknown, it can be estimated directly as a smooth function of θ by estimating Pr[y = 1] as the sample proportion of individuals making the choice while both the conditional and unconditional densities can be estimated nonparametrically. If these densities are estimated by kernel methods with the same window being employed for conditional and unconditional densities, then the estimate of (5.9) will be the usual kernel estimate of the expected value of y conditioned on the index. Also, since v aggregates the information in the possibly high vector x into a scalar, it is needed only univariate density estimation. Using $\hat{P}_i(\theta)$ to denote $P_i(\theta)$ as defined in (5.9) the estimation of θ_0 is to choose θ to maximize the estimated quasi-likelihood:[1]

$$(5.10) \qquad Q(\hat{P}(\theta)) = \sum_{i=1}^{n}\left[y_i \ln[\hat{P}_i(\theta)^2] + (1-y_i)\ln[(1-\hat{P}_i(\theta))^2]\right].$$

Note for both y_i and $(1-y_i)$ terms, the argument of the log function is squared.

With some required condition, the estimated functions are sufficiently well-behaved. The following conditions assumed to hold:

(A.1) The data consist of a random sample (y_i, x_i), $i = 1,\dots,N$. The random variable y is binomial with realizations 1 and 0.

(A.2) There exists a scalar aggregator, $a(x;\theta_0)$, such that for any x: $\Pr[y=1 \mid x] \equiv E(y \mid x) = E[y \mid a(x;\theta_0)] \equiv \Pr[y=1 \mid a(x;\theta_0)]$.

[1] Klein and Spady indicated that the quasi-likelihood is more appropriately to formulate semiparametric likelihood.

(A.3) There exist $\underline{P}, \overline{P}$ that do not depend on x such that $0 < \underline{P} \leq \Pr(y = 1 \mid x) \leq \overline{P} < 1$, which serves to bound the probability function away form zero and one. With $H[v(x;\theta_0)] \equiv \Pr[u_0 < v(x;\theta_0)]$, assume that $H(t)$ is continuously differentiable in t for all t and that $| \partial H(t) / \partial t | < c$.

(A.4) The index v is smooth in that for θ in a neighborhood of θ_0 and all t:

$$\left\{ | D_\theta^{[r]} v(t;\theta) |, | \partial(D_\theta^{[r]} v(t;\theta)) / \partial x | \right\} < c \qquad (r = 0,1,2,3,4).$$

(A.5) Letting $g_{yv}(v_i;\theta) \equiv \Pr(y) g_{v|y}(v_i;\theta)$, define the kernel estimator for g_{yv} as

$$(5.11) \qquad \hat{g}_{yv}(v_i;\theta;h) \equiv \sum_{j \neq i}^{N} \frac{1\{y_j = y\}}{h} K\left[\frac{v_i - v_j}{h}\right] \Big/ (N-1) \qquad (y = 0,1),$$

where $1\{\cdot\}$ indicate the indicator function.

The kernel function, $K(z)$ is a symmetric function that integrates to one, has bounded second moment, and $\left\{ | D_z^r K(z) |, \int | D_z^r K(z) | dz \right\} < c$ $(r = 0,1,2,3,4)$, where c is a positive generic constant and $D_z^r f(z)$ denote the rth order partial of $f(z)$ with respect to z the parameter h is $h = h_N \hat{\sigma}_y(\theta)$ with a nonstochastic window, $N^{-1/6} < h_N < N^{-1/8}$ and $\hat{\sigma}_y$, the sample standard deviation of $v(x;\theta)$ conditioned on y. The kernel is also $\int z^2 K(z) dz = 0$ with bias reducing window h. This bias reducing kernel has the constant window at each sample size. The other method is adaptive kernel estimator with a variable and data dependent window size. Silverman (1986) obtains a uniform bias of order h_N^4 for the density estimated with local smoothing

Now, parameter estimates are obtained by maximizing an estimated quasi-likelihood function. The true probability function is given as

(5.12) $$P_i(\theta) \equiv g_{1v}(v_i;\theta)/g_v(v_i;\theta), \quad g_v \equiv g_{1v} + g_{0v}.$$

It would seem natural to define an estimated probability function by replacing all of these components with the corresponding kernel estimates given in (5.11).[2] With $\hat{g}_v \equiv \hat{g}_{1v} + \hat{g}_{0v}$ and an adjustment factors $\hat{\delta}_N \equiv \hat{\delta}_{0N} + \hat{\delta}_{1N}$, the estimated probability function is defined as $\hat{P}(v;\theta) \equiv [\hat{g}_{1v}(v;\theta,h) + \hat{\delta}_{1N}(v;\theta)]/[\hat{g}_N(v;\theta) + \hat{\delta}_N(v;\theta)]$. The purpose of the adjustment factors in the estimated probability is to control the rate at which numerator and denominator of estimated probability functions tend to zero. With $\hat{P}_i \equiv \hat{P}(v_i;\theta)$, it can be shown that $|\hat{P}_i(\theta) - P_i(\theta)|$ converges in probability, uniformly in i, θ to zero except for v_i in a region with vanishing probability. The quasi-likelihood converges to a function that is uniquely maximized at θ_0. The estimator that maximizes this objective function would then be consistent. Under the appropriate regularity conditions and with established consistency, $\hat{\theta}$ is distributed asymptotically as square root N normal with mean zero and variance-covariance matrix. The proof of consistency and normality for the estimator maximizing the estimated quasi-likelihood function is in Appendix of Klein and Spady (1993).[3] The gradient of the quasi-likelihood function,

(5.12) $$\sum \hat{p}(v(x_i,\hat{\theta})^{-1}\left[1-\hat{p}(v(x_i,\hat{\theta}))\right]^{-1}[\partial\hat{p}(v(x_i,\theta))/\partial\theta]\left[y_i - \hat{p}(v(x_i,\hat{\theta}))\right] = 0.$$

The covariance matrix of $n^{1/2}(\hat{\theta}-\theta_0)$ will be $N(0,V)$, where

---

[2] For technical difficulties with uniform convergence argument when estimated densities become too small, it is introduced an adjustment factors. $\hat{\delta}_{yN} \equiv h^a[e^z/(1+e^z)$, $\quad z \equiv [(h^b - \hat{g}_{yv}(v;\theta))/h^c$ where (a, b, c) are constant parameters. For detail, see Klein and Spady (1993).

[3] To prove the consistency, Klein and Spady used the trimmed quasi-likelihood function. However, they provide some simulation evidence on the properties of their estimator and found little difference between the performance of the trimmed and untrimmed estimators.

(5.13) $$V \equiv E\left\{\left[\frac{\partial P}{\partial \theta}\right]\left[\frac{\partial P}{\partial \theta}\right]'\left[\frac{1}{P(1-P)}\right]\right\}^{-1}_{\theta=\theta_0}$$ [4].

This is the equivalent of the variance of a parametric estimator where p_i replaces F_i. All the elements in V can be consistently estimated; p_i by $\hat{p}_i$ and the middle term either directly or with some robust estimator to take account of heterogeneity or dependence.

5.3. Numerical Experiment

5.3.1. Application on the Congressional Voting data

The first experiment is a semiparametric estimation of two-equation simultaneous probit model with the actual data. The actual data was previously used by Stratmann (1992) to analyze 'The Effects of Logrolling on Congressional Voting'. Stratmann sets a three-equation simultaneous probit model to analyze vote trading among congressmen for agricultural issues. Stratmann's data contains 406 observations on congressmen's votes cast in 1985 concerning the farmer's interests in different agricultural amendments to the Farm Bill.[5] For simplicity of calculation, we only use two equations, the PEANUT equation and SUGAR equation. The dependent variables are $y_{1i}=1$ if one vote for PEANUT amendment, $y_{2i}=1$ if one vote for SUGAR amendment and zero otherwise. The independent variables are constituency interests (PFA for peanut, and SFA for sugar), campaign contributions (PNCT for peanut and SGCT for sugar), party affiliation (PARTY), and ideological interests (ACU). The two-equation simultaneous probit model of congressional voting data follows:

[4] This asymptotic normality is proved by Klein and Spady.

[5] The data was kindly supplied by Professor Stratmann.

$$y_{1i} = 1, \quad \text{if } y_{1i}^* > 0; \quad y_{1i} = 0, \quad \text{if } y_{1i}^* \leq 0$$
$$y_{2i} = 1, \quad \text{if } y_{2i}^* > 0; \quad y_{2i} = 0, \quad \text{if } y_{2i}^* \leq 0,$$

where

$$\begin{aligned} y_{1i}^* &= \alpha_{12} y_{2i}^* + \beta_{10} + \beta_{11} ACU + \beta_{12} PFA + \beta_{13} PNCT + \beta_{14} PARTY + u_{1i} \\ y_{2i}^* &= \alpha_{21} y_{1i}^* + \beta_{20} + \beta_{21} ACU + \beta_{22} SFA + \beta_{23} SGCT + \beta_{24} PARTY + u_{2i} \end{aligned}.$$

For nonparametric density estimation we choose the Epanechnikov kernel $K(u) = (3/4\sqrt{5})(1 - \frac{1}{5}u^2)$ for $|u| < \sqrt{5}$ and bandwidth is $h = 2.34 stdc(u) n^{-1/5}$, where n is the number of sampled observations.

The two-stage probit MLE and semiparametric estimates of coefficients are shown in the Table 1. There is no constant term in the semiparametric estimates of coefficients.[6] Comparing the two sets of results, we find that almost of the parameter estimates are same sign with one exception although there are some of parameter estimates are far apart. The coefficients of the Contribution for PEANUT have different sign between probit ML estimates and semiparametric estimates. The predictive power of the models shows that the semiparameteric estimates are quite comparable.

Figure 1 shows estimates of $\hat{p}_i$ and derivative of $\hat{p}_i$ obtained from probit ML estimates of y_i's on the $v_i = x_i\hat{\beta}$. Epanechnikov kernel and constant bandwidth $h = 2.34 stdc(u) n^{-1/5}$ are used to draw the density function. The two top panel graphs show the p_i and dP/dv for the first equation and bottom panel graph show those of the second equation. As showed, dP/dv is not a unimodal. This contradicts the parametric model, which assumes that dP/dv is a unimodal (normal) pdf.

[6] Any constant terms are absorbed into the explanatory terms.

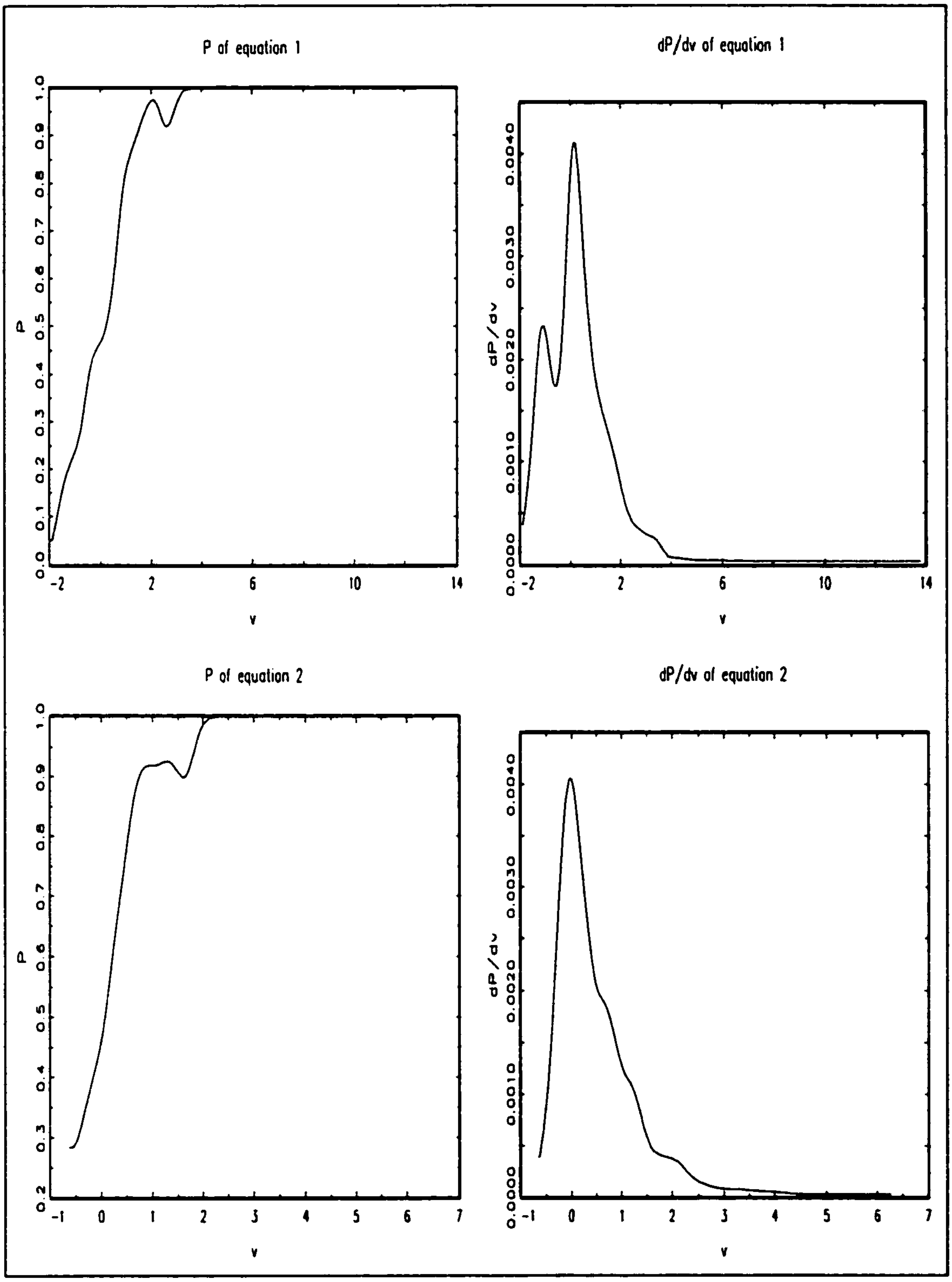


Figure 5.1. Estimates of p_i and dP/dv for the Each Equation

have more accurate predictive value for the first equation which has bimodal dP/dv, and less predictive value but not much par apart from the probit ML estimates for the second equation which has unimodal and almost normal shape. [7]

To gain additional insight into whether the semiparametric estimates reflect genuine features of the sampled population or are artifacts of our choices of bandwidths and other tuning parameters, we carried out a Monte Carlo experiment in which simulated data sets of size 100 were generated by sampling (Y, X) randomly without replacement from the congressional voting data. Each simulated data set is a random sample from the distribution that generated the congressional voting data, rather than from an assumed model that may not capture essential features of this distribution.

The last column of Table 5.1 shows the means and standard errors of the parameter estimates obtained in 400 Monte Carlo replications. The standard errors indicate the variability of the Monte Carlo estimates. The Monte Carlo parameter estimates are close to those obtained from the full data set of the first equation. The Monte Carlo estimates of coefficients are much closer to the full-data semiparametric estimate than to the probit ML estimates. Thus it appears that the main features of the semiparametric estimates are not artifacts of the choices of tuning parameters.

From this numerical experiment we studied two-stage maximum likelihood estimation of two simultaneous equation models with nonparametric density function. The results of semiparametric estimates are not quite better than the two-stage maximum likelihood estimation method, although we have the superior theoretical approach from Klein and Spady (1993).

[7] Horowitz and Hardle (1996) also found the similar bimodal pattern from their empirical work.

Table 5.1. Estimated Coefficients for Two-Equation Simultaneous Model

Equation	Variable name	Probit MLE Model	Semiparametric Model	Monte Carlo experiment
Equation of peanut	Constant	0.046 (0.008)		
	y_2^*	1.281 (0.008)	0.743 (0.002)	0.451 (0.181)
	ACU	0.434 (0.021)	0.286 (0.005)	0.250 (0.006)
	Farmer	2.678 (0.028)	0.061 (0.008)	0.171 (0.043)
	Contribution	-0.129 (0.019)	0.328 (0.005)	0.198 (0.108)
	Party	-1.290 (0.014)	-0.761 (0.004)	0.408 (0.079)
Predicted value	Actual value (220)	254	232	
Log likelihood		-211.035	-418.763	
Equation of milk	Constant	-0.059 (0.005)		
	y_1^*	0.527 (0.005)	2.855 (0.013)	0.508 (0.059)
	ACU	0.084 (0.013)	1.903 (0.031)	0.221 (0.014)
	Farmer	0.141 (0.025)	0.006 (0.067)	0.052 (0.048)
	Contribution	0.488 (0.006)	2.331 (0.015)	0.032 (0.146)
	Party	0.285 (0.012)	0.317 (0.031)	0.868 (0.077)
Predicted value	Actual value (264)	265	259	
Log likelihood		-206.199	-405.318	

Note: The number in the parenthesis reveals standard errors. The standard errors are computed by (5.13) for semiparametric estimates. For Monte Carlo the standard error is calculated by variation.

5.3.2. Monte Carlo Simulation

There are many aspects of estimator performance that we could study, and obviously not all can be considered at any one time. We study the finite sample properties of the simultaneous two-step probit MLE model when the model identification is correct. In this section we report the results that deal with three categories of questions, (1) a normal and χ^2 distributed explanatory variables design, (2) three different error correlation designs, (3) three different correlation between two equation, and (4) a homoscedasticity and heteroscedasticity design.

The numbers of sample of size n=100 that are drawn control the inherent sampling error of the Monte Carlo experiment. In the experiment we want the average variability of the parameter estimates to be close to the true variance of the estimator. We set the number of samples at 400. For simplicity, the identification condition is assumed to hold. That is, each equation has different explanatory variables that are independently and identically distributed. Our simulation equation, therefore, is:

$$\begin{aligned} y_{1i} &= 1, \quad \text{if } y_{1i}^{*} > 0; \quad y_{1i} = 0, \quad \text{if } y_{1i}^{*} \le 0 \\ y_{2i} &= 1, \quad \text{if } y_{2i}^{*} > 0; \quad y_{2i} = 0, \quad \text{if } y_{2i}^{*} \le 0, \end{aligned}$$

where

$$\begin{aligned} y_{1i}^{*} &= \alpha_{12} y_{2i}^{*} + \beta_{11} x_{11} + u_{1i} \\ y_{2i}^{*} &= \alpha_{21} y_{1i}^{*} + \beta_{22} x_{22} + u_{2i} \end{aligned}$$

The explanatory variable is generated from normal distribution truncated at ±2, standardized to have zero mean and unit variance by subtracting their means and dividing by standard deviation for the normal design. The explanatory variable for next design is chi-squared distribution with 3 degree of freedom truncated at 6 and

standardized similarly[8]. For the parameter vector, we set $\beta_{11} = \beta_{22} = 1$. The dependent variable is determined by drawing a uniform random number on the unit interval, v_i, where $v_i = x'_{1i}\gamma_{1i} + x'_{2i}\gamma_{2i}$ for *i*th equation, and assigning

$$\begin{aligned} y_i &= 1 \quad \text{if } v_i \in [0, P_i] \\ y_i &= 0 \quad \text{if } v_i \in (P_i, 1]\text{'} \end{aligned}$$

where $P_i = F(x'_i\,\beta)$ is the probability that $y_i = 1$. (Griffiths, Hill, and Pope (1987)).

The error terms have standard bivariate normal distribution for homoscedasticity design. For the heteroscedasticity design the error terms are bivariate normal with mean zero and variance $0.25(1+v_i^2)^2$. In all designs, we estimated densities with constant bandwidth, $h = n^{-1/7.5} stdc(v_i)$.

The computations were carried out in GAUSS using GAUSS pseudo-random number generators. The numerical optimization module (CO) of the GAUSS package was used to obtain the solutions, using Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.[9]

The estimation of reduced form is the probit ML estimation for each equation with all exogenous variables in both equations, and the structural form is estimated by quasi-maximum likelihood method for the semiparametric estimation and probit ML estimation for the two-stage probit ML estimation. The starting values are the OLS estimators of discrete dependent variable on explanatory variables for the reduced form estimation.

[8] Klein and Spady (1993) used this setup for semiparametric estimation of single probit model.
[9] The GAUSS optimization module is the minimization of the object function. We minimize the negative log-likelihood function.

Table 5.2. Monte Carlo Parameter Estimates, Homoscedasticity, Normal Design

Normal Design	Semiparametric						Probit MLE					
	ρ=0		ρ=0.25		ρ=0.75		ρ=0		ρ=0.25		ρ=0.75	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\alpha_{12} = \alpha_{21} = 0.25$												
α_{12}	-0.125	0.040	0.091	0.115	-0.058	0.067	0.007	0.073	-0.089	0.046	0.053	0.103
$\beta_{11} = 1$	-0.635	0.444	-0.284	0.289	-0.641	0.455	-0.237	0.251	-0.656	0.472	-0.391	0.277
α_{21}	-0.093	0.047	0.056	0.131	-0.079	0.049	0.072	0.108	-0.088	0.067	0.036	0.076
$\beta_{22} = 1$	-0.662	0.479	-0.342	0.286	-0.655	0.471	-0.246	0.219	-0.671	0.489	-0.419	0.281
$\alpha_{12} = \alpha_{21} = 0.5$												
α_{12}	-0.172	0.099	-0.156	0.122	-0.153	0.112	0.154	0.172	0.128	0.143	0.035	0.126
$\beta_{11} = 1$	-0.694	0.543	-0.717	0.565	-0.713	0.550	-0.319	0.325	-0.459	0.344	-0.557	0.400
α_{21}	-0.210	0.106	-0.161	0.103	-0.171	0.159	0.130	0.162	0.128	0.160	0.095	0.131
$\beta_{22} = 1$	-0.702	0.530	-0.722	0.566	-0.744	0.630	-0.327	0.292	-0.461	0.347	-0.536	0.391
$\alpha_{12} = \alpha_{21} = 0.85$												
α_{12}	-0.353	0.266	-0.368	0.274	-0.323	0.296	0.219	0.225	0.185	0.216	0.108	0.180
$\beta_{11} = 1$	-0.892	0.827	-0.882	0.828	-0.870	0.836	-0.761	0.682	-0.796	0.732	-0.843	0.779
α_{21}	-0.342	0.243	-0.310	0.234	-0.350	0.304	0.235	0.259	0.216	0.234	0.101	0.170
$\beta_{22} = 1$	-0.885	0.836	-0.884	0.833	-0.895	0.850	-0.759	0.704	-0.790	0.747	-0.848	0.775

5.3.3. Monte Carlo Simulation Results

Table 5.2 shows estimates of the bias and mean squared error (MSE) for the homoscedasticity error terms of normal generating x's with three different correlation coefficients, ρ=0, 0.25 and 0.75, and three different correlation between two equations, $\alpha_{12} = \alpha_{21}$ =0.25, 0.5, and 0.85. When the error is homoscedasticity error, the probit ML estimates well behave. Only with few exceptions the two-step probit ML estimates have smaller bias and MSE than semiparametric estimates. The absolute sizes of bias for the probit MLE and semiparametric estimates increase when the degree of correlation of equation α is increasing. Although the biases of Semiparametric estimates are larger than those of two-stage probit ML estimates, the MSE of semiparametric estimates are not much different form those of probit ML estimates. Unlike probit ML estimates, the bias of semiparametric estimates of coefficients for exogenous variables do not have monotonic increasing pattern when the degree of correlation ρ is larger. From Table 2, the semiparametric estimates have the larger bias and similar magnitude of MSE to the probit ML estimates. That is, the efficiency loss in normal and homoscedasticity design from using the semiparametric estimates is quite tolerable. These results confirm that the Klein and Spady's single equation results. The bias of probit ML estimates and semiparametric estimates for the exogenous variables show downward bias.

Table 5.3 represents the design of normal with heteroscedasticity error term. Since with heteroscedasticity the pribit model is inconsistent, we use the normalization for the two-stage probit ML estimates. It is similar to homoscedasticity design in the case of the increasing of the absolute sizes of bias for the probit and semiparametric

Table 5.3. Monte Carlo Parameter Estimates, Heteroscedasticity, Normal Design

Normal Design	Semiparametric						Probit MLE					
	ρ=0		ρ=0.25		ρ=0.75		ρ=0		ρ=0.25		ρ=0.75	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\alpha_{12} = \alpha_{21} = 0.25$												
α_{12}	0.039	0.742	0.038	0.448	0.008	0.313	0.138	0.173	0.140	0.139	0.103	0.121
$\beta_{11} = 1$	-0.696	0.593	-0.743	0.648	-0.654	0.564	-0.697	0.520	-0.723	0.540	-0.718	0.534
α_{21}	0.003	0.627	-0.150	0.626	-0.080	2.014	0.071	0.177	0.017	0.107	-0.105	0.426
$\beta_{22} = 1$	-0.639	0.540	-0.727	0.653	-0.652	0.539	-0.703	0.515	-0.718	0.532	-0.676	0.485
$\alpha_{12} = \alpha_{21} = 0.5$												
α_{12}	-0.145	0.364	-0.144	0.672	-0.059	0.552	0.064	0.120	0.117	0.105	0.035	0.144
$\beta_{11} = 1$	-0.788	0.876	-0.786	0.753	-0.815	0.738	-0.778	0.635	-0.807	-0.818	-0.557	0.683
α_{21}	-0.091	0.543	-0.226	1.434	-0.130	0.743	0.093	0.176	0.045	-0.012	0.095	0.155
$\beta_{22} = 1$	-0.752	0.633	-0.769	0.726	-0.720	0.676	-0.780	0.626	-0.801	-0.763	-0.536	0.614
$\alpha_{12} = \alpha_{21} = 0.85$												
α_{12}	-0.492	4.316	-0.089	1.615	-0.328	0.492	0.039	1.119	0.029	0.127	0.028	0.109
$\beta_{11} = 1$	-0.816	1.304	-0.933	0.996	-0.895	0.873	-0.932	0.997	-0.937	0.889	-0.936	0.886
α_{21}	-0.229	0.545	-0.400	2.520	-0.325	2.326	0.078	0.213	-0.079	1.911	-0.043	0.139
$\beta_{22} = 1$	-0.919	0.940	-0.870	0.962	-0.841	0.996	-0.944	0.917	-0.921	0.901	-0.897	0.823

estimates increase when the degree of correlation of equation α is increasing. In this design the pattern of semiparametric and probit ML estimates are quite different from the homoscedasticity design. First, the biases of estimates of coefficient for the x's using semiparametric estimates are quite smaller than probit ML estimates. Second, the MSE of both estimates are not quite different for these coefficients. Third, MSE of semiparametric estimates are relatively larger than those of homoscedasticity design and than probit ML estimates for the endogenous explanatory variables.

Table 5.4 and Table 5.5 show the results of χ^2 distributed explanatory variables design with homoscedasticity and heteroscedasticity error terms, respectively. The results of probit ML estimates and semiparametric estimates are similar with normal design. For example, the coefficients estimates of x's are increasing when correlation between equations and in heteroscadesticity design the bias of semiparametric estimates are smaller than those of probit ML estimates.

Also, biases of semiparametric estimation for the coefficients estimates of x's of all coefficients are smaller in the χ^2 design than those in the normal design. The one of the interesting feature of normal and χ^2 designs with heteroscedasticity is that the increasing degree of correlation causes the increasing the MSE for both probit MLE and semiparametric estimation. The χ^2 generating designs have larger MSE than normal generating designs. All bias of the coefficients estimates of x's are downward for all of the designs.

From the above Monte Carlo simulation, both estimators are quite biased. In the homoscedasticity design the semiparametric ML estimator and probit maximum likelihood estimation are quite similar in the both bias and MSE respects. However, In

Table 5.4. Monte Carlo Parameter Estimates, Homoscedasticity, χ^2 Design

χ^2 Design	Semiparametric						Probit MLE					
	ρ=0		ρ=0.25		ρ=0.75		ρ=0		ρ=0.25		ρ=0.75	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\alpha_{12} = \alpha_{21} = 0.25$												
α_{12}	-0.086	0.048	-0.076	0.045	-0.072	0.210	0.201	0.236	0.177	0.194	0.091	0.130
$\beta_{11} = 1$	-0.595	0.453	-0.598	0.450	-0.673	0.516	0.173	0.552	0.097	0.440	-0.668	0.486
α_{21}	-0.056	0.067	-0.095	0.055	-0.474	2.143	0.170	0.224	0.135	0.189	-0.328	1.008
$\beta_{22} = 1$	-0.577	0.425	-0.595	0.481	-0.525	0.458	0.101	0.389	0.099	0.427	-0.564	0.435
$\alpha_{12} = \alpha_{21} = 0.5$												
α_{12}	-0.167	0.138	-0.143	0.138	-0.426	2.100	0.484	0.550	0.492	0.622	0.097	0.250
$\beta_{11} = 1$	-0.617	0.523	-0.606	0.544	-0.663	0.690	0.125	0.388	0.132	0.516	-0.743	0.614
α_{21}	-0.129	0.147	-0.141	0.168	-0.388	3.438	0.426	0.504	0.448	0.561	-0.174	0.428
$\beta_{22} = 1$	-0.587	0.511	-0.597	0.525	-0.517	0.897	0.124	0.410	0.136	0.490	-0.637	0.524
$\alpha_{12} = \alpha_{21} = 0.85$												
α_{12}	-0.275	0.277	-0.238	0.268	-0.607	2.109	0.902	1.169	0.947	1.288	0.144	2.462
$\beta_{11} = 1$	-0.802	0.800	-0.782	0.809	-0.760	1.026	-0.356	0.564	-0.273	0.497	-0.941	1.183
α_{21}	-0.252	0.299	-0.263	0.340	-0.661	2.926	0.889	1.193	0.905	1.240	0.144	6.707
$\beta_{22} = 1$	-0.836	0.818	-0.810	0.801	-0.624	1.000	-0.386	0.576	-0.364	0.504	-0.935	2.249

the heteroscedaticity, however, semiparametric ML estimator is much smaller MSE than probit maximum likelihood estimator.

From the above Monte Carlo simulation, both estimators are quite biased. In the homoscedasticity design the semiparametric ML estimator and probit maximum likelihood estimation are quite similar in the both bias and MSE respects. In the heteroscedaticity, however, semiparametric ML estimator is much smaller MSE than probit maximum likelihood estimator.

5.4. Conclusions

This chapter has described a semiparametric maximum likelihood type method for estimating the parameters of two-equation simultaneous model. We suggest a method to use the semiparametric estimation in the second stage. The resulting estimators of application for the congressional voting data are significantly different from parametric two-stage probit ML estimators. The non-unimodal graph in the Figure 5.1 gives to us the possibility of use of semiparametric estimator (Horowitz and Hardle, 1996). The predicted value for the first equation which has the bimodal density is much closer to the actual value than probit ML estimates and for the second equation the predicted value of probit ML estimates are better than semiparametric estimates but not much different.

For small sample properties we carried out a Monte Carlo experiment. When the error terms are both homoscedasticity and heteroscedasticity, the semiparametric estimators have smaller and steadier MSE than parametric two-stage estimation method. The biases are quite downward. Also, we could see the increasing MSE when the degree of correlation between equations is increasing.

Table 5.5. Monte Carlo Parameter Estimates, Heteroscedasticity, χ^2 Design

χ^2 Design	Semiparametric						Probit MLE					
	ρ=0		ρ=0.25		ρ=0.75		ρ=0		ρ=0.25		ρ=0.75	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
$\alpha_{12} = \alpha_{21} = 0.25$												
α_{12}	-0.118	0.832	0.018	2.509	-0.056	0.066	0.070	0.288	0.134	0.318	0.120	0.209
$\beta_{11} = 1$	-0.594	0.501	-0.628	0.743	-0.616	0.451	-0.662	0.502	-0.663	0.504	-0.158	0.353
α_{21}	-0.186	0.448	-0.217	0.903	-0.068	0.073	0.073	0.213	0.044	0.197	0.100	0.182
$\beta_{22} = 1$	-0.622	0.523	-0.585	0.519	-0.614	0.453	-0.677	0.527	-0.658	0.512	-0.192	0.312
$\alpha_{12} = \alpha_{21} = 0.5$												
α_{12}	-0.046	2.077	-0.252	1.386	-0.151	0.137	0.189	1.797	0.141	0.294	0.278	0.361
$\beta_{11} = 1$	-0.685	0.660	-0.660	0.698	-0.660	0.526	-0.755	0.925	-0.722	0.596	-0.249	0.373
α_{21}	-0.277	1.956	-0.284	0.760	-0.132	0.231	0.060	0.408	0.089	0.285	0.357	9.850
$\beta_{22} = 1$	-0.628	0.751	-0.623	0.591	-0.620	0.508	-0.717	0.626	-0.701	0.603	-0.301	1.996
$\alpha_{12} = \alpha_{21} = 0.85$												
α_{12}	-0.485	1.437	-0.427	2.477	-0.297	0.875	0.165	0.915	0.017	5.569	0.428	0.631
$\beta_{11} = 1$	-0.764	0.873	-0.750	1.115	-0.815	0.917	-0.890	0.979	-0.834	2.267	-0.576	0.641
α_{21}	-0.504	2.414	-0.535	1.314	-0.216	1.957	0.084	2.337	0.101	0.452	0.325	1.879
$\beta_{22} = 1$	-0.778	0.999	-0.748	0.836	-0.869	1.173	-0.895	1.218	-0.926	0.995	-0.585	0.843

In this chapter we use the nonparametric kernel density estimator with the constant bandwidth rule which is fixed for all sample size. Variable bandwidth can be used to get nonparametric density estimator also, and it will give us more accurate empirical density function.

REFERENCES

Amemiya, T. (1974), Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal, *Econometrica*, 42, 999-1012.

Andrews, D.W.K (1995), Nonparametric Kernel Estimation for Semiparametric Models, *Econometric theory*, 11, 560-596.

Basu, R. and A. Ullah (1992), Chinese Earning-Age Profile: A Nonparametric analysis, *Journal of International Trade and Economic Development*, 1, 151-165.

Biwman, A. (1984), An Alternative Method of Cross-Validation for the Smoothing of Density Estimates, *Biometrika*, 71, 353-360.

Bound, J. and G. Johnson (1992), Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations, *The American Economic Review*, Vol. 82, 371-392.

Bratsberg, B. and D. Terrell (1997), Experience, Tenure, and Wage Growth of Young Black and White Men, *Journal of Human Resources, forthcoming.*

Breiman, L., Meise, W. and E. Purcell (1977), Variable kernel estimates of multivariate densities. *Technometrics*, 19, 135-144.

Bresnahan, T.F, and P. C. Reiss (1991), Empirical models of Discrete Games, *Journal of Econometrics*, 48, 57-81.

Buchinsky, M. (1994), Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression, *Econometrica*, 62, 405-458.

Carrasco, R. (2001), Binary Choice with Binary Endogenous Regressors in Panel data: Estimating the Effect of Fertility on Female Labor Participation, *Journal of Business & Economic Statistics*, October, Vol. 19, 385-394.

Dart, J. D. (1997), *Nonparametric Smoothing and Lack-of –Fit Tests*, Springer-Verlag, New York.

DiNardo, J., N. M. Fortin, and T. Lemieus (1996), Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach, *Econometrica*, 64, 1001-1044.

Fan, J. (1992), Design-adaptive Nonparametric regression, *Journal of the American Statistical Association*, 87, 998-1004.

Fan, J. and I. Gijbels (1992), Variable bandwidth and local linear regression smoothers, *The Annals of Statistics*, 20, 2008-2036.

_______________(1995), Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation, *Journal of the American Statistical Association*, 86, 643-652.

_______________(1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall, London.

Fan, Y. (1995), Bootstrapping a consistent nonparametric goodness-of-fit test, *Econometric Reviews*, 14(3), 367-382.

Ginther, D. K. (2000), Nonparametric Analysis of the U.S. Male Earnings Distribution, *Applying Kernel and Nonparametric Estimation to Economic Topics*, edited by Fomby, T. B. and R. C. Hill, 275-308.

Griffiths, W.E, R. C. Hill, and P. J. Pope (1987), Small Sample Properties of Probit Model Estimators, *Journal of American Statistical Association*, 82, 929-937.

Hardle, W. (1991), *Applied Nonparametric Regression*, Cambridge University Press, New York.

Horowitz, J. L. (1997), Bootstrap methods in econometrics: Theory and numerical Performance, in *Advances in Economics and Econometrics : Theory and Applications, Vol. III.*

Horowitz, J.L. and N.E. Savin (2000), Empirically relevant critical values for hypothesis tests: A Bootstrap approach, *Journal of Econometrics*, 95, 375-389.

Horowitz, J.L. and W. Hardle (1996), Direct semiparametric estimation of Single-index models with discrete covariate, *Journal of the American Statistical Association*, 91, 1632-1640.

Iwata, S. H. Murao, and Q. Wang (1999), Nonparametric assessment of the effects of neighborhood land uses on residential house values, in *Applying Kernel and Nonparametric Estimation to Economic Topics*, edited by Fomby, T. B. and R. C. Hill, 201-228.

Katz, L. F. (1999), Technological Change, Computerization, and the Wage Structure, *Conference on "Understanding the Digital Economy: Data, Tools, and Research"*, Washington, DC, May 25-26, 1999.

Klein, R.W. and R.H. Spady (1993), An efficient semiparametric estimator for binary response models, *Econometrica*, 61, 387-421.

Lee, M. J. (1995), Semiparametric Estimation of Simulataneous Equations with Limited Dependent Variables: A Case Study of Female Labor Supply, *Journal of Applied Econometrics*, 10, 487-200.

Lee, M. J. (1996), *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*, Springer-Verlag, New York.

Lee, Thomas C.M. and Victor Solo (1999), Bandwidth selection for local linear regression: A simulation study, *Computational Statistics*, 14, 515-532.

Maddala, G.S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

Mallar, C. D. (1977), The Estimation of Simultaneous Probability Models, *Econometrica* 45(7), 1717-22.

Manski, C. (1975), Maximum Score Estimation of the Stochastic Utility Model of Choice, *Journal of Econometrics*, 3, 205-228.

McMillen, D.P. and P. Thorsnes (1999), The reaction of housing prices to information on superfund sites: A semiparametric analysis of the Tacoma, Washington Market, in *Applying Kernel and Nonparametric Estimation to Economic Topics*, edited by Fomby, T. B. and R. C. Hill, 229-258.

Mincer, J. (1974), *Schooling, Experience, and Earnings*, New York: National Bureau of Economic Research.

Murphy, K.M. and F. Welch (1990), Empirical Age-Earnings Profiles, *Journal of Labor Economics*, 8, 202-229.

Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Park, B.U. and B.A. Turlach (1992), Practical Performance of Several Data Driven Bandwidth Selectors, *Computational Statistics*, 7, 251-270.

Pudney, S. (1993), Income and wealth inequality and the life cycle: A nonparametric analsysi for China, *Journal of Applied Econometrics*, 8, 249-276.

Rudemo, M. (1982), Empirical choice of histograms and kernel density estimators, *Scandinavian Journal of Statistics*, 9, 65-78.

Ruppert, D., S.J. Sheather, and M. P. Wand (1995), An effective bandwidth selector for local least squares regression, *Journal of the American Statistical Association*, 90, 1257-1270.

Seather, S. J. (1992), The Performance of Six Popular Bandwidth Selection Methods on Some Real Data Sets, *Computational Statistics*, 7, 225-250.

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

Stock, James H. (1989), Nonparametric Policy Analysis, *Journal of the American Statistical Association*, 84, 567-575.

Stratmann, T. (1992), The effects of logrolling on congressional voting, *American Economic Review*, 82(5), 1162-1176.

Wand, M. P. and M. C. Jones (1995), *Kernel Smoothing*, Chapman and Hall, London.

Zhang, W. and S. Y. Lee (2000), Variable bandwidth selection in varying-coefficient Models, *Journal of Multivariate Analysis*, 74, 116-134.

Zheng, J. X. (2000), Specification Testing and Nonparametric Estimation of the Human Capital Model, *Applying Kernel and Nonparametric Estimation to Economic Topics*, edited by Fomby, T. B. and R. C. Hill, 129-154.

APPENDIX

Table A.1. Percent of Mean Estimates that Lie Outside of Semiparametric 95 Percent Confidence Intervals, 1997

1979	Quadratic Model	Quartic Model
White, Male	75.5	18.4
White, Female	75.0	22.9
Nonwhite, Male	51.9	05.8
Nonwhite, Female	64.3	02.0
1988		
White, Male	74.5	31.9
White, Female	77.1	22.9
Nonwhite, Male	28.0	14.0
Nonwhite, Female	23.4	10.6
1997		
White, Male	43.5	17.4
White, Female	50.0	15.2
Nonwhite, Male	26.1	13.4
Nonwhite, Female	15.2	10.9

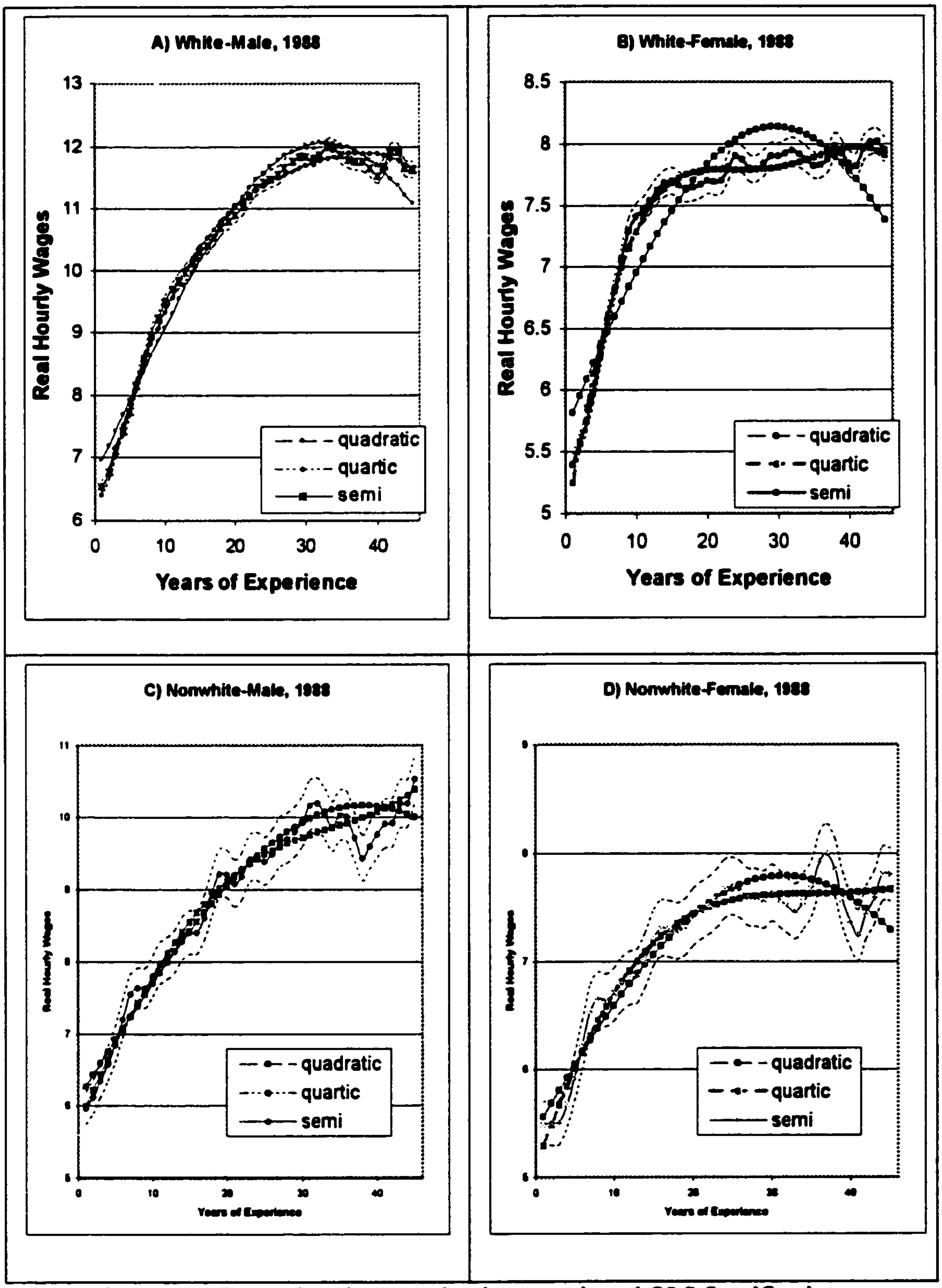


Figure A.1. Comparison between Semiparametric and OLS Specifications

Figure A.1.1. Comparison between Semiparametric and OLS Specifications for 1988

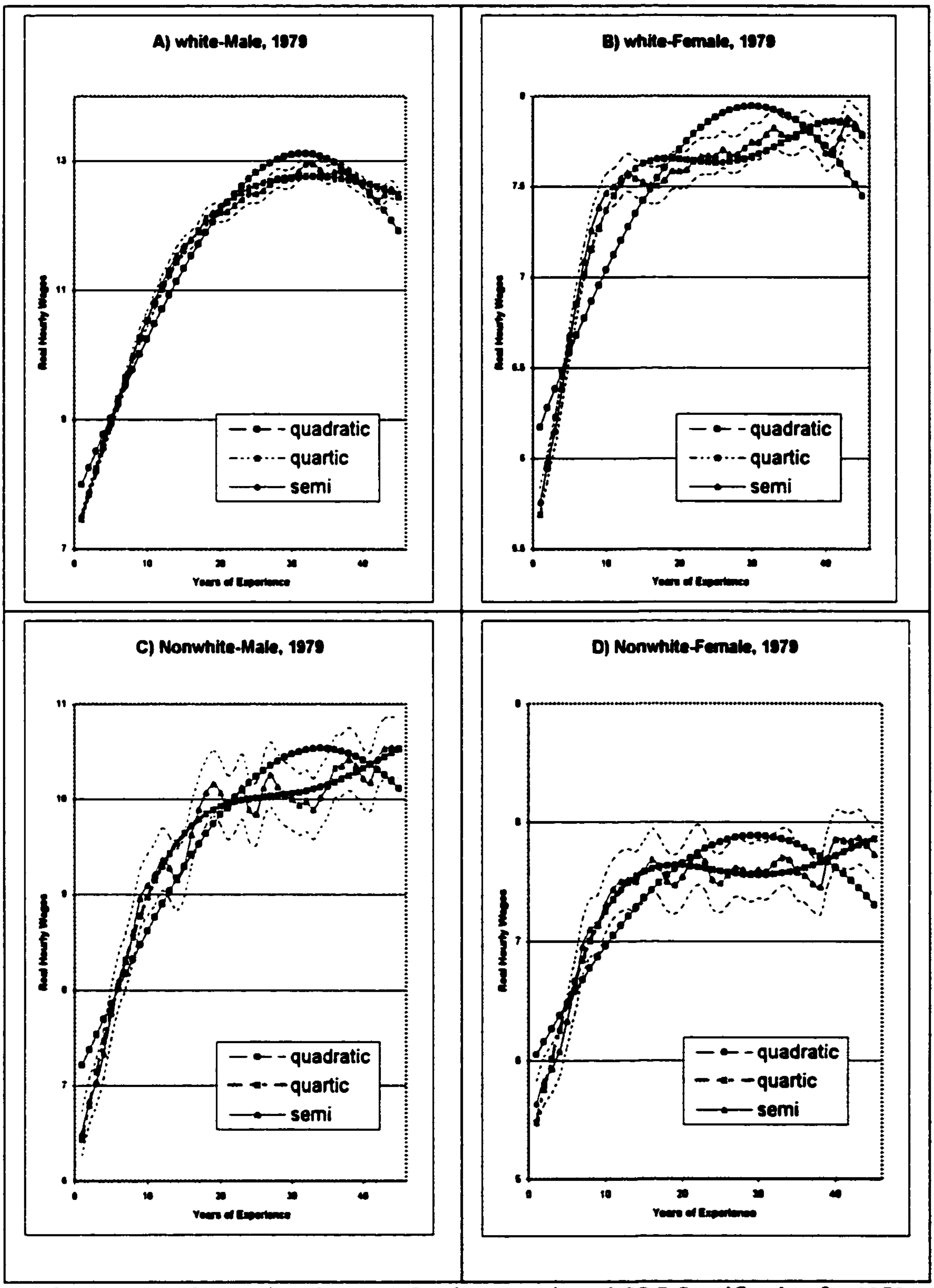


Figure A.1.2. Comparison between Semiparametric and OLS Specification for 1979

Table A.2. Coefficient Estimates of Each Subsamples

A.2.1. Coefficient Estimates 1997, Female-White

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.268(0.018)	0.219(0.045)	
Region2	-0.068(0.006)	-0.068(0.015)	-0.069(0.018)
Region3	-0.085(0.006)	-0.085(0.013)	-0.086(0.015)
Region4	-0.030(0.006)	-0.029(0.014)	-0.030(0.017)
FTPT	0.137(0.005)	0.135(0.015)	0.129(0.014)
SMSA	0.164(0.005)	0.163(0.013)	0.151(0.014)
Marry	0.033(0.004)	0.028(0.009)	0.034(0.010)
Dind1	-0.075(0.019)	-0.076(0.065)	-0.076(0.029)
Dind2	-0.023(0.011)	-0.023(0.026)	-0.023(0.027)
Dind3	-0.071(0.011)	-0.072(0.025)	-0.072(0.028)
Dind4	-0.010(0.014)	-0.010(0.030)	-0.009(0.029)
Dind5	0.122(0.026)	0.123(0.064)	0.122(0.040)
Dind6	-0.086(0.014)	-0.087(0.041)	-0.087(0.035)
Dind7	-0.286(0.009)	-0.283(0.022)	-0.285(0.027)
Dind8	-0.051(0.012)	-0.051(0.028)	-0.051(0.040)
Dind9	-0.113(0.012)	-0.114(0.027)	-0.113(0.031)
Dind10	-0.282(0.014)	-0.282(0.029)	-0.283(0.038)
Dind11	-0.200(0.018)	-0.195(0.043)	-0.196(0.052)
Dind12	-0.093(0.010)	-0.093(0.023)	-0.095(0.042)
Dind13	0.020(0.012)	0.020(0.025)	0.019(0.041)
Dind14	-0.334(0.015)	-0.333(0.030)	-0.335(0.053)
Dind15	-0.163(0.010)	-0.159(0.023)	-0.162(0.034)
Dind16	0.011(0.012)	0.011(0.034)	0.011(0.045)
Dind17	0.008(0.012)	0.009(0.024)	0.008(0.031)
Schooling	0.105(0.001)	0.105(0.002)	0.106(0.003)
Experience	0.026(0.001)	0.042(0.007)	
Experience^2	-4.45E-05(1.5E-05)	-0.002(2.83E-04)	
Experience^3		2.43E-06(9.93E-07)	
Experience^4		-1.29E-08(1.15E-08)	
R-square	0.349	0.350	0.350
Hausman test: OLS vs. Semiparametric	322.121	26.719	

A.1.2. Coefficient Estimates 1997, Male-Nonwhite

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.392(0.047)	0.312(0.051)	
Region2	-0.024(0.018)	-0.023(0.018)	-0.022(0.018)
Region3	-0.092(0.015)	-0.092(0.015)	-0.093(0.015)
Region4	0.062(0.016)	0.062(0.016)	0.061(0.017)
FTPT	0.212(0.024)	0.199(0.024)	0.179(0.014)
SMSA	0.100(0.016)	0.100(0.016)	0.086(0.014)
Marry	0.098(0.011)	0.093(0.011)	0.096(0.010)
Dind1	0.053(0.029)	0.053(0.029)	0.053(0.029)
Dind2	0.015(0.025)	0.017(0.025)	0.016(0.027)
Dind3	-0.043(0.027)	-0.043(0.027)	-0.045(0.028)
Dind4	0.025(0.027)	0.026(0.027)	0.025(0.029)
Dind5	0.117(0.042)	0.121(0.042)	0.118(0.040)
Dind6	-0.037(0.034)	-0.040(0.034)	-0.038(0.035)
Dind7	-0.227(0.025)	-0.224(0.025)	-0.224(0.027)
Dind8	0.112(0.040)	0.111(0.040)	0.113(0.040)
Dind9	-0.155(0.029)	-0.155(0.029)	-0.156(0.031)
Dind10	-0.226(0.036)	-0.228(0.036)	-0.228(0.038)
Dind11	-0.183(0.046)	-0.182(0.046)	-0.183(0.052)
Dind12	-0.111(0.042)	-0.110(0.042)	-0.115(0.042)
Dind13	-0.093(0.038)	-0.090(0.037)	-0.088(0.041)
Dind14	-0.195(0.055)	-0.196(0.055)	-0.189(0.053)
Dind15	-0.087(0.031)	-0.085(0.031)	-0.089(0.034)
Dind16	0.070(0.039)	0.070(0.039)	0.070(0.045)
Dind17	0.067(0.028)	0.068(0.028)	0.065(0.031)
Schooling	0.091(0.002)	0.091(0.002)	0.092(0.003)
Experience	0.023(0.002)	0.049(0.008)	
Experience^2	-3.15E-04(4.0E-05)	-0.002(0.001)	
Experience^3		5.05E-05(2.52E-05)	
Experience^4		-4.12E-07(2.83E-07)	
R-square	0.308	0.310	0.313
Hausman test: OLS vs. Semiparametric	190.590	122.606	

A.1.3. Coefficient Estimates 1997, Female-Nonwhite

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.369(0.042)	0.329(0.045)	
Region2	-0.062(0.015)	-0.062(0.015)	-0.063(0.015)
Region3	-0.136(0.013)	-0.136(0.013)	-0.138(0.013)
Region4	0.003(0.014)	0.004(0.014)	0.001(0.015)
FTPT	0.162(0.015)	0.158(0.015)	0.150(0.012)
SMSA	0.112(0.013)	0.112(0.013)	0.098(0.011)
Marry	0.043(0.009)	0.041(0.009)	0.044(0.008)
Dind1	0.087(0.065)	0.086(0.065)	0.084(0.062)
Dind2	-0.047(0.026)	-0.047(0.026)	-0.048(0.027)
Dind3	-0.114(0.025)	-0.115(0.025)	-0.117(0.024)
Dind4	0.034(0.030)	0.033(0.030)	0.033(0.031)
Dind5	0.096(0.064)	0.096(0.064)	0.101(0.057)
Dind6	-0.076(0.041)	-0.076(0.041)	-0.078(0.037)
Dind7	-0.271(0.022)	-0.271(0.022)	-0.273(0.021)
Dind8	-0.048(0.028)	-0.048(0.028)	-0.048(0.028)
Dind9	-0.135(0.027)	-0.137(0.027)	-0.137(0.026)
Dind10	-0.247(0.029)	-0.248(0.029)	-0.251(0.027)
Dind11	-0.167(0.043)	-0.166(0.043)	-0.168(0.048)
Dind12	-0.187(0.023)	-0.188(0.023)	-0.189(0.023)
Dind13	-0.037(0.025)	-0.037(0.025)	-0.037(0.024)
Dind14	-0.299(0.030)	-0.300(0.030)	-0.301(0.033)
Dind15	-0.101(0.023)	-0.100(0.023)	-0.103(0.023)
Dind16	-0.032(0.034)	-0.031(0.034)	-0.032(0.034)
Dind17	0.049(0.024)	0.049(0.024)	0.048(0.023)
Schooling	0.097(0.002)	0.097(0.002)	0.098(0.003)
Experience	0.021(0.001)	0.036(0.007)	
Experience^2	-3.25E-05(3.5E-06)	-0.002(0.001)	
Experience^3		4.17E-05(2.28E-05)	
Experience^4		-4.18E-07(2.62E-07)	
R-square	0.331	0.331	0.334
Hausman test: OLS vs. Semiparametric	249.264	196.258	

A.2.4. Coefficient Estimates 1988, Male-White

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.834(0.012)	0.750(0.013)	
Region2	-0.073(.004)	-0.072(0.004)	-0.070(0.004)
Region3	-0.108(.004)	-0.107(0.004)	-0.106(0.004)
Region4	-0.007(.005)	-0.007(0.005)	-0.007(0.005)
FTPT	-0.296(.006)	-0.281(0.006)	-0.257(0.007)
SMSA	0.156(.003)	0.155(0.003)	0.163(0.002)
Marry	0.126(.004)	0.117(0.004)	0.109(0.002)
Dind1	0.086(0.008)	0.084(0.008)	0.084(0.008)
Dind2	0.066(0.007)	0.066(0.007)	0.067(0.008)
Dind3	0.023(0.008)	0.022(0.008)	0.023(0.008)
Dind4	0.034(0.008)	0.034(0.008)	0.033(0.009)
Dind5	0.160(0.011)	0.159(0.011)	0.161(0.011)
Dind6	-0.048(0.009)	-0.049(0.009)	-0.049(0.009)
Dind7	-0.215(0.007)	-0.212(0.007)	-0.214(0.008)
Dind8	0.090(0.012)	0.091(0.012)	0.091(0.012)
Dind9	-0.054(0.010)	-0.055(0.010)	-0.057(0.012)
Dind10	-0.301(0.014)	-0.302(0.014)	-0.305(0.016)
Dind11	-0.186(0.016)	-0.186(0.016)	-0.191(0.019)
Dind12	-0.103(0.016)	-0.107(0.016)	-0.108(0.018)
Dind13	-0.097(0.013)	-0.100(0.013)	-0.101(0.013)
Dind14	-0.274(0.019)	-0.278(0.019)	-0.281(0.024)
Dind15	-0.141(0.009)	-0.140(0.009)	-0.141(0.010)
Dind16	-0.073(0.010)	-0.074(0.010)	-0.075(0.013)
Dind17	0.025(0.009)	0.025(0.008)	0.026(0.009)
Schooling	0.075(0.001)	0.075(0.001)	0.075(0.001)
Experience	0.035(0.001)	0.061(0.002)	
Experience^2	-0.001(1.12E-05)	-0.003(0.000192)	
Experience^3		5.39E-05(6.66E-06)	
Experience^4		-4.55E-07(7.57E-08)	
R-square	0.414	0.416	0.416
Hausman test: OLS vs. Semiparametric	746.781	34.813	

A.2.5. Coefficient Estimates 1988, Female-White

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.792(0.014)	0.698(0.014)	
Region2	-0.100(0.004)	-0.100(0.004)	-0.100(0.004)
Region3	-0.106(0.004)	-0.106(0.004)	-0.106(0.004)
Region4	-0.014(0.005)	-0.015(0.005)	-0.015(0.005)
FTPT	-0.210(0.004)	-0.204(0.004)	-0.202(0.004)
SMSA	0.161(0.004)	0.161(0.004)	0.160(0.002)
Marry	0.034(0.003)	0.022(0.004)	0.022(0.002)
Dind1	-0.049(0.015)	-0.052(0.015)	-0.053(0.015)
Dind2	-0.009(0.008)	-0.011(0.008)	-0.011(0.008)
Dind3	-0.097(0.008)	-0.099(0.008)	-0.099(0.009)
Dind4	0.029(0.012)	0.028(0.012)	0.027(0.013)
Dind5	0.115(0.020)	0.113(0.020)	0.112(0.019)
Dind6	-0.082(0.011)	-0.083(0.011)	-0.084(0.011)
Dind7	-0.365(0.007)	-0.361(0.007)	-0.362(0.008)
Dind8	-0.046(0.009)	-0.045(0.009)	-0.044(0.009)
Dind9	-0.103(0.010)	-0.102(0.010)	-0.102(0.010)
Dind10	-0.321(0.011)	-0.321(0.010)	-0.321(0.011)
Dind11	-0.238(0.017)	-0.237(0.017)	-0.239(0.019)
Dind12	-0.119(0.008)	-0.120(0.008)	-0.121(0.009)
Dind13	0.077(0.008)	0.074(0.008)	0.073(0.008)
Dind14	-0.277(0.011)	-0.276(0.011)	-0.276(0.012)
Dind15	-0.130(0.008)	-0.125(0.008)	-0.126(0.008)
Dind16	-0.064(0.010)	-0.063(0.010)	-0.064(0.010)
Dind17	0.018(0.009)	0.016(0.009)	0.016(0.009)
Schooling	0.080(0.001)	0.079(0.001)	0.080(0.001)
Experience	0.024(0.000)	0.062(0.002)	
Experience^2	-0.00041(1.14E-05)	-0.0035(0.000198)	
Experience^3		8.56E-05(7.00E-06)	
Experience^4		-7.51E-07(8.10E-08)	
R-square	0.384	0.390	0.391
Hausman test: OLS vs. Semiparametric	487.406	70.727	

A.2.6. Coefficient Estimates 1988, Male-Nonwhite

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.903(0.036)	0.853(0.039)	
Region2	-0.082(0.015)	-0.081(0.015)	-0.083(0.004)
Region3	-0.168(0.013)	-0.169(0.013)	-0.167(0.004)
Region4	0.012(0.014)	0.012(0.014)	0.014(0.005)
FTPT	-0.271(0.015)	-0.265(0.015)	-0.247(0.004)
SMSA	0.151(0.012)	0.150(0.012)	0.160(0.002)
Marry	0.080(0.010)	0.075(0.010)	0.069(0.002)
Dind1	0.094(0.025)	0.093(0.025)	0.092(0.015)
Dind2	0.031(0.022)	0.031(0.022)	0.032(0.008)
Dind3	-0.007(0.024)	-0.007(0.024)	-0.004(0.009)
Dind4	0.089(0.024)	0.089(0.024)	0.088(0.013)
Dind5	0.145(0.032)	0.144(0.032)	0.146(0.019)
Dind6	-0.044(0.030)	-0.047(0.030)	-0.042(0.011)
Dind7	-0.250(0.023)	-0.247(0.023)	-0.248(0.008)
Dind8	0.058(0.039)	0.055(0.039)	0.052(0.009)
Dind9	-0.141(0.027)	-0.142(0.027)	-0.145(0.010)
Dind10	-0.251(0.032)	-0.251(0.032)	-0.251(0.011)
Dind11	-0.049(0.043)	-0.047(0.043)	-0.046(0.019)
Dind12	-0.147(0.040)	-0.146(0.040)	-0.149(0.009)
Dind13	-0.080(0.030)	-0.080(0.030)	-0.079(0.008)
Dind14	-0.217(0.043)	-0.215(0.043)	-0.218(0.012)
Dind15	-0.063(0.026)	-0.064(0.026)	-0.066(0.008)
Dind16	-0.041(0.034)	-0.040(0.034)	-0.042(0.010)
Dind17	0.081(0.025)	0.080(0.025)	0.079(0.009)
Schooling	0.070(0.002)	0.071(0.002)	0.071(0.001)
Experience	0.026(0.001)	0.037(0.006)	
Experience^2	-0.00035(3.10E-05)	-0.00089(0.000519)	
Experience^3		5.34E-06(1.73E-05)	
Experience^4		5.44E-08(1.90E-07)	
R-square	0.374	0.375	0.377
Hausman test: OLS vs. Semiparametric	264.937	232.278	

A.2.7. Coefficient Estimates 1988, Female-Nonwhite

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	0.904(0.034)	0.852(0.037)	
Region2	-0.078(0.013)	-0.078(0.013)	-0.077(0.013)
Region3	-0.141(0.011)	-0.142(0.011)	-0.141(0.012)
Region4	0.017(0.013)	0.017(0.013)	0.017(0.014)
FTPT	-0.205(0.010)	-0.202(0.010)	-0.199(0.009)
SMSA	0.147(0.011)	0.149(0.011)	0.149(0.005)
Marry	-0.002(0.008)	-0.006(0.008)	-0.007(0.007)
Dind1	0.005(0.051)	0.000(0.051)	-0.002(0.064)
Dind2	-0.079(0.022)	-0.078(0.022)	-0.078(0.022)
Dind3	-0.223(0.021)	-0.222(0.021)	-0.221(0.021)
Dind4	0.029(0.026)	0.027(0.026)	0.027(0.027)
Dind5	0.083(0.049)	0.084(0.049)	0.084(0.042)
Dind6	-0.148(0.035)	-0.146(0.035)	-0.146(0.034)
Dind7	-0.373(0.020)	-0.368(0.020)	-0.369(0.020)
Dind8	-0.087(0.025)	-0.087(0.025)	-0.086(0.024)
Dind9	-0.216(0.023)	-0.215(0.023)	-0.215(0.024)
Dind10	-0.359(0.024)	-0.361(0.024)	-0.360(0.023)
Dind11	-0.290(0.049)	-0.290(0.049)	-0.292(0.053)
Dind12	-0.283(0.022)	-0.283(0.021)	-0.283(0.021)
Dind13	-0.084(0.020)	-0.084(0.020)	-0.084(0.020)
Dind14	-0.328(0.025)	-0.329(0.025)	-0.329(0.026)
Dind15	-0.080(0.020)	-0.078(0.020)	-0.078(0.020)
Dind16	-0.063(0.029)	-0.063(0.028)	-0.063(0.028)
Dind17	-0.011(0.021)	-0.010(0.021)	-0.010(0.021)
Schooling	0.074(0.002)	0.074(0.002)	0.074(0.002)
Experience	0.022(0.001)	0.037(0.005)	
Experience^2	-3.63E-04(2.9E-06)	-0.001(4.92E-04)	
Experience^3		2.15E-06(1.71E-06)	
Experience^4		-1.21E-08(1.96E-08)	
R-square	0.389	0.390	0.392
Hausman test: OLS vs. Semiparametric	42.975	14.363	

A.2.8. Coefficient Estimates 1979, Male-White

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	1.097(0.011)	1.029(0.012)	
Region2	0.041(0.004)	0.042(0.004)	0.043(0.004)
Region3	-0.021(0.004)	-0.020(0.004)	-0.018(0.004)
Region4	0.108(0.004)	0.108(0.004)	0.108(0.004)
FTPT	-0.242(0.006)	-0.230(0.006)	-0.202(0.008)
SMSA	0.090(0.003)	0.091(0.003)	0.097(0.002)
Marry	0.130(0.004)	0.115(0.004)	0.095(0.002)
Dind1	0.120(0.007)	0.121(0.007)	0.123(0.008)
Dind2	0.058(0.007)	0.059(0.007)	0.061(0.007)
Dind3	0.010(0.007)	0.010(0.007)	0.012(0.007)
Dind4	0.100(0.008)	0.099(0.008)	0.099(0.009)
Dind5	0.084(0.011)	0.084(0.011)	0.087(0.010)
Dind6	-0.039(0.008)	-0.039(0.008)	-0.038(0.009)
Dind7	-0.186(0.007)	-0.183(0.007)	-0.184(0.008)
Dind8	0.077(0.012)	0.074(0.012)	0.075(0.014)
Dind9	-0.120(0.012)	-0.121(0.012)	-0.124(0.015)
Dind10	-0.280(0.015)	-0.281(0.015)	-0.283(0.017)
Dind11	-0.149(0.016)	-0.147(0.016)	-0.153(0.022)
Dind12	-0.109(0.017)	-0.110(0.017)	-0.111(0.019)
Dind13	-0.150(0.012)	-0.151(0.012)	-0.151(0.013)
Dind14	-0.478(0.014)	-0.479(0.014)	-0.479(0.019)
Dind15	-0.191(0.008)	-0.193(0.008)	-0.193(0.009)
Dind16	0.031(0.011)	0.030(0.011)	0.030(0.012)
Dind17	-0.003(0.008)	-0.003(0.008)	-0.002(0.008)
Schooling	0.064(0.001)	0.065(0.001)	0.065(0.001)
Experience	0.032(0.000)	0.055(0.002)	
Experience^2	-5.30E-05(9.9E-07)	-0.002(1.64E-05)	
Experience^3		3.74E-06(5.50E-07)	
Experience^4		-2.63E-08(6.10E-09)	
R-square	0.352	0.355	0.355
Hausman test: OLS vs. Semiparametric	244.538	21.327	

A.2.9. Coefficient Estimates 1979, Female-White

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	1.059(0.013)	0.984(0.013)	
Region2	0.007(0.004)	0.007(0.004)	0.007(0.004)
Region3	-0.032(0.004)	-0.032(0.004)	-0.032(0.004)
Region4	0.083(0.005)	0.082(0.004)	0.082(0.004)
FTPT	-0.160(0.004)	-0.156(0.004)	-0.154(0.004)
SMSA	0.097(0.003)	0.096(0.003)	0.096(0.002)
Marry	0.020(0.003)	0.009(0.003)	0.008(0.002)
Dind1	-0.016(0.015)	-0.016(0.015)	-0.016(0.017)
Dind2	-0.001(0.008)	-0.002(0.008)	-0.001(0.008)
Dind3	-0.063(0.008)	-0.064(0.008)	-0.064(0.008)
Dind4	0.038(0.012)	0.036(0.012)	0.036(0.014)
Dind5	0.060(0.020)	0.058(0.020)	0.058(0.019)
Dind6	-0.080(0.011)	-0.081(0.011)	-0.081(0.011)
Dind7	-0.269(0.007)	-0.264(0.007)	-0.264(0.008)
Dind8	-0.078(0.009)	-0.077(0.009)	-0.076(0.009)
Dind9	-0.131(0.012)	-0.132(0.012)	-0.131(0.013)
Dind10	-0.242(0.011)	-0.246(0.011)	-0.245(0.012)
Dind11	-0.141(0.017)	-0.134(0.016)	-0.135(0.022)
Dind12	-0.107(0.008)	-0.108(0.008)	-0.107(0.009)
Dind13	0.009(0.008)	0.007(0.008)	0.007(0.008)
Dind14	-0.172(0.012)	-0.174(0.012)	-0.174(0.014)
Dind15	-0.113(0.007)	-0.113(0.007)	-0.113(0.008)
Dind16	-0.047(0.011)	-0.049(0.011)	-0.049(0.011)
Dind17	0.045(0.009)	0.045(0.009)	0.045(0.010)
Schooling	0.063(0.001)	0.063(0.001)	0.063(0.001)
Experience	0.017(0.000)	0.050(0.002)	
Experience^2	-2.92E-04(1.0E-06)	-0.003(1.76E-05)	
Experience^3		7.85E-06(6.2E-07)	
Experience^4		-7.10E-08(7.0E-09)	
R-square	0.276	0.282	0.282
Hausman test: OLS vs. Semiparametric	434.648	134.113	

A.2.10. Coefficient Estimates 1979, Male-Nonwhite

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	1.205(0.034)	1.090(0.036)	
Region2	0.060(0.015)	0.059(0.015)	0.059(0.015)
Region3	-0.100(0.013)	-0.101(0.013)	-0.099(0.013)
Region4	0.097(0.014)	0.097(0.014)	0.101(0.014)
FTPT	-0.204(0.016)	-0.189(0.016)	-0.171(0.016)
SMSA	0.074(0.010)	0.076(0.010)	0.084(0.010)
Marry	0.103(0.010)	0.085(0.010)	0.069(0.009)
Dind1	0.138(0.024)	0.137(0.024)	0.138(0.024)
Dind2	0.083(0.022)	0.082(0.022)	0.085(0.022)
Dind3	0.053(0.023)	0.053(0.023)	0.057(0.023)
Dind4	0.130(0.025)	0.127(0.025)	0.126(0.025)
Dind5	0.091(0.032)	0.096(0.032)	0.100(0.032)
Dind6	0.026(0.029)	0.023(0.028)	0.023(0.028)
Dind7	-0.177(0.023)	-0.173(0.023)	-0.174(0.023)
Dind8	0.023(0.042)	0.026(0.042)	0.027(0.042)
Dind9	-0.180(0.032)	-0.186(0.032)	-0.189(0.032)
Dind10	-0.190(0.035)	-0.190(0.035)	-0.192(0.035)
Dind11	-0.145(0.049)	-0.148(0.048)	-0.156(0.048)
Dind12	-0.060(0.046)	-0.057(0.046)	-0.053(0.046)
Dind13	-0.051(0.030)	-0.049(0.030)	-0.050(0.030)
Dind14	-0.074(0.039)	-0.076(0.039)	-0.076(0.038)
Dind15	0.006(0.026)	0.005(0.026)	0.005(0.026)
Dind16	0.106(0.042)	0.104(0.041)	0.107(0.041)
Dind17	0.097(0.024)	0.098(0.024)	0.099(0.024)
Schooling	0.054(0.002)	0.055(0.002)	0.055(0.001)
Experience	0.023(0.001)	0.058(0.005)	
Experience^2	-3.50E-05(2.8E-05)	-0.003(4.54E-05)	
Experience^3		5.9E-05 (1.5E-05)	
Experience^4		-4.4E-07 (1.5E-07)	
R-square	0.308	0.314	0.315
Hausman test: OLS vs. Semiparametric	321.595	215.690	

A.2.11. Coefficient Estimates 1979, Female-Nonwhite

Variable	OLS (quadratic)	OLS (quartic)	Semiparametric
Constant	1.116(0.031)	1.005(0.033)	
Region2	-0.012(0.013)	-0.010(0.013)	0.000(0.028)
Region3	-0.113(0.011)	-0.113(0.011)	-0.011(0.013)
Region4	0.049(0.013)	0.050(0.013)	-0.114(0.011)
FTPT	-0.141(0.011)	-0.133(0.011)	0.049(0.012)
SMSA	0.060(0.009)	0.060(0.009)	-0.133(0.011)
Marry	0.028(0.008)	0.019(0.008)	0.060(0.009)
Dind1	0.070(0.057)	0.077(0.056)	0.019(0.008)
Dind2	-0.003(0.020)	0.003(0.020)	0.079(0.056)
Dind3	-0.113(0.019)	-0.108(0.019)	0.006(0.020)
Dind4	0.090(0.033)	0.094(0.033)	-0.106(0.019)
Dind5	0.065(0.051)	0.066(0.051)	0.097(0.033)
Dind6	-0.034(0.035)	-0.033(0.035)	0.070(0.051)
Dind7	-0.220(0.019)	-0.208(0.019)	-0.031(0.035)
Dind8	-0.076(0.025)	-0.070(0.025)	-0.208(0.019)
Dind9	-0.132(0.030)	-0.123(0.030)	-0.068(0.025)
Dind10	-0.225(0.023)	-0.222(0.023)	-0.119(0.029)
Dind11	-0.100(0.049)	-0.087(0.049)	-0.219(0.023)
Dind12	-0.137(0.022)	-0.130(0.022)	-0.087(0.049)
Dind13	-0.028(0.019)	-0.022(0.019)	-0.128(0.021)
Dind14	-0.130(0.025)	-0.126(0.025)	-0.020(0.018)
Dind15	-0.057(0.019)	-0.050(0.018)	-0.124(0.025)
Dind16	-0.003(0.031)	-0.002(0.031)	-0.048(0.018)
Dind17	0.066(0.020)	0.073(0.020)	0.000(0.031)
Schooling	0.060(0.002)	0.061(0.002)	0.075(0.020)
Experience	0.018(0.001)	0.051(0.005)	0.061(0.002)
Experience^2	-3.2E-04(2.8E-06)	-0.003(4.52E-05)	
Experience^3		6.38E-06 (1.56E-06)	
Experience^4		-5.0E-08(1.76E-08)	
R-square	0.295	0.302	0.304
Hausman test: OLS vs. Semiparametric	71.766	10.940	

Table A.3. Estimates of Real Hourly Wages Conditional on Year of Schooling and Experience

Table A.3.1 Estimates of Real Hourly Wages Conditional on Year of Schooling and Experience of 1997

M/W	Years of Experience					
School	5	10	15	20	25	30
10	5.19 (4.99,5.22)	6.44 (6.20,6.85)	7.23 (6.79,7.69)	7.85 (7.57,8.35)	7.97 (7.70,8.49)	8.14 (7.86, 8.67)
12	6.32 (6.05,6.74)	7.83 (7.52,8.36)	8.79 (8.45,9.39)	9.54 (9.18,10.19)	9.70 (9.33,10.37)	9.91 (9.53,10.59)
14	7.68 (7.33,8.23)	9.53 (9.11,10.21)	10.70 (10.24,11.47)	11.61 (11.13,12.45)	11.80 (11.31,12.66)	12.05 (11.55,12.93)
16	9.35 (8.89,10.06)	11.58 (11.05,12.48)	13.01 (12.42,14.01)	14.12 (13.49,15.21)	14.35 (13.71,15.47)	14.65 (14.00,15.80)
18	11.37 (10.78,12.28)	14.09 (13.39,15.24)	15.82 (15.05,17.12)	17.17 (16.35,18.58)	17.45 (16.62,18.89)	17.82 (16.97,19.29)
F/W						
10	4.30 (4.14,4.64)	5.12 (4.94,5.53)	5.40 (5.21,5.82)	5.67 (5.47,6.11)	5.73 (5.53,6.17)	5.78 (5.59,6.23)
12	5.34 (5.12,5.79)	6.36 (6.12,6.90)	6.70 (6.44,7.26)	7.04 (6.77,7.62)	7.11 (6.84,7.70)	7.18 (6.91,7.77)
14	6.63 (6.34,7.22)	7.90 (7.56,8.60)	8.32 (7.97,9.06)	8.74 (8.37,9.51)	8.83 (8.46,9.61)	8.91 (8.55,9.70)
16	8.23 (7.84,9.02)	9.81 (9.36,10.73)	10.33 (9.86, 11.30)	10.85 (10.36,11.87)	10.96 (10.46,11.98)	11.07 (10.57,12.10)
18	10.21 (9.69,11.24)	12.17 (11.57,13.39)	12.82 (12.19,14.10)	13.47 (12.81,14.80)	13.61 (12.94,14.95)	13.74 (13.08,15.09)

(Table A.3.1. continued)

M/NW	Years of Experience					
School	5	10	15	20	25	30
10	4.74 (4.17,5.46)	5.97 (5.25,6.88)	6.28 (5.54,7.23)	6.57 (5.81,7.56)	6.81 (6.03,7.84)	7.06 (6.24,8.12)
12	5.74 (5.00,6.67)	7.22 (6.30,8.40)	7.60 (6.64,8.83)	7.95 (6.97,9.25)	8.23 (7.23,9.58)	8.54 (7.48,9.93)
14	6.94 (5.99,8.16)	8.73 (7.55,10.27)	9.19 (7.96,10.80)	9.62 (8.35,11.30)	9.96 (8.66,11.71)	10.32 (8.97,12.14)
16	8.39 (7.18,9.97)	10.56 (9.05,12.55)	11.11 (9.54,13.20)	11.64 (10.01,13.81)	12.05 (10.38,14.31)	12.49 (10.75,14.83)
18	10.15 (8.60,12.18)	12.78 (10.84,15.34)	13.44 (11.43,16.13)	14.08 (11.99,16.88)	14.57 (12.44,17.49)	15.11 (12.88,18.13)
F/NW						
10	4.28 (3.78,4.99)	4.91 (4.34,5.72)	5.13 (4.55,5.97)	5.34 (4.73,6.21)	5.63 (4.99,6.56)	5.61 (4.99,6.52)
12	5.24 (4.58, 6.18)	6.01 (5.26,7.09)	6.28 (5.51,7.40)	6.54 (5.73,7.70)	6.89 (6.05,8.13)	6.87 (6.05,8.07)
14	6.42 (5.56,7.66)	7.36 (6.38,8.79)	7.70 (6.68,9.16)	8.01 (6.95,9.54)	8.45 (7.34,10.07)	8.42 (7.33,10.00)
16	7.86 (6.73,9.49)	9.02 (7.73,10.89)	9.43 (8.10,11.35)	9.81 (8.42,11.82)	10.35 (8.89,12.48)	10.31 (8.88,12.39)
18	9.63 (8.16,11.76)	11.05 (9.37,13.49)	11.55 (9.81,14.07)	12.01 (10.21,14.64)	12.67 (10.78,15.46)	12.63 (10.77,15.36)

Table A.3.2 Estimates of Real Hourly Wages Conditional on Year of Schooling and Experience of 1988

M/W	Years of Experience					
School	5	10	15	20	25	30
10	5.70 (5.54,5.99)	7.23 (7.09,7.66)	7.94 (7.81,8.43)	8.46 (8.34,9.00)	8.95 (8.82,9.51)	9.25 (9.13,9.84)
12	6.67 (6.47,7.04)	8.46 (8.28,9.00)	9.29 (9.12,9.90)	9.90 (9.74,10.57)	10.47 (10.30,11.17)	10.82 (10.66,11.56)
14	7.80 (7.56,8.27)	9.90 (9.67,10.57)	10.87 (10.65,11.63)	11.59 (11.37,12.41)	12.25 (12.03,13.12)	12.66 (12.46,13.58)
16	9.13 (8.83,9.71)	11.59 (11.30,12.41)	12.72 (12.44,13.66)	13.56 (13.28,14.58)	14.33 (14.05,15.41)	14.82 (14.55,15.95)
18	10.68 (10.31,11.40)	13.56 (13.20,14.58)	14.88 (14.53,16.04)	15.86 (15.51,17.12)	16.77 (16.41,18.10)	17.34 (16.99,18.73)
F/W						
10	4.75 (4.59,5.04)	5.66 (5.47,6.00)	5.89 (5.70,6.25)	5.90 (5.71,6.25)	6.03 (5.84,6.40)	6.05 (5.86,6.41)
12	5.58 (4.36,5.93)	6.64 (6.40, 7.07)	6.92 (6.67,7.36)	6.92 (6.67,7.37)	7.08 (6.84,7.53)	7.10 (6.86,7.55)
14	6.55 (6.27,6.99)	7.80 (7.48,8.33)	8.12 (7.79,8.67)	8.12 (7.80,8.67)	8.31 (7.99,8.87)	8.34 (8.02,8.89)
16	7.68 (7.33,8.23)	9.15 (8.75,9.81)	9.53 (9.12,10.21)	9.53 (9.13,10.22)	9.76 (9.35,10.45)	9.78 (9.38,10.48)
18	9.02 (8.58,9.70)	10.74 (10.23,11.55)	11.18 (10.66,12.03)	11.19 (10.67,12.03)	11.45 (10.93,12.31)	11.48 (10.97,12.34)

(Table A.3.2. continued)

M/NW	Years of Experience					
School	5	10	15	20	25	30
10	5.34 (4.82,5.98)	6.22 (5.64,7.01)	6.77 (6.17,7.66)	7.47 (6.83,8.45)	7.64 (6.99,8.63)	8.10 (7.42,9.14)
12	6.20 (5.55,7.00)	7.22 (6.51,8.20)	7.86 (7.12,8.96)	8.67 (7.87,9.89)	8.86 (8.06,10.11)	9.39 (8.56,10.69)
14	7.19 (6.41,8.19)	8.37 (7.50,9.60)	9.12 (8.21,10.49)	10.06 (9.08,11.58)	10.28 (9.30,11.83)	10.90 (9.87,12.52)
16	8.35 (7.39,9.59)	9.71 (8.65,11.24)	10.58 (9.46,12.28)	11.67 (10.47,13.56)	11.93 (10.72,13.85)	12.64 (11.38,14.66)
18	9.68 (8.52,11.23)	11.27 (9.98,13.16)	12.27 (10.91,14.38)	13.54 (12.07,15.87)	13.84 (12.36,16.22)	14.67 (13.12,17.16)
F/NW						
10	4.64 (4.19,5.20)	5.28 (4.79,5.92)	5.71 (5.18,6.41)	5.87 (5.33,6.57)	6.09 (5.54,6.81)	6.05 (5.51,6.76)
12	5.40 (4.84,6.10)	6.14 (5.52,6.94)	6.64 (5.98,7.51)	6.82 (6.15,7.71)	7.08 (6.39,7.98)	7.03 (6.36,7.93)
14	6.28 (5.58,7.15)	7.14 (6.37,8.14)	7.73 (6.89,8.81)	7.93 (7.09,9.03)	8.23 (7.37,9.36)	8.18 (7.34,9.29)
16	7.30 (6.44,8.38)	8.30 (7.35,9.54)	8.98 (7.95,10.33)	9.22 (8.18,10.59)	9.57 (8.50,10.97)	9.51 (8.46,10.90)
18	8.49 (7.42,9.82)	9.65 (8.48,11.19)	10.45 (9.17,12.11)	10.72 (9.43,12.42)	11.13 (9.81,12.86)	11.06 (9.76,12.78)

Table A.3.3 Estimates of Real Hourly Wages Conditional on Year of Schooling and Experience of 1979

M/W	Years of Experience					
School	5	10	15	20	25	30
10	7.11 (6.91,7.43)	8.67 (8.49,9.13)	9.71 (9.53,10.24)	10.16 (9.98,10.71)	10.44 (10.25,11.01)	10.66 (10.47,11.24)
12	8.12 (7.88,8.52)	9.91 (9.68,10.47)	11.10 (10.87,11.74)	11.61 (11.38,12.28)	11.93 (11.69,12.62)	12.18 (11.95,12.89)
14	9.28 (8.99,9.76)	11.33 (11.04,12.00)	12.69 (12.40,13.46)	13.27 (12.98,14.08)	13.63 (13.34,14.47)	13.92 (13.62,14.78)
16	10.61 (10.25,11.19)	12.94 (12.59,13.76)	14.50 (14.14,15.43)	15.17 (14.80,16.15)	15.58 (15.21,16.59)	15.91 (15.54,16.94)
18	12.12 (11.69,12.83)	14.79 (14.36,15.77)	16.57 (16.13,17.70)	17.33 (16.88,18.52)	17.80 (17.35,19.02)	18.18 (17.72,19.43)
F/W						
10	4.28 (3.78,4.99)	4.91 (4.34,5.72)	5.13 (4.55,5.97)	5.34 (4.73,6.21)	5.63 (4.99,6.56)	5.61 (4.99,6.52)
12	5.24 (4.58,6.18)	6.01 (5.26,7.09)	6.28 (5.51,7.40)	6.54 (5.73,7.70)	6.89 (6.05,8.13)	6.87 (6.05,8.07)
14	6.42 (5.56,7.66)	7.36 (6.38,8.79)	7.70 (6.68,9.16)	8.01 (6.95,9.54)	8.45 (7.34,10.07)	8.42 (7.33,10.00)
16	7.86 (6.73,9.49)	9.02 (7.73,10.89)	9.43 (8.10,11.35)	9.81 (8.42,11.82)	10.35 (8.89,12.48)	10.31 (8.88,12.39)
18	9.63 (8.16,11.76)	11.05 (9.37,13.49)	11.55 (9.81,14.07)	12.01 (10.21,14.64)	12.67 (10.78,15.46)	12.63 (10.77,15.36)

(Table A.3.3. continued)

M/NW	Years of Experience					
School	5	10	15	20	25	30
10	6.70 (6.02,7.48)	8.02 (7.27,9.02)	8.27 (7.51,9.30)	9.01 (8.21,10.12)	8.83 (8.06,9.92)	8.98 (8.19,10.07)
12	7.51 (6.70,8.45)	9.00 (8.09,10.19)	9.27 (8.36,10.52)	10.10 (9.14,11.44)	9.90 (8.98,11.21)	10.06 (9.12,11.38)
14	8.42 (7.46,9.56)	10.09 (9.01,11.52)	10.40 (9.31,11.89)	11.32 (10.17,12.93)	11.10 (10.00,12.68)	11.28 (10.16,12.87)
16	9.44 (8.31,10.81)	11.31 (10.04,13.03)	11.66 (10.37,13.44)	12.69 (11.33,14.62)	12.44 (11.13,14.33)	12.65 (11.31,14.55)
18	10.58 (9.25,12.22)	12.68 (11.17,14.73)	13.07 (11.54,15.20)	14.23 (12.61,16.53)	13.95 (12.39,16.20)	14.18 (12.59,16.45)
F/NW						
10	5.36 (4.84,5.97)	6.21 (5.63,6.93)	6.45 (5.86,7.19)	6.42 (5.83,7.15)	6.38 (5.80,7.09)	6.47 (5.89,7.19)
12	6.06 (5.43,6.80)	7.03 (6.32,7.90)	7.30 (6.58,8.20)	7.26 (6.54,8.16)	7.21 (6.51,8.09)	7.32 (6.62,8.20)
14	6.86 (6.09,7.76)	7.95 (7.09,9.01)	8.26 (7.38,9.35)	8.21 (7.34,9.30)	8.16 (7.31,9.23)	8.28 (7.43,9.35)
16	7.75 (6.84,8.85)	8.99 (7.96,10.28)	9.34 (8.29,10.66)	9.29 (8.24,10.61)	9.23 (8.20,10.52)	9.37 (8.34,10.66)
18	8.77 (7.68,10.10)	10.17 (8.94,11.72)	10.57 (9.30,12.16)	10.51 (9.25,12.10)	10.44 (9.21,12.00)	10.60 (9.36,12.16)

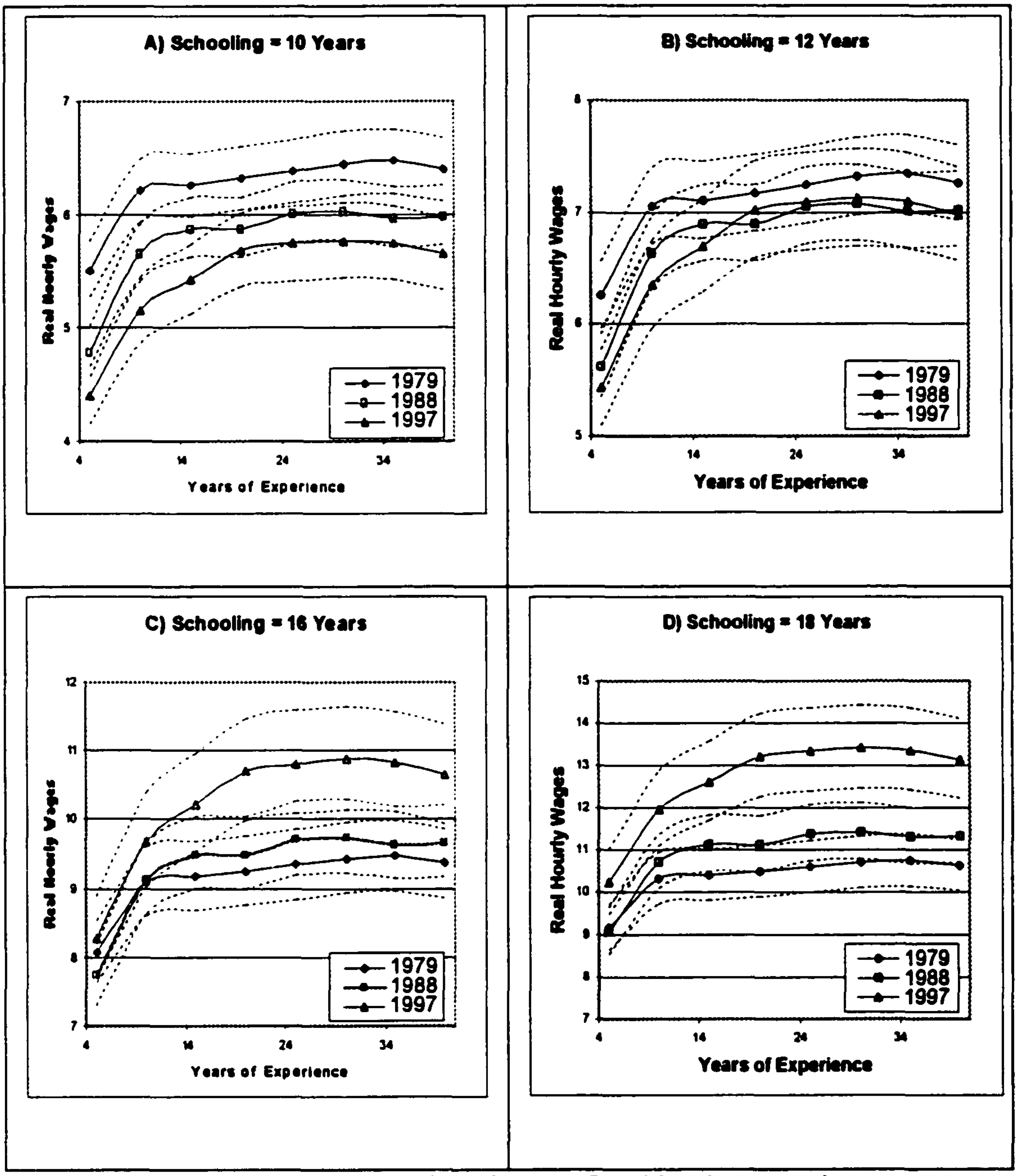
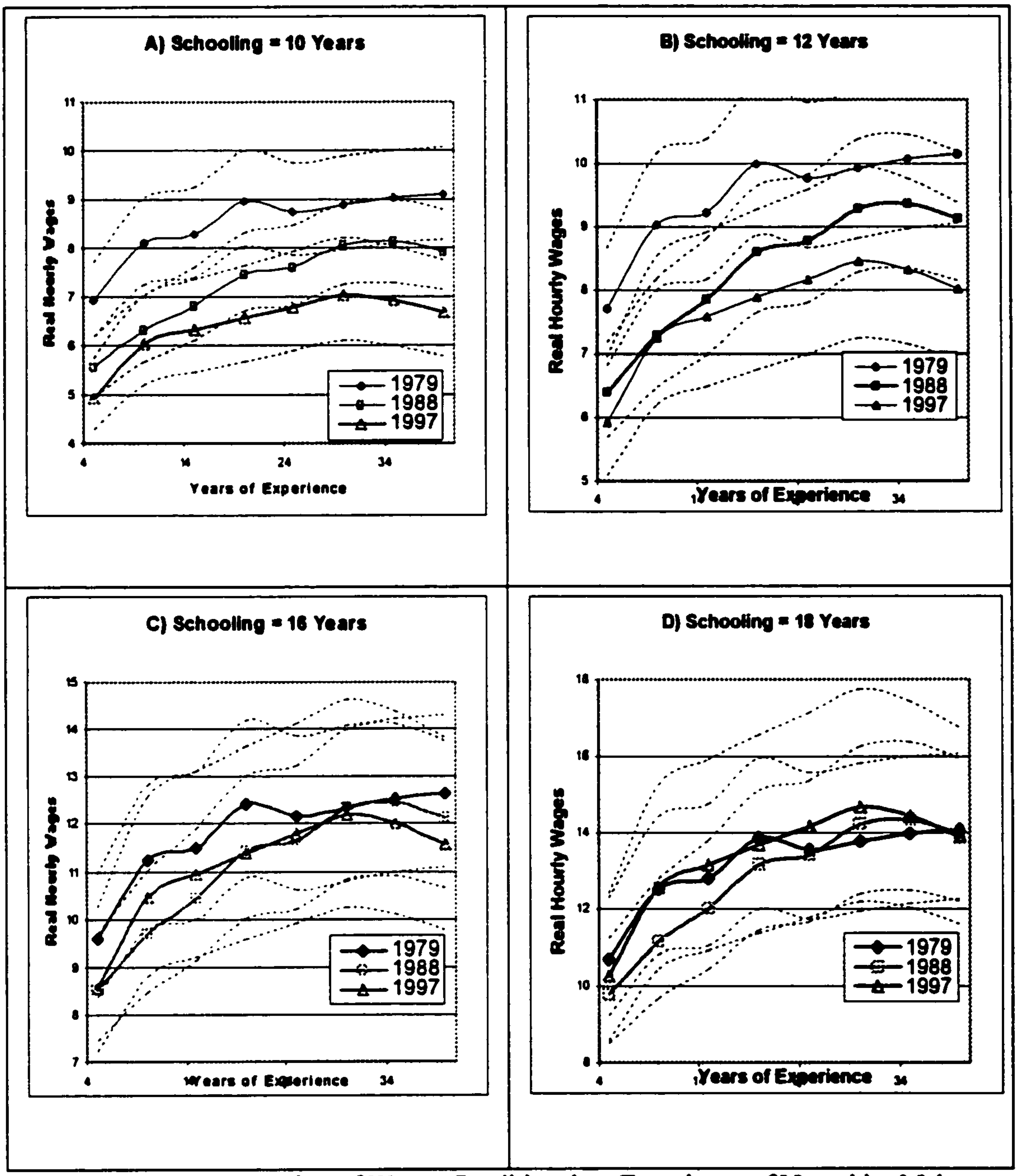
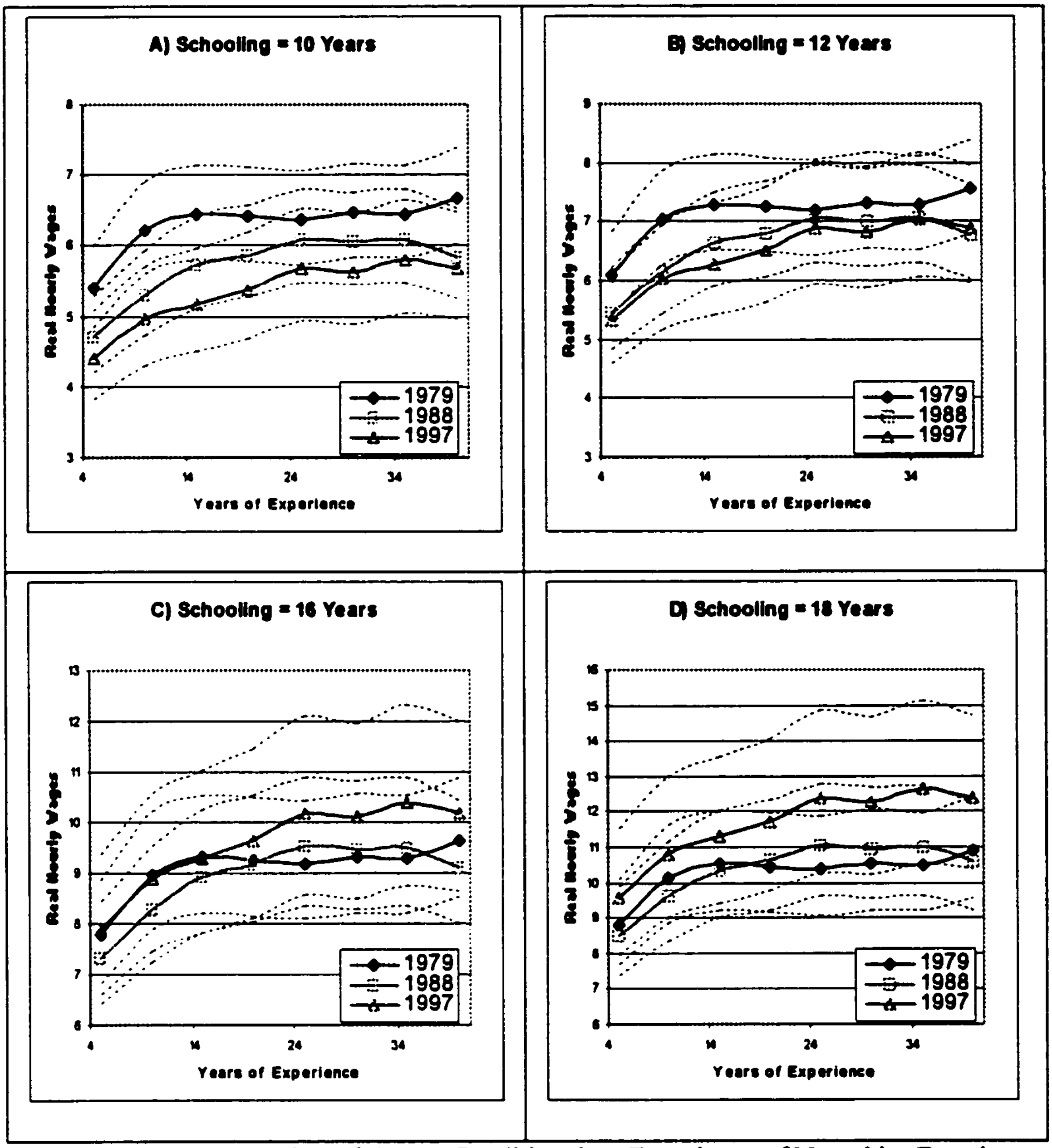


Figure A.2. Mean Regression of Wages Conditional on Experience

A.2.1. Mean Regression of Wages Conditional on Experience of White-Female



A.2.2. Mean Regression of Wages Conditional on Experience of Nonwhite-Male



A.2.3. Mean Regression of Wages Conditional on Experience of Nonwhite-Female

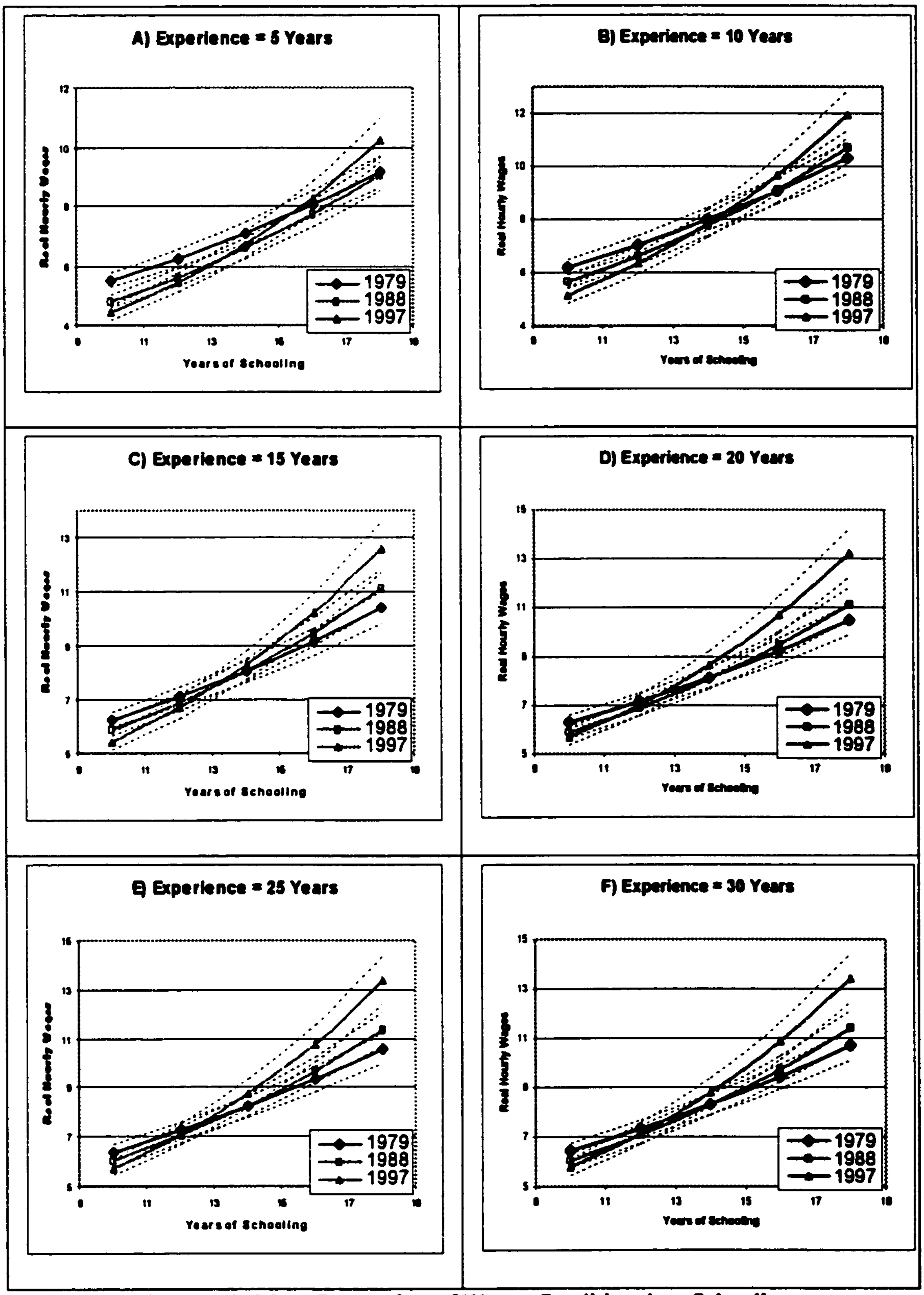
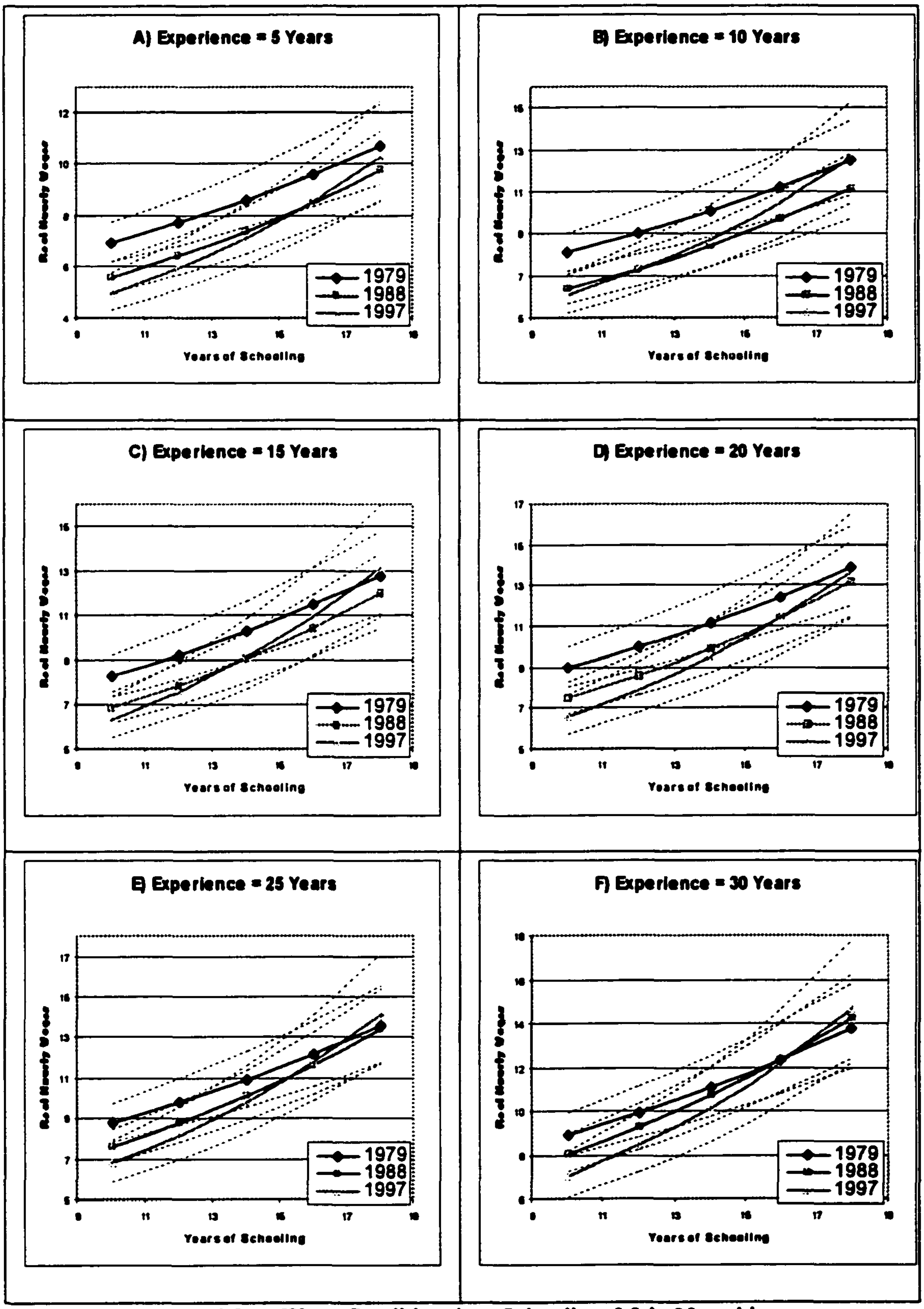
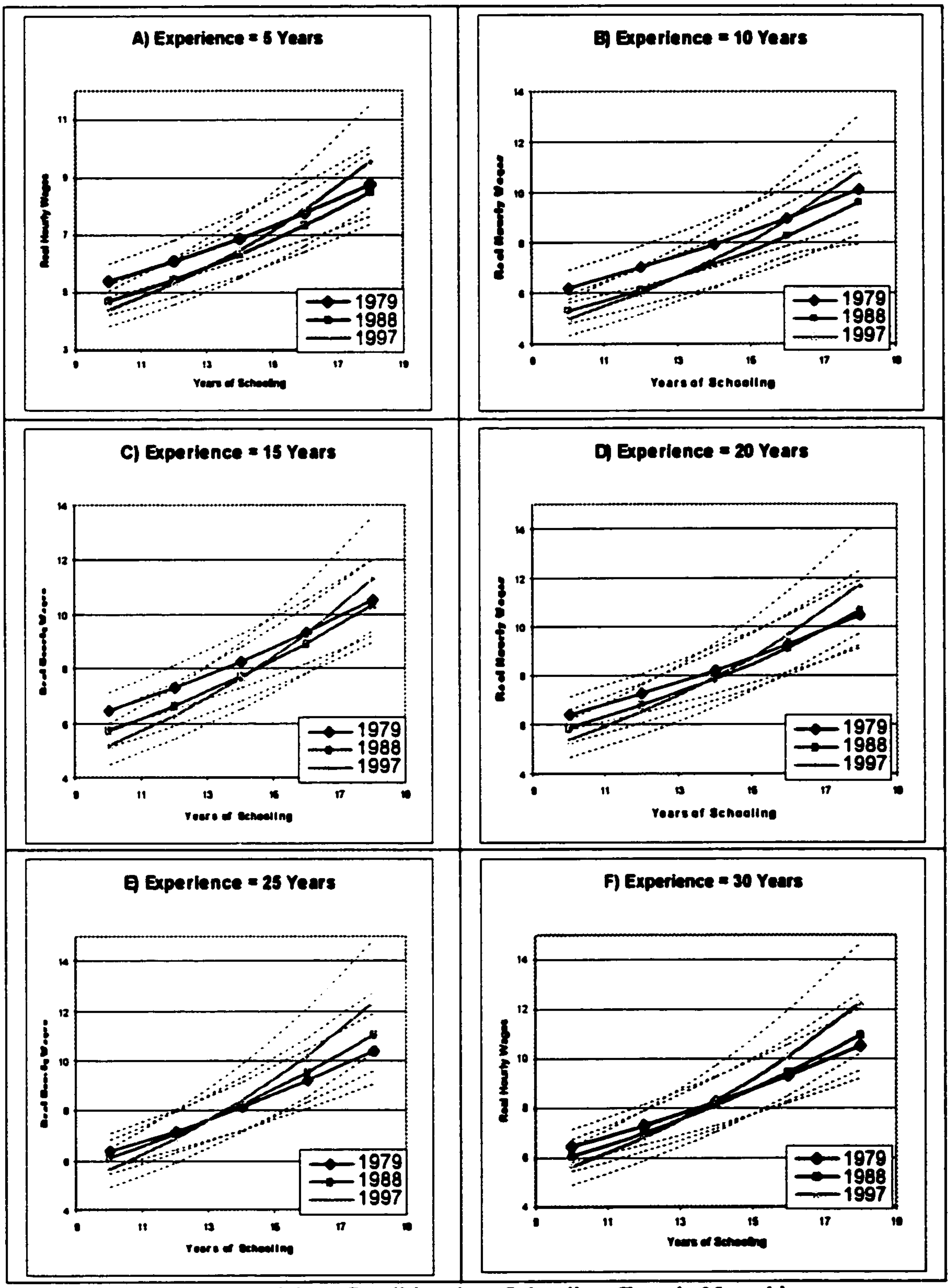


Figure A.3. Mean Regression of Wages Conditional on Schooling

A.3.1. Mean Wage Conditional on Schooling, Female-White



A.3.2. Mean Wage Conditional on Schooling, Male-Nonwhite



A.3.3. Mean Wage Conditional on Schooling, Female-Nonwhite

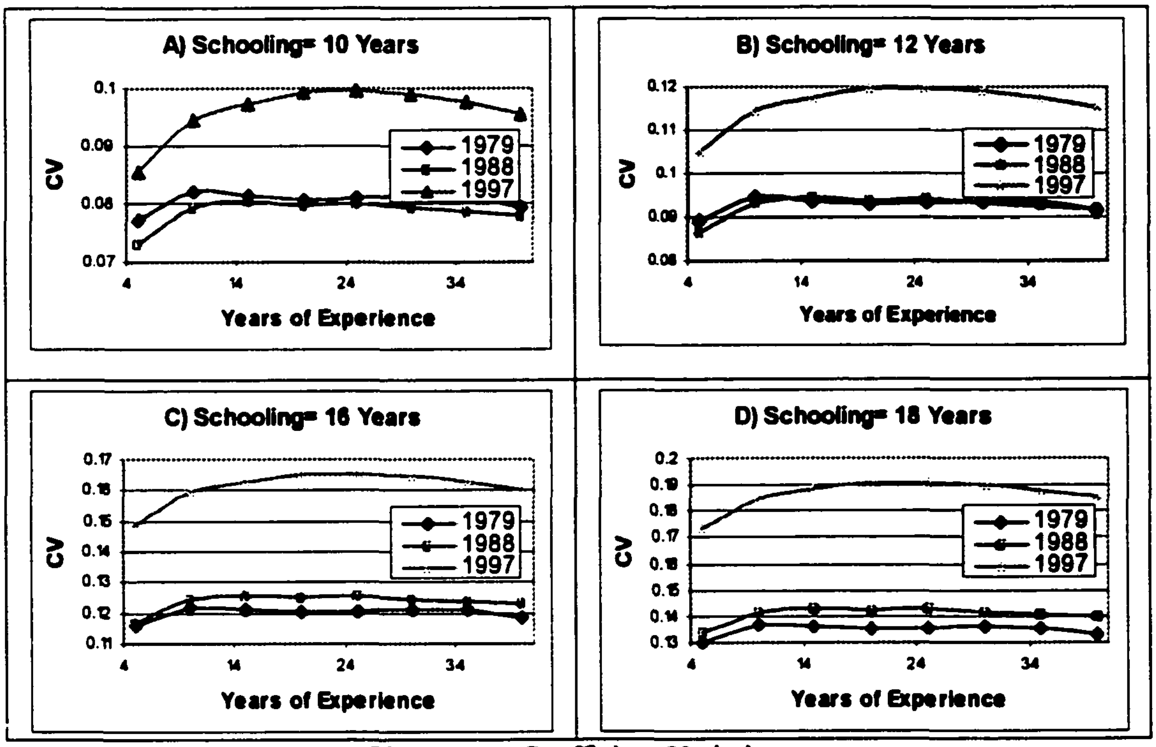
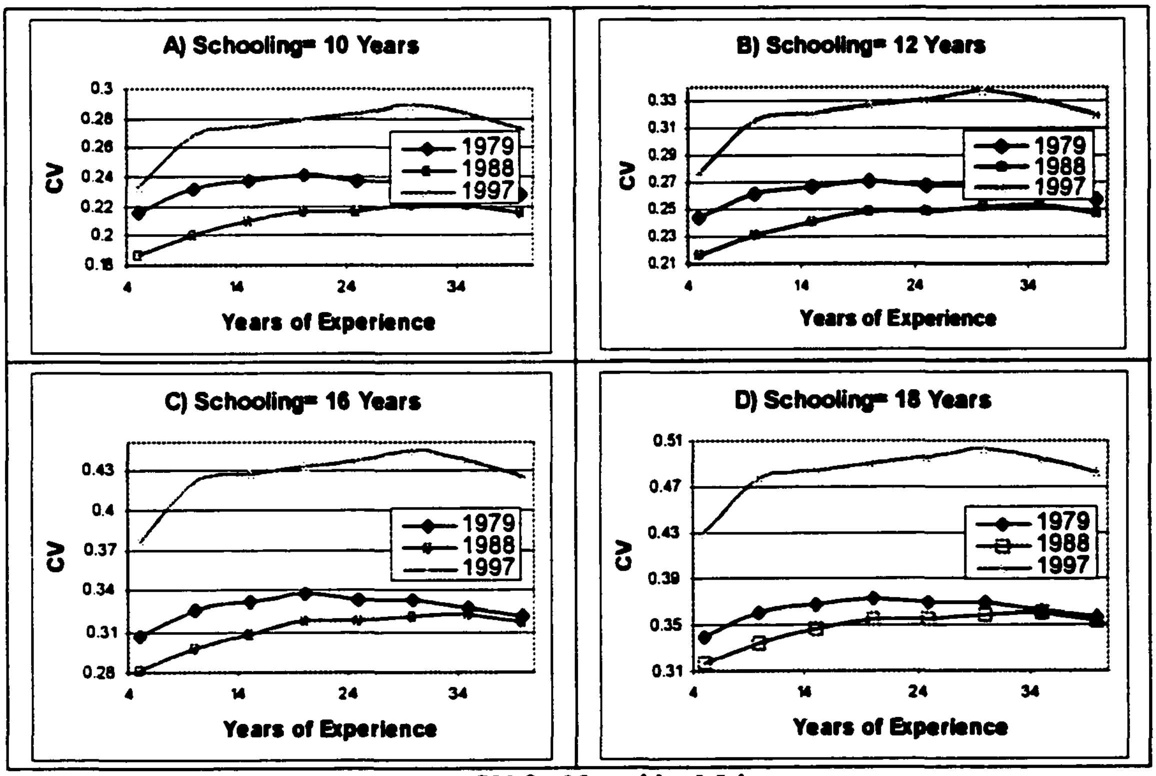
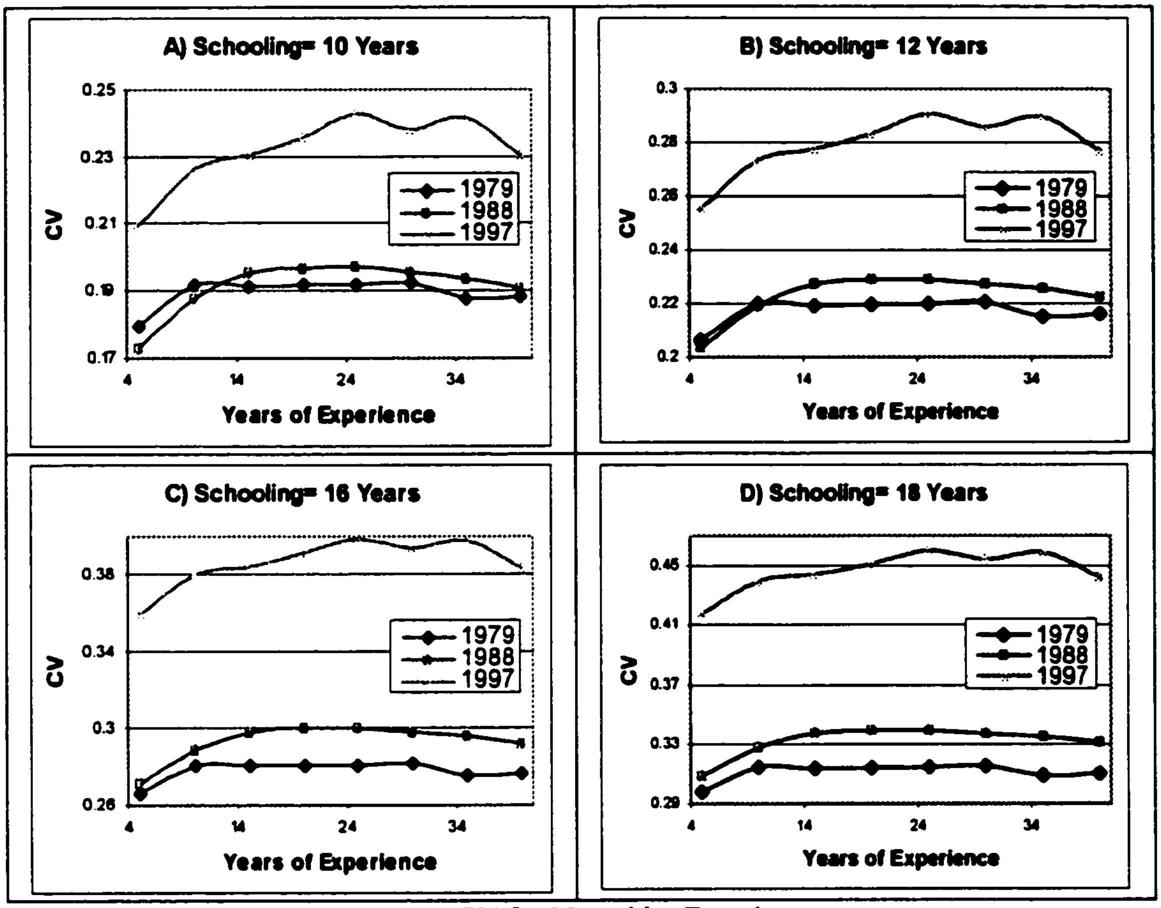


Figure A.4. Coefficient Variation

A.4.1. CV of Female-White



A.4.2 CV for Nonwhite-Male



A.4.3 CV for Nonwhite-Female

Appendix Table A.4. Results for 13 Dummy Variables Estimation

Variable		Estimate 1979-1988	Standard error	Estimate 1988-1997	Standard error
Experience (men)	X<10	-0.054	0.147	-0.212	0.161
	10<x<20	0.039	0.135	-0.213	0.662
	20<x<30	0.067	0.135	0.001	0.145
women	X<10	-0.008	0.139	-0.394	0.155
	10<x<20	0.206	0.136	-0.104	0.153
	20<x<30	-0.003	0.135	-0.031	0.152
Schooling (men)	S<12	-0.696	0.264	-0.236	0.279
	S=12	-0.245	0.139	-0.200	0.143
	12<S<16	-0.186	0.139	-0.130	0.143
women	S<12	-0.068	0.141	-0.144	0.157
	S=12	-0.358	0.133	-0.060	0.141
	12<S<16	-0.158	0.132	-0.084	0.140
Men/women		-0.122	0.134	-0.030	0.141

Table A.5. The Estimates of General Technical Changes with and without Dummy Variables

Variable	Estimate (w/o dummies)	Estimate (w dummies)
Constant	-0.015 (0.003)	-0.015 (0.004)
$d^2 N_a$	-0.558 (0.350)	-0.499 (0.427)
Men-dropout		0.015 (0.022)
Women-dropout		0.004 (0.023)
Men-high school		0.0004 (0.022)
Women-high school		0.001 (0.023)
Men-dropout w/ 10<X<20		0.006 (0.022)

Table A.6. Estimated Relative Wage Changes and Employment Distributions

		Relative Wage Changes		Employment Distributions		
Experience	Education	1979-88	1988-97	1979	1988	1997
Men:0-9	DO	-0.104	-0.083	0.016	0.011	0.012
	HS	-0.041	-0.058	0.075	0.056	0.04
	SC	0.060	-0.068	0.052	0.042	0.04
	CO	0.070	-0.036	0.053	0.045	0.035
10-19	DO	0.031	-0.043	0.018	0.015	0.012
	HS	0.088	-0.004	0.051	0.063	0.053
	SC	0.260	-0.007	0.031	0.037	0.042
	CO	0.240	0.042	0.035	0.05	0.043
20-29	DO	-0.110	-0.108	0.022	0.013	0.013
	HS	-0.087	-0.067	0.037	0.041	0.05
	SC	0.146	-0.086	0.017	0.023	0.039
	CO	0.057	-0.041	0.022	0.028	0.041
30+	DO	0.074	-0.053	0.05	0.026	0.015
	HS	0.066	0.000	0.049	0.04	0.04
	SC	0.419	-0.004	0.016	0.015	0.025
	CO	0.262	0.058	0.015	0.017	0.019
Women:	DO	-0.014	-0.133	0.009	0.006	0.007
0-9	HS	0.021	-0.047	0.064	0.051	0.032
	SC	0.307	-0.111	0.05	0.05	0.044
	CO	0.121	-0.014	0.042	0.045	0.038
10-19	DO	0.207	-0.077	0.012	0.009	0.008
	HS	0.192	0.033	0.048	0.055	0.045
	SC	0.628	-0.027	0.023	0.037	0.041
	CO	0.340	0.101	0.019	0.039	0.04
20-29	DO	-0.291	-0.111	0.015	0.011	0.009
	HS	-0.181	-0.042	0.038	0.046	0.048
	SC	0.126	-0.084	0.013	0.023	0.04
	CO	-0.079	-0.008	0.012	0.02	0.035
30+	DO	-0.102	-0.042	0.028	0.018	0.011
	HS	-0.038	0.040	0.047	0.044	0.044
	SC	0.436	0.018	0.013	0.014	0.025
	CO	0.112	0.109	0.008	0.009	0.014

VITA

Kang-Sun Lee received his following degrees: B.A. (1991) and M.A. (1993) in Economics from Dankook University, Seoul, Korea (ROK); M.S. (1998) in Economics in Louisiana State University. He had worked as a research assistant to Dr. Seung-Joon Park at Dankook University in Seoul, Korea, from 1992 to 1994. At Louisiana State University, he worked as a research assistant to Dr. Hill from 1997 to 1999, Dr. Kaz Miyagiwa in 2000, and Dr. M. Dek Terrell in 2001. Currently he is a candidate for the degree of Doctor of Philosophy at Louisiana State University, which will be awarded at the December 2001, Commencement.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Kang-Sun Lee

Major Field: Economics

Title of Dissertation: Essays on Semiparametric Estimation

Approved:

R. Carter Hill

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

James P. Geaghan

M. Dekalb Terrell

Date of Examination:

November 2, 2001