Parametric and Semiparametric Estimations of the Return to Schooling of Wage Workers in Canada

By

Jin John Lei
Bachelor of Arts in Economics
University of Colorado at Denver, 2002

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

In the
Department
of
Economics

SIMON FRASER UNIVERSITY
Spring 2005

APPROVAL

Name: **Jin John Lei**

Degree: **M. A. (Economics)**

Title of Project: **Parametric And Semiparametric Estimations Of The Return To Schooling Of Wage Workers In Canada**

Examining Committee:

Chair: **Gord Myers**

Krishna Pendakur
Senior Supervisor

Simon Woodcock
Supervisor

Jane Friesen
Internal Examiner

Date Approved: Wednesday, March 30, 2005

ABSTRACT

This project presents estimates of the return to education for wage workers in Canada using 1994 SLID microdata. To correct for non-random sampling, the parametric two-stage estimators of wage equation proposed by Heckman (1974, 1979) and its semiparametric counterparts introduced by Newey (1999) and Robinson (1988) are employed. Specification tests proposed by Horowitz (1993) and Hausman (1978) suggest semiparametric approaches are superior to parametric ones in terms of flexibility (consistency or at least efficiency gains), generality and predictability. The endogeneity problem of schooling is considered and schooling is instrumented by parental education. The estimated return for female paid-workers is 4.4~4.9% and 3.9% for male paid-workers in Canada.

DEDICATION

老妈, 感谢你多年来对我的养育与教导. 没有你, 这份厚厚的毕业论文也不能成为永久的历史摆放在学校的图书馆里边供人参详. 我深知, 这一切只是我人生中的一个阶段性的标记. 我会继续努力, 以报答来自你的生生不息的爱.

To my parents, who have sustained me.

To my lovely 小粉猪, *(D.F.Ha, HaHa, D.F.Bao)*

Shan Chen,

who has given WaWa her love and support.

To my cutest 小驴驴, *Jie Zhang,*

for her continuing encouragement.

ACKNOWLEDGEMENTS

I owe particular thanks to Professor Krishna Pendakur and Professor Simon Woodcock for their insight, guidance and support. I am also very grateful to Professor Jane Friesen for her helpful suggestions and comments.

I offer my sincere gratitude to the faculty, staff and my fellow students at the SFU Department of Economics, who have combined to create a stimulating synergy for studies and research in our field.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Over the past decade there have been a series of studies attempting to understand the causal links between education and labour market success. To policy makers, it is crucial to comprehend the questions: are the higher earnings observed for better educated people determined only by their higher education level, or do they reflect inherent ability differences that correlate with educational attainment; treating schooling as a way to increase one's (non)market productivity, is it a meaningful thing to increase funding for public education and particularly to which group of people under tight government's budget constraints?

This paper applies both traditional parametric and flexible semiparametric regressions to estimate the return to schooling of wage workers in Canada. There has been a common practice that most researchers attempting to estimate the return to schooling only worry about measurement error and endogeneity of schooling but not the non-random sample problem. In reality, an individual chooses to be in the observed sample of workers if her offered wage exceeds her reservation wage. Since workers and non-workers may have systematically different characteristics, using non-random sample cannot yield representative estimates for the population. Thus, without correcting for self-selection problem, the estimate of return to schooling is inconsistent. In contrast to other studies, the aim of this paper is to obtain a consistent estimate of the return to education in Canada.

In this paper, Heckman (1974) and Heckman (1979) are used to handle the censored sample problem. They are, however, sensitive to the assumed parametric distribution of the unobserved disturbance terms. If the model is misspecified, this may bias estimates of the parameters of interest. A series of specification tests are also used to prove the necessity of semiparametric approach, which include Horowitz (1993), Bera and Jarque and Lee (1987) and Hausman (1978). It is shown that a semiparametric approach is preferable to the parametric alternative when education is exogenous. Semiparametric estimators considered include Newey (1999) and Robinson (1988). They give reasonable estimates of a structural Buchinsky (1998) wage equation and labour force participation equation.

Within the parametric context, first, conventional OLS and IV approaches are revisited to correct for endogeneity and measurement error in the schooling variable (See Card (2001)). Given the potential endogeneity of schooling choice, the relevance of an instrument is examined by Davidson and MacKinnon (1993, Ch7, page 237-242) and Durbin-Wu-Hausman (1978) test. The F test for exclusion restrictions and Sargan test (1958) are carried out to verify the validity of parental education as instruments.

For female workers, Newey (1999) and Robinson (1988) are the preferred specification.

For male workers, 2SLS is the preferred specification. The estimated return is 4.4~4.9% for working females and 3.9% for working males in Canada.

There are very few studies in the literature that use semiparametric methods to estimate the return to education (See Andrews, Schafgans (1998), Martins (2001) and Christofides, Li, Liu and Min (2003)). Moreover, there has been no serious attention paid to the endogeneity problem of schooling when one estimates the return to education in this framework. In this paper, I address and show, without proposing any remedy, the potential failure of the semiparametric approach when schooling is endogenous. Blundell and Powell (2004), proposes a nonparametric control function and pairwise differencing approach to solve the problem of endogenous regressors in a semiparametric single index model. This is beyond the scope of this paper.

The layout of the paper is as follows. Chapter 2 describes the econometric model. Chapter 3 presents the data and results. Chapter 4 gives conclusions and extensions.

CHAPTER 2

THE MODEL

2.1 Parametric Model

Sample selection models (see Heckman 1974, 1979) are applied when the individuals in the sample are not randomly chosen from the population from which researchers and policy makers would like to draw inferences and extrapolate it to the entire population. The estimation procedure of selectivity model is composed of two steps. The first step is to estimate a binary decision equation which takes account of the non-representative nature of the sample. The second step estimates the parameters of an outcome equation on which the researcher's interest is centred. In the study of the determinants of wages of wage workers, workers might be self-selected into the sample on the basis of explanatory variables that are correlated with the dependent variable of the outcome equation. The typical sample selection model is proposed by Heckman (1979). It has the form

Participation equation: $$D_i^* = Z_i^T \gamma + u_i \text{; for } i = 1, \ldots\ldots, N \qquad (1)$$

$$D_i = 1 \text{ if } D_i^* > 0 \text{; } D_i = 0 \text{ otherwise} \qquad (2)$$

Outcome equation: $$Y_i = Y_i^* D_i \qquad (3)$$

$$Y_i^* = \alpha_0 + X_i^T \beta_0 + \varepsilon_i = X_i^T \beta + \varepsilon_i \text{; for } i = 1, \ldots\ldots, N, \qquad (4)$$

where (D_i, Z_i, Y_i, X_i) are observed random variables, in particular, (Z_i, X_i) are vectors of exogenous variables, Y_i^* is a latent (potential) logged wage, D_i^*, is the latent propensity to select into the sample, with associated indicator function D_i, $(\gamma, \alpha_0, \beta_0)$ are unknown parameter vectors, and (u_i, ε_i) are zero mean error terms. The subsample where $D_i = 1$ has size denoted n in contrast to N, the size of complete sample. The self-selection problem arises if u_i (the impact of unobservable individuals' characteristic (motivation or ability) on the decision to select into the sample) and ε_i (the impact of unobservable characteristic on the outcome of interest, the wage) are correlated. Then if higher motivation leads to higher wages and to an increased propensity to work, u_i and ε_i would be positively correlated and the observed wages of the working subpopulation, n, would be too high to represent the whole population, ceteris paribus. Econometric analysis of labour force participation has been traditionally undertaken by Generalized Linear Models (GLM) (See Appendix A).

2.1.1 Heckman Maximum Likelihood Estimation (1974)

The methodology of maximum likelihood estimation solving selectivity bias was first introduced by Heckman (1974). Under joint normality of the errors ε, u, the maximum likelihood estimator maximizes average log likelihood function:

$$\ln L(\beta, \gamma, \sigma_{\varepsilon u}, \sigma_\varepsilon^2) = \frac{1}{N}\sum_{i=1}^{N}\left[D_i \cdot \ln \int_{-Z_i^{\mathrm{T}}\gamma}^{\infty} \phi_{\varepsilon u}(Y_i - X_i^{\mathrm{T}}\beta, u)du + (1 - D_i) \cdot \ln \int_{-\infty}^{\infty} \int_{-Z_i^{\mathrm{T}}\gamma}^{\infty} \phi_{\varepsilon u}(\varepsilon, u)dud\varepsilon\right] \quad (5)$$

where $\phi_{\varepsilon u}$ is the bivariate normal density function, see Vella (1998). The MLE reaches the Cramer Rao bound, producing efficient estimates $\hat{\beta}, \hat{\gamma}, \hat{\sigma}_{\varepsilon u}$ and $\hat{\sigma}_{\varepsilon}^2$.

2.1.2 Heckman Two Step (1979)

The Heckit procedure was introduced by Heckman (1979). It is a second-best alternative to maximum likelihood in terms of efficiency. Harking back to equation (1) to (4), since $\begin{pmatrix} \varepsilon_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon u} \\ \sigma_{\varepsilon u} & 1 \end{pmatrix}\right)$ by assumption, where σ_u^2 is normalized to 1, ε_i, u_i are assumed independently and identically distributed and are independent of Z_i, the outcome equation is:

$$E(Y_i \mid X_i) = E(Y_i^* \mid X_i, D_i^* > 0) = E(X_i^{\mathrm{T}}\beta \mid X_i, D_i^* > 0) + E(\varepsilon_i \mid X_i, D_i^* > 0)$$

$$= X_i^{\mathrm{T}}\beta + E(\varepsilon_i \mid X_i, Z_i^{\mathrm{T}}\gamma > -u_i) \tag{6}$$

$$= X_i^{\mathrm{T}}\beta + \sigma_{\varepsilon u} \cdot \frac{\phi(Z_i^{\mathrm{T}}\gamma)}{\Phi(Z_i^{\mathrm{T}}\gamma)},$$

where $\lambda(\bullet) = \dfrac{\phi(\bullet)}{\Phi(\bullet)}$ is the inverse Mill's ratio, $\phi(\bullet)$ and $\Phi(\bullet)$ are the univariate probability density and cumulative distribution function respectively of the standard normal distribution, and $\sigma_{\varepsilon u}$ is the covariance between ε and u. The parameters of the model β, γ and $\sigma_{\varepsilon u}$ can be consistently estimated by the following two-step procedure proposed in Heckman (1979).

1. Probit step: estimate γ by fitting the Probit model $\Pr ob(D=1)=\Phi(Z^{\mathrm{T}}\gamma)$ using all observations, N.

2. OLS step: using only observations with $D=1$ estimate the regression function $E(Y_i \mid X_i)=X_i^{\mathrm{T}}\beta+\sigma_{\varepsilon u}\cdot\dfrac{\phi(Z_i^{\mathrm{T}}\gamma)}{\Phi(Z_i^{\mathrm{T}}\gamma)}$ by an OLS regression of the observed Y_i on X_i and $\dfrac{\phi(Z_i^{\mathrm{T}}\hat{\gamma})}{\Phi(Z_i^{\mathrm{T}}\hat{\gamma})}$ where $\hat{\gamma}$ is the first step estimate of γ.

2.2 Semiparametric Model

In the sample selection model deviations from normality can lead to biased and inconsistent estimates (See Schafgans (1997)). A semiparametric estimation technique described in this section relaxes the normality assumption and lets the data indicate what the distribution of the error term is.

Semiparametric models combine components of parametric and nonparametric models. Hence, they possess the virtue of easy interpretability of parametric models and flexibility of nonparametric models[1]. They avoid the assumptions of parametric models and functional form since they do not assume a known link function, and solve the curse of dimensionality[2] of nonparametric models. However, if the distribution of parametric

[1] Like parametric models and unlike nonparametric models, semiparametric models permit intermediate capability for extrapolation that captures the a priori assumptions about the data generation process (see Horowitz 1998). That is, it provides predictions of $E(D \mid z)$ at points z that are out of the support of Z.

[2] The precision of a nonparametric estimator decreases rapidly as the number of continuously distributed components of independent variables increases.

distributional assumption is valid, a semiparametric estimator will be less efficient than its parametric counterpart.

This section presents two semiparametric estimators of β_1 for a single structural equation:

$$Y_i = \beta_0 + X_i^{\mathrm{T}} \beta_1 + \lambda(Z_i^{\mathrm{T}} \gamma) + \xi_i \text{ with } E(\xi \mid X, Z, D = 1) = 0 \tag{7}$$

This corresponds a selection model, semiparameterized by a linear index $\theta(Z_i^{\mathrm{T}} \gamma) = Z_i^{\mathrm{T}} \gamma$ and linear functional relationship between Y_i and X_i while $\lambda(\bullet)$ and ξ_i are unspecified. The intercept and its standard error in Robinson (1988) and Newey (1999) are not identified, since it is subsumed in the link function

$$Y_i = \beta_0 + X_i^{\mathrm{T}} \beta_1 + \lambda(Z_i^{\mathrm{T}} \gamma) + \xi_i = X_i^{\mathrm{T}} \beta_1 + \tilde{\lambda}(Z_i^{\mathrm{T}} \gamma) + \xi_i \tag{8}$$

Andrews and Schafgans (1998) derive a consistent estimator of the constant term under some regularity conditions. Schafgans (2004) investigated the finite sample properties of the semiparametric estimator of the intercept of a censored regression model. Robinson's estimator of β_1 is identified by differencing out the selection bias $\lambda(\bullet)$ by Kernel method. In contrast, Newey's estimator of β_1 is identified by replacing the selection bias $\lambda(\bullet)$ by an infinite series approximation.

2.2.1 Newey (1999)

2.2.1.1 Step 1: Klein and Spady (1993)

Klein and Spady (1993) proposed an estimator of the selection equation $D_i = \varphi(\theta(Z_i;\gamma), u_i)$ (See equation (1) and (2)). They specify the continuous index function $\theta(\bullet)$ whose range is [0,1] and leave the distribution of the error term u_i unrestricted. They specify $\varphi(\bullet)$ as the usual threshold crossing indicator function, and normalize the unidentified threshold to zero. Then the selection equation is

$$D_i = 1(\theta(Z_i;\gamma) > u_i) \tag{9}$$

The single index restriction $E(D \mid Z) = E(D \mid Z^{\mathrm{T}}\gamma)$ permits multiplicative heteroscadasticity of a general but known form, and heteroscadasticity of unknown form only if it depends entirely on the assumed index $\theta(Z_i;\gamma)$, for example $\theta(Z_i;\gamma) = Z_i^{\mathrm{T}}\gamma$. Klein and Spady's estimation algorithm is motivated by the parametric ML approach where the coefficients γ are estimated by maximizing the likelihood

$$\begin{aligned}\underset{\gamma}{Max}\, L(\gamma) &= \sum_{i=1}^{N} D_i \cdot \ln \Pr ob(D_i = 1 \mid \theta_i) + (1 - D_i) \cdot \ln(1 - \Pr ob(D_i = 1 \mid \theta_i)) \\ &= \sum_{i=1}^{N} D_i \cdot \ln \Pr ob(Z_i^{\mathrm{T}}\gamma > u_i \mid Z_i^{\mathrm{T}}\gamma) + (1 - D_i) \cdot \ln(1 - \Pr ob(Z_i^{\mathrm{T}}\gamma > u_i \mid Z_i^{\mathrm{T}}\gamma))\end{aligned} \tag{10}$$

Unlike the parametric approach, the density of the error term u_i is unspecified. A quasi likelihood function with an estimated error term density is proposed instead. That is, a smooth function of γ, $\theta_n(\bullet)$ [3], obtained with iteration using a nonparametric kernel estimator of the density of $Z_i^{\mathrm{T}}\gamma$, approximates the parametric likelihood. Maximizing this quasi likelihood yields $\hat{\gamma}_{Klein\&Spady}$ via

$$\underset{\gamma}{Max}\, QuasiL(\gamma) = \sum_{i=1}^{N} w(Z_i)\left\{D_i \cdot \ln[\Pr\hat{o}b(D_i = 1 \mid Z_i)]^2 + (1-D_i)\cdot \ln[1-\Pr\hat{o}b(D_i = 1 \mid Z_i)]^2\right\} \tag{11}$$

(See more details in Appendix B).

2.2.1.2 Step 2

In the second step, the wage equation $Y_i = \beta_0 + X_i^{\mathrm{T}}\beta_1 + \lambda(Z_i^{\mathrm{T}}\gamma) + \xi_i$ is estimated for the uncensored subsample using $\hat{\gamma}_{Klein\&Spady}$ from the first step. That is,

$$Y_i = \beta_0 + X_i^{\mathrm{T}}\beta + \lambda(Z_i^{\mathrm{T}}\hat{\gamma}_{Klein\&Spady}) + \xi_i, \tag{12}$$

where $\lambda(Z_i^{\mathrm{T}}\hat{\gamma}_{Klein\&Spady})$ is an unknown selectivity term. Newey (1999) approximates $\lambda(\bullet)$ with a series based on orthogonal polynomials, in which the number of terms increases with the sample size. With the series added, equation (12) can be written as:

$$Y_i = X_i^{\mathrm{T}}\beta + \sum_{k=1}^{K}\omega_k \cdot \tau_k(Z_i^{\mathrm{T}}\hat{\gamma}_{Klein\&Spady}) + \tilde{\xi}_i \text{ with } \tilde{\xi}_i = \sum_{k=K+1}^{\infty}\omega_k \cdot \tau_k(Z_i^{\mathrm{T}}\hat{\gamma}_{Klein\&Spady}) + \xi_i, \tag{13}$$

where ω_k are unknown coefficients and $\tau_k(\bullet)$ are known smooth basis functions depending on the index. The number of basis functions, k, serves a role similar to the

[3] Klein and Spady recommend using either a higher order Kernel or an adaptive locally smoothed Kernel to obtain $\theta_n(\bullet)$.

bandwidth h in kernel estimation. The parameter vectors β_1 and ω can then be estimated by OLS, for given k. In this paper, I choose k using method of Vella (1998), which chooses the number of approximation terms based on the t-statistics of the additional terms.

2.2.2 Robinson (1988)

Robinson (1988) estimates the selectivity model without any index restriction on the error terms. Instead, he assumes strong independence of the errors and regressors. Robinson subtracts the statistical expectations of the structural equation (12) from the observed values for each individual to eliminate the selection term $\lambda(Z_i, \gamma)$. That is,

$$(Y_i - E[Y_i \mid Z_i]) = (X_i - E[X_i \mid Z_i])^{\mathrm{T}} \beta + \xi_i - E(\xi_i \mid Z_i) \;, \tag{14}$$

where $E[Y_i \mid Z_i] = E[X_i \mid Z_i]^{\mathrm{T}} \beta + E[\lambda(Z_i,\gamma) \mid Z_i] + E[\xi_i \mid Z_i] = E[X_i \mid Z_i]^{\mathrm{T}} \beta + \lambda(Z_i,\gamma)$.

Note that $E(\xi_i \mid X_i, Z_i) = 0$. The law of iterated expectations implies $E[\xi_i - E(\xi_i \mid Z_i)] = 0$. Inserting nonparametric Nadaraya-Watson estimates of $E[Y_i \mid Z_i]$ and $E[X_i \mid Z_i]$ in equation (14), OLS on the censored sample yields $\sqrt{n}$ consistent and asymptotically normal estimates of β :

$$\hat{\beta}_{Robinson} = \\ [\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{E}[X_i \mid Z_i]) \cdot (X_i - \hat{E}[X_i \mid Z_i)]^{\mathrm{T}}]^{-1} \cdot [\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{E}[X_i \mid Z_i]) \cdot (Y_i - \hat{E}[Y_i \mid Z_i)]^{\mathrm{T}}] \tag{15}$$

Note that a multivariate Kernel $K(\frac{Z_i - Z_j}{h})$ with bandwidth h that weighs the distance of the two vectors Z_i and Z_j can be invoked to estimate $E[X_i \mid Z_i]$ and $E[Y_i \mid Z_i]$.

Also note that there is no theory on how to choose the bandwidth. Due to Robinson's strong independence assumption, his estimation mechanism does not allow for heteroscedastic disturbance terms conditional on X_i, Z_i. However, Hardle and Linton (page 2330~2331, Ch 38 (1994)), Schafgans (1998) and Yatchew (Ch 3 and 4 (2003)) proposed varies ways to fix the nonspherical error problem, which include for instance a weighted least squares approach.

CHAPTER 3

DATA SET AND EMPIRICAL RESULTS

3.1 Data and Variables

The analysis uses data from the Canadian Survey of Labour and Income Dynamics for the year 1994 (94 SLID). The original sample consists of 29632 observations. Individuals excluded from the sample are full-time students (Indicated by FLLPR20C in 94 SLID), individuals less than 25 years of age or above 59, and those with relevant information missing. The resulting samples involve 7495 females, 5039 of whom are paid employees, and 6963 males, 5288 of whom are paid employees. Detailed variables and definition as well as Means of the variables and standard deviations of the means are summarized in Appendix C.

Note that, from the descriptive statistics, we observe that the average wage rate for female workers is lower than that of male workers. The former with larger standard deviations of the means is about $ 10.9 CND/hour, and the latter is about $ 12.1 CND/hour. The average years of education of female workers are higher than that of male workers. And the average years of education of paid-workers are higher than that of non-paid workers for both males and females.

3.2 Empirical Results for Sample Selection Corrected Estimation

Table 1 compares the results of parametric probit and semiparametric probit (Klein and Spady (1993)). The dependent variable in the work participation equation is a dichotomous indicator for wageworker. The identifying independent variables in the paid work participation equation are non-labour income[4] and household size, since they only affect people's participation decision rather than the offered wage. The rest of the control variables include marital status, number of young children, number of old children, age, age squared, years of schooling, a dummy variable indicating individual's long-term disability condition, mother tongue, group status and region of residence. The optimal smoothness parameter h_n is chosen via generalized cross validation. The Hausman tests reject [5] the probit specification for both males and females, indicating that the parametric estimator may not be consistent. Bera, Jarque and Lee tests reject the normality of residuals. For identification, the coefficient of age in semiparametric probit is normalized to 1. For comparison, the coefficient of age in parametric probit is also normalized to 1. For males, probit model produces a wrong sign on the estimated age coefficient. As expected, nonlabour income and number of children have a statistically significant negative impact on labour force participation for both males and females. Women with large household size devote less time to work. It is opposite for men, the breadwinner.

[4] NLINCOME has influence on one's reservation wage but does not affect one's offered wage.

[5] According to Schafgans (2000), the null is rejected when the test statistic is negative.

Table 1 Parametric and Semiparametric results for labour force participation equation (Females and Males)

Method	Female		Male	
Variables	*Probit*	*K & S* $h_n = 0.26$	*Probit*	*K & S* $h_n = 0.27$
Dependent Variable	WAGE WORKER	WAGE WORKER	WAGE WORKER	WAGE WORKER
Constant	-1.451 ***(0.717)	---	8.680 (20.346)	—
NLINCOME	-0.040***(0.010)	-0.038*** (0.010)	-2.144*** (6.459)	-1.315*** (6.107)
HHSIZE	-0.089***(0.030)	-0.111*** (0.029)	0.458 (1.447)	1.358** (6.314)
MARRIED	-0.053 (0.050)	-0.061 (0.044)	5.250***(15.751)	1.667 (7.860)
N0004C	-0.493***(0.126)	-0.432***(0.108)	-2.542*** (7.709)	-2.745* (12.875)
N0519C	-0.083* (0.036)	-0.054* (0.029)	-0.547 (1.891)	-0.925 (4.158)
AGE10	1	1	-1	1
AGE10SQ	-0.157***(0.009)	-0.154*** (0.007)	-0.325 (1.334)	-0.309 (0.898)
EDU	0.110***(0.024)	0.112*** (0.024)	0.902*** (2.713)	0.386*** (1.810)
DISABLED	-0.998***(0.215)	-1.137*** (0.227)	-12.196***(36.725)	-9.158*** (42.565)
ENGLISH	0.082 (0.074)	0.083 (0.059)	4.055***(12.205)	2.620*** (12.301)
FRENCH	0.213* (0.105)	-0.246*** (0.080)	6.808***(20.485)	3.869* (18.180)
ABORIGINAL	-0.247* (0.125)	-0.328*** (0.101)	-0.138 (1.881)	-0.566* (3.898)
VISIBLEM	-0.026 (0.117)	0.128 (0.095)	0.678 (2.651)	1.172 (6.077)
ATLANTIC	-0.178*** (0.065)	-0.167*** (0.057)	-1.015 (3.183)	-1.721 (8.116)
QUEBEC	-0.390*** (0.113)	-0.363*** (0.094)	-2.686* (8.198)	-2.688 (12.627)
PRAIRIES	0.036 (0.055)	0.123** (0.051)	-1.979* (6.020)	-1.543 (7.257)
BC	-0.013 (0.072)	0.111 (0.070)	-0.217 (1.298)	-0.617 (3.318)
Log L	-1321.184	-4178.130	-3352.445	-3714.981
Hausman Test		[-17.511]		[30.144]
Bera, Jarque normality test	[812.377] { 0.000 } df=2	---	[1283.345] { 0.000 } df=2	---
Sample Size	7495	7495	6963	6963

Notes:

(1) K&S refers to the semiparametric estimates of Klein and Spady (1993).

(2) The value for bandwidth h_n satisfies (i) $n^{-1/6} < h_n < n^{-1/8}$ and (ii) bias reducing condition for the kernel $\int Z^2 K(Z) dZ = 0$.

(3) The Hausman (1978) test, tests for equality of the parametric and Klein-Spady semiparametric estimates (a test for normal disturbances of the binary response of sample selection equation). Under the null hypothesis that the probit estimator is consistent and efficient but inconsistent under the alternative, but the Klein-Spady estimator is consistent under both the null and the alternative but inefficient under the null. The test statistic is $(b_{k\&s} - b_p)^T INV(V_{k\&s} - V_p)(b_{k\&s} - b_p) \sim Chi^2(df)$, where $b_{k\&s}$ and b_p are the Klein-Spady and Probit estimates respectively, $V_{k\&s}$ and V_p are their (asymptotic) covariance matrices, and df is the rank of the matrix $(V_{k\&s} - V_p)$. The cut-off value is $\chi^2_{5\%,16} = 26.30$.

(4) The numbers in parenthesis are standard errors, the numbers in square brackets are test statistics and the numbers in large brackets are p-values.

(5) Significance: ***: 0.01, **: 0.05, *:0.1.

To correct for endogeneity of experience, I adopted methods proposed in Buchinsky (1998). His model assumes that the main alternative use for women's time is child rearing (and the home activities related to this task). Then instead of considering Mincer's human capital model, I adopt the following specification:

$$LogY_i = X_i^T \beta + r_{edu} S_i + \alpha * PEXP_i + \delta * PEXP_i^2 + \pi * PEXP * ToTCHD + \tau * PEXPSQ * ToTCHD + u_i, \tag{16}$$

where Y_i is potential hourly wage, $PEXP$ is potential experience, $PEXPSQ$ is potential experience squared, $ToTCHD$ is total number of children, $(\beta\ , r_{edu}, \alpha, \delta, \pi, \tau)$ are unknown parameter vectors and u_i is an error term.

Table 2 gives parametric and semiparametric estimations of wage equations using various corrections for sample selection bias for females. Inverse Mill's ratio and correlation errors in Heckit and MLE model respectively indicate that there is a selection bias. The selectivity term in Newey's model also indicates a selection bias. A pattern of diminishing return to experience is found. Disabled people earn less on average. English or French speaking workers earn more than others, which is shown in the semiparametric models. The highest-paid females reside in British Columbia. For females, the estimated return to schooling is 4.9% in Robinson's model and 4.4% in Newey's model. Parametric models underestimate the return.

Table 2: Parametric and Semiparametric estimations of wage equations using various corrections for sample selection bias for females

Method	Female				
	Parametric			Semi-parametric	
	OLS[1]	*Heckit*[2]	*MLE*	*Robinson*	*Newey*
Dependent Variable	LWAGE	LWAGE	LWAGE	LWAGE	LWAGE
Constant	0.470*** (0.022)	0.379*** (0.034)	0.433*** (0.028)	---	---
REXP10	0.106*** (0.012)	0.128*** (0.014)	0.115*** (0.012)	0.224*** (0.024)	0.128*** (0.013)
REXP10SQ	-0.016*** (0.003)	-0.024*** (0.003)	-0.019***(0.003)	-0.057***(0.005)	-0.011*** (0.003)
PEXP10TC	-0.002 (0.005)	-0.016***(0.006)	-0.007 (0.005)	-0.071***(0.010)	0.028*** (0.006)
PEXP10SQTC	0.000 (0.002)	0.004* (0.002)	0.002 (0.002)	0.019***(0.004)	-0.008*** (0.002)
EDU	0.040*** (0.001)	0.042*** (0.001)	0.041*** (0.001)	0.049*** (0.002)	0.044*** (0.001)
DISABLED	-0.029*** (0.012)	-0.074***(0.019)	-0.050***(0.015)	-0.282***(0.021)	-0.113*** (0.017)
ENGLISH	0.001 (0.010)	0.007 (0.010)	0.004 (0.010)	0.035 (0.021)	0.043*** (0.010)
FRENCH	-0.011 (0.013)	0.001 (0.014)	-0.006 (0.014)	0.064** (0.028)	0.013 (0.014)
ABORIGINAL	-0.015 (0.017)	-0.024 (0.018)	-0.020 (0.017)	-0.046 (0.034)	0.050 (0.018)
VISIBLEM	-0.019 (0.015)	-0.021 (0.016)	-0.020 (0.016)	-0.005 (0.034)	0.007 (0.016)
ATLANTIC	-0.090*** (0.008)	-0.096***(0.008)	-0.092***(0.008)	-0.103*** (0.016)	-0.058*** (0.008)
QUEBEC	-0.019*** (0.011)	-0.032** (0.013)	-0.026** (0.012)	-0.074*** (0.024)	0.039*** (0.012)
PRAIRIES	-0.044*** (0.007)	-0.042***(0.007)	-0.043***(0.007)	-0.141*** (0.016)	-0.042*** (0.007)
BC	0.020*** (0.009)	0.022** (0.010)	0.020** (0.010)	0.050** (0.021)	0.041*** (0.010)
MARRIED	0.023*** (0.006)	0.014** (0.007)	0.18*** (0.003)	0.292*** (0.011)	0.063*** (0.006)
Inverse Mill's Ratio		0.096*** (0.030)			
$(z_i^T \gamma)$					0.120*** (0.010)
Correlation Errors, RHO(1,2)			0.238** (0.107)		
Std error of regression SIGMA (1)	0.179	0.189	0.187 (0.124)	0.544	0.184
Adjusted R-squared	0.289	0.292		0.085	0.248
R-squared	0.290	0.294		0.088	0.250
Sample Size	5039	5039	5039	5039	5039
Sum of squared residuals	160.156	158.787			

Notes:

(1) *OLS*[1] : OLS estimation with no selectivity but with robust covariance matrix for heteroscadasticity.

(2) *Heckit*[2] : Heckman Two-Step Estimation Procedure with Greene Corrected SE.

(3) The numbers in parenthesis are standard errors.

(4) Significance: ***: 0.01, **: 0.05, *:0.1.

Table 3: Parametric and Semiparametric estimations of wage equations using various corrections for sample selection bias for males

Method	Male				
	Parametric			Semi-parametric	
	OLS^1	$Heckit^2$	*MLE*	*Robinson*	*Newey*
Dependent Variable	LWAGE	LWAGE	LWAGE	LWAGE	LWAGE
Constant	0.685*** (0.019)	0.681*** (0.020)	0.682*** (0.019)	---	---
REXP10	0.106*** (0.011)	0.103*** (0.010)	0.103*** (0.010)	0.098*** (0.022)	0.137***(0.012)
REXP10SQ	-0.010*** (0.002)	-0.011*** (0.002)	-0.011*** (0.002)	-0.024*** (0.005)	-0.002 (0.003)
PEXP10TC	0.030*** (0.004)	0.018*** (0.004)	0.018*** (0.004)	0.002 (0.009)	0.065*** (0.005)
PEXP10SQTC	-0.008*** (0.002)	-0.005*** (0.002)	-0.005*** (0.001)	0.003 (0.003)	-0.017*** (0.002)
EDU	0.026*** (0.001)	0.025*** (0.001)	0.025*** (0.001)	0.027*** (0.002)	0.018*** (0.001)
DISABLED	-0.068** (0.011)	-0.067*** (0.011)	-0.066*** (0.010)	-0.324*** (0.019)	0.177*** (0.014)
ENGLISH	0.033*** (0.010)	0.035*** (0.010)	0.035*** (0.009)	0.082*** (0.021)	0.113*** (0.010)
FRENCH	0.017 (0.013)	0.020 (0.012)	0.019* (0.011)	0.099*** (0.026)	0.057*** (0.013)
ABORIGINAL	-0.007 (0.016)	-0.004 (0.016)	-0.004 (0.017)	-0.049 (0.034)	0.085*** (0.017)
VISIBLEM	-0.075*** (0.014)	-0.075*** (0.014)	-0.076*** (0.015)	-0.077** (0.031)	-0.029* (0.015)
ATLANTIC	-0.076*** (0.007)	-0.078*** (0.007)	-0.078*** (0.007)	-0.092*** (0.015)	-0.026*** (0.007)
QUEBEC	-0.017 (0.011)	-0.011 (0.010)	-0.011 (0.009)	-0.044** (0.022)	0.087*** (0.011)
PRAIRIES	-0.032*** (0.007)	-0.032*** (0.007)	-0.032*** (0.007)	-0.145*** (0.015)	-0.040*** (0.007)
BC	0.037*** (0.009)	0.039*** (0.009)	0.039*** (0.010)	0.048** (0.020)	0.030*** (0.010)
MARRIED	0.050*** (0.006)	0.050*** (0.006)	0.050*** (0.006)	0.216*** (0.010)	0.093*** (0.007)
Inverse Mill's Ratio		0.007 (0.016)			
$(z_i^T \gamma)$					0.210*** (0.009)
Correlation Errors, RHO(1,2)			0.028 (0.079)		
Std error of regression SIGMA (1)	0.167	0.167	0.167*** (0.001)	0.517	0.179
Adjusted R-squared	0.250	0.250		0.068	0.145
R-squared	0.252	0.252		0.070	0.148
Sample Size	5288	5288	5039	5288	5288
Sum of squared residuals	147.436	146.956			

Notes:

(1) OLS^1 : OLS estimation with no selectivity but with robust covariance matrix for heteroscadasticity.

(2) $Heckit^2$: Heckman Two-Step Estimation Procedure with Greene Corrected SE.

(3) The numbers in parenthesis are standard errors.

(4) Significance: ***: 0.01, **: 0.05, *:0.1.

For males, results of parametric and semiparametric estimations of wage equations using various corrections for sample selection bias are shown in table 3. The highest-paid males also reside in British Columbia. Both Robinson (1988) and Newey (1999) produce significant and positive estimates for mother tongue, indicating language skills are key determinants of wages for males in Canada. Inverse Mill's ratio and correlation errors in Heckit and MLE model respectively indicate that there is no selection bias. The selectivity term in Newey's model, however, indicates a selection bias. Newey's model produces a wrong sign[6] on the estimated coefficient of long-term disability and ethnicity status (aboriginal). The estimated return to schooling is 2.7% and 1.8% in Robinson (1988) and Newey (1999)'s model respectively.

Table 4: Hausman specification test (Distributed as Chi-squared) for females

Difference	Statistic	Degrees of freedom	5% Critical values	Interpretation
OLS Versus *Robinson*	701.181	15	25.00	Reject null of no selectivity bias
Heckit Versus *Newey*	-287.311	15	25.00	Reject parametric normality assumption
MLE Versus *Newey*	46093.97	15	25.00	Reject parametric normality assumption

To further confirm the existence of selectivity bias, in table 4 and 5, we compare OLS and Robinson model by Hausman specification test. The test statistic is 701.181 and 1303.899 for females and males respectively, which are greater than the critical value at 5% level.

[6] For males, schooling is proved endogenous in table 6. So it might be the source of the wrong sign. It also casts some doubts on the reliability of the results of the semiparametric model.

So the test rejects the null hypothesis that OLS model is correctly specified. This indicates strong evidence for the presence of a selectivity bias. MLE and Heckit model are also compared against Newey's model via Hausman test. It further confirms our previous finding in table 1 that normality is rejected at 5% level.

Table 5 : Hausman specification test (Distributed as Chi-squared) for males

Difference	Statistic	Degrees of freedom	5% Critical values	Interpretation
OLS Versus *Robinson*	1303.899	15	25.00	Reject null of no selectivity bias
Heckit Versus *Newey*	-803.570	15	25.00	Reject parametric normality assumption
MLE Versus *Newey*	-736.862	15	25.00	Reject parametric normality assumption

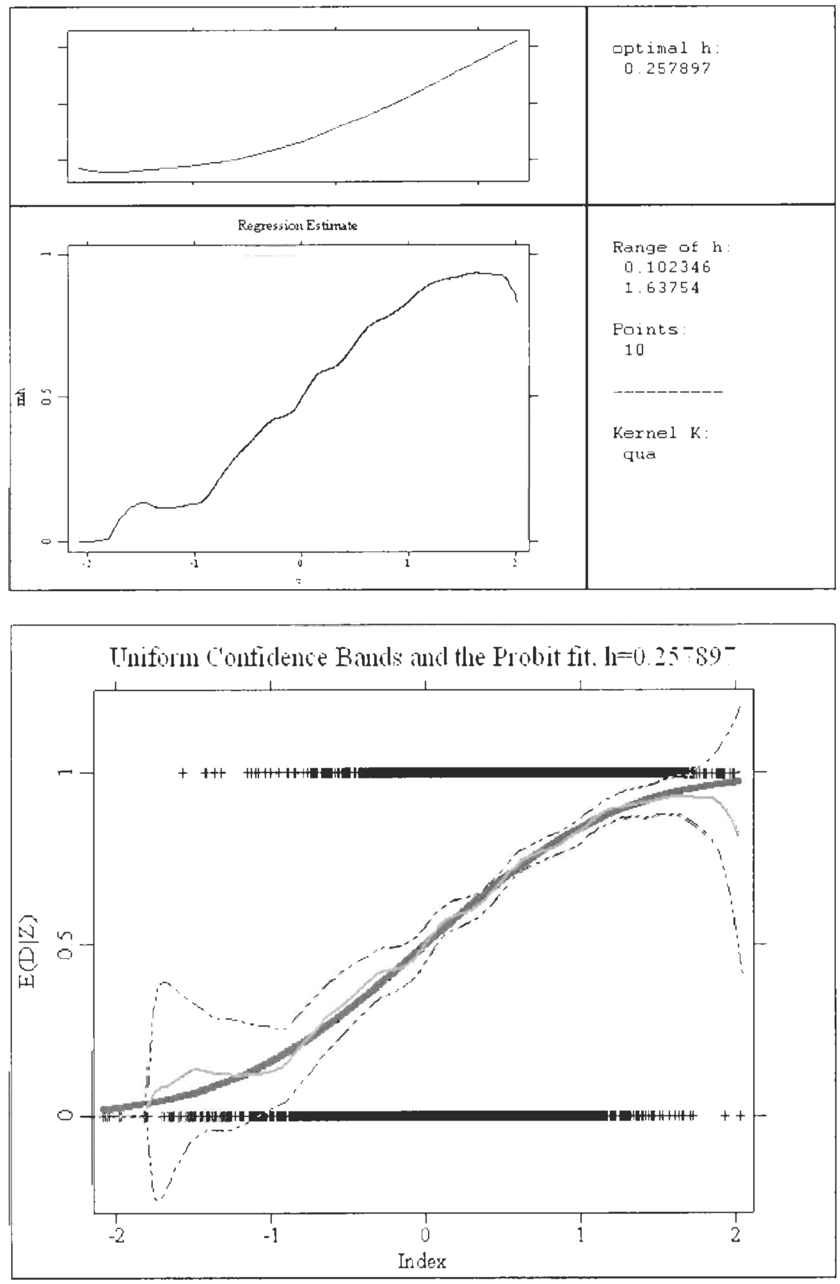
3.3 Specification Checks

3.3.1 Horowitz (1993)

To determine whether the distributional assumptions made by the probit model are consistent with the data, the specification tests proposed by Horowitz (1993) is used. These tests are based on the difference between the probit link, Φ, and the nonparametric regression curve for the general model, $E(D \mid Z) = F(Z^T \gamma)$,where F is an unknown function. Horowitz's (1993) test is based on a non-parametric regression of D on $Z^T \hat{\gamma}$, call it $\tilde{F}(Z^T \hat{\gamma})$, and a uniform confidence band for the regression function (where $\hat{\gamma}$ is the parametric probit maximum likelihood estimate of γ). The probit model is rejected if the standard normal distribution function does not lie within this band. The test is explained in more details in Martins (2001). The test for females is displayed in figure 1, which includes a plot of the quartic kernel regression of D on $Z^T \hat{\gamma}$, the uniform 95% confidence band for the regression, and $\Phi(Z^T \hat{\gamma})$, for $h_{Optimal} = 0.257897$ (results for other values are shown in Appendix D). Like the finding in Martins (2001), the results depend on the choice of the particular bandwidth [7]. The probit fit seems to be within the uniform confidence band only for $0.4 < h < 0.6$. The optimal binwidth ($h_{Optimal} = 0.257897$) is chosen by Generalized Cross-validation (See Appendix E). The graph also suggests that the greatest departure from the probit model seems to occur in the range of the low index value people. At the optimal binwidth, the probit link function is rejected. The test for males is displayed in figure 2. The probit fit seems to be within the uniform confidence band only for $0.6 < h < 0.8$. The optimal binwidth ($h_{Optimal} = 0.269368$) is chosen by Generalized Cross-validation (results for other values are shown in Appendix D). At the optimal binwidth, the probit link function is rejected again. Notice that for males the kernel regression curve is non-monotone in the range of low index. We need to trim the small estimated density that maybe disturbs the outcomes, because the non-monotone occurs when a small bandwidth is chosen. All in all, at the optimal binwidth the results support a semiparametric specification of the sample selection equation.

[7] Too small bandwidths result in inaccurate estimates whereas too large bandwidths give estimates that are too small to reveal structural features.

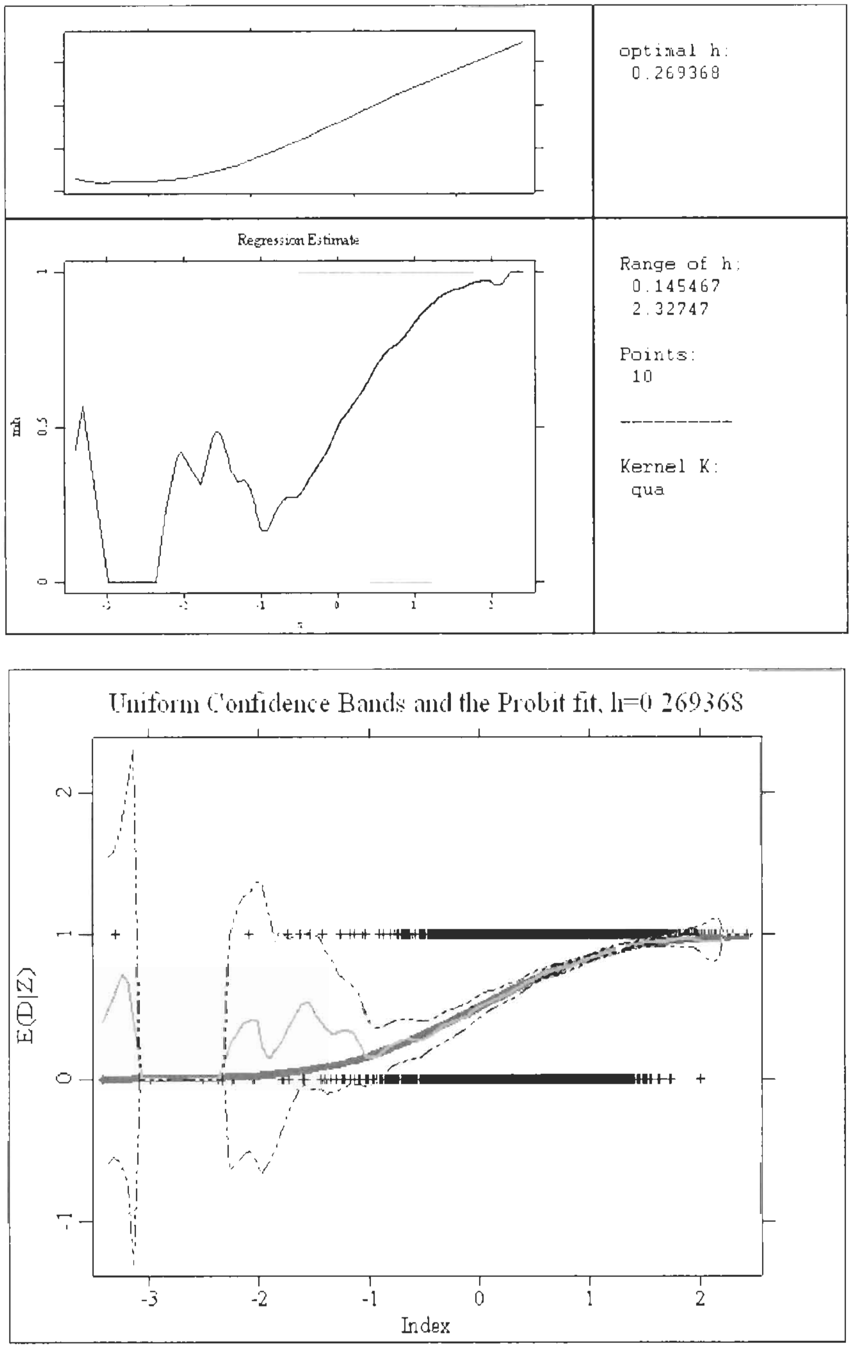
Figure 1: GCV and Horowitz (1993) test for females



The graphical displays on the top show the cross-validation criterion in the upper left, the chosen optimal bandwidth in the upper right, the resulting kernel regression in the lower left, and information about the search grid and the kernel in the lower right.

Uniform confidence bands and Probit fit for the optimal bandwidth are shown at the bottom. Dashed thin line: 95% confidence bands, thick solid line: Cumulative Normal distribution, solid line: Kernel regression estimates. The Probit specification cannot be rejected when Cumulative Normal distribution function lies within the confidence band.

Figure 2: GCV and Horowitz (1993) test for males



The graphical displays on the top show the cross-validation criterion in the upper left, the chosen optimal bandwidth in the upper right, the resulting kernel regression in the lower left, and information about the search grid and the kernel in the lower right.

Uniform confidence bands and Probit fit for the optimal bandwidth are shown at the bottom. Dashed thin line: 95% confidence bands, thick solid line: Cumulative Normal distribution, solid line: Kernel regression estimates. The Probit specification cannot be rejected when Cumulative Normal distribution function lies within the confidence band.

3.3.2 Measurement Error and Endogeneity of Schooling

3.3.2.1 Measurement error

Assume that the independent variable S_i in equation (16) is measured with error. For each observation, S_i equals the true value for $TRUE\_S_i$ plus classical measurement error ω_i. That is, $S_i = TRUE\_S_i + \omega_i$. Substituting it into (16), we find that the OLS estimator of the intercept and slope coefficient is inconsistent, since S_i depends on ω_i and so does ε_i. It can be shown that

$P\lim \hat{r}_{edu} = r_{edu} + \dfrac{E(S_i, \varepsilon_i - r_{edu} \cdot \omega_i)}{Var(S_i)} = r_{edu} \cdot (1 - \dfrac{\sigma_\omega^2}{\sigma_{TRUE\_S}^2 + \sigma_\omega^2})$. Thus, $\hat{r}_{edu}$ is consistent only if there is no measurement error. It is asymptotically biased towards zero if σ_ω^2, the variance of measurement errors, is positive.

3.3.2.2 Endogenous schooling: unobserved ability bias

If individuals are heterogeneous, unobserved ability that affects schooling choice and wage outcomes means schooling is endogenous. To see this, rewrite equation (16) as follows:

$$LogY_i = r_{edu} S_i + W_i^T \beta + \varepsilon_i \tag{17}$$

$$S_i = M_i^T \pi + \rho_i, \tag{18}$$

where W_i^T is a vector of observed covariates including people's experience, experience squared, a constant and so forth, M_i^T is a vector of exogenous factors that influence the schooling decision and ρ_i is the heterogeneity term. In the wage equation (17),

ε_i includes all unobservables that affect a person's wage, including things like ability. The coefficient r_{edu} represents the true causal effect of education, which can be interpreted as the expected increase in average earnings if one more year of education was assigned to a random sample of the population. However, the estimated coefficient on r_{edu} is inconsistent if ε_i and ρ_i are correlated. This correlation arises if, for example, individuals who earn higher wages also acquire more schooling.

So $p\lim \hat{r}_{edu} - r_{edu} = \dfrac{Cov(S_i, \varepsilon_i)}{Var(S_i)} \neq 0$, which means the estimated schooling coefficient is biased upward relative to the average return to education in the population if ability is ignored assuming $Cov(S_i, \varepsilon_i) > 0$. Card (1999) argued that if the individual specific returns to schooling are higher for individuals with low levels of schooling, the unobserved component will be negatively correlated with schooling. In this case, there would be a downward bias. To see this, rewrite equation (17) as the following

$$LogY_i = r_{edu,i}S_i + W_i^T \beta + \varepsilon_i = r_{edu}S_i + W_i^T \beta + \tau_i, \tag{19}$$

where $\tau_i = \varepsilon_i + (r_{edu,i} - r_{edu}) \cdot S_i$ and $(r_{edu,i} - r_{edu}) \cdot S_i$ is the unobserved discrepancy component. Deschenes (2002) modelled schooling and earnings with heterogeneous returns to education. His model allows individual heterogeneity in earnings capacity to affect both the intercept and the slope of the earnings function. Card (2001) further illustrated that individual heterogeneity in the optimal schooling choice can arise from differences in economic benefits of schooling and differences in the economic costs of schooling; therefore, the OLS estimate for r_{edu} would reflect the difference in the expected wage of two arbitrary people with the same observed characteristics but having say $S-1$ and S years of education (a ceteris paribus interpretation). It does not measure the expected wage difference if an arbitrary person decides to increase his/her years of schooling from $S-1$ to S (a causal interpretation).

3.3.2.3 The instrumental variable solution

A common solution to the problems of ability bias and measurement error bias is the method of instrumental variables [8]. In this paper, family background variables such as mother's years of schooling and father's years of schooling are used as instruments [9] for schooling. This is assuming that they affect the costs/choice of schooling rather than wage earnings directly in the wage equation and are not a linear combination of the other variables in the model. Other instruments have been proposed recently; for instance, individual's quarter of birth in Angrist and Krueger (1991), tuition at 2 and 4-year state colleges in Kane and Rouse (1993), sibling composition with respect to gender in Butcher and Case (1994) distance to nearest high school and indicator for local private high school in Maluccio (1997), potential eligibility for the Veteran's Rehabilitation Act benefits dummy for Ontario men age 19-22 in 1946 Lemieux and Card (2001). However, it should be noted that a convincing or desirable instrument for education is very hard to be found as it can be easily correlated with the omitted ability.

Table 6 presents the estimation results of the wage equations and the reduced form equation by gender. The column (1) and (4) give the OLS results and the column (3) and (6) give the 2SLS results. An exogeneity test by Davidson and MacKinnon (1993, Ch7)

[8] I will stick to IV estimation in this paper, although GMM is a viable alternative and the former might suffer from weak IV problems with large standard errors.

[9] Parental education is the element in M_i^T but not in W_i^T (See equation 17 and 18).

on education is performed to make sure the unbiasedness of OLS estimates. 1st step: regress children's education on all exogenous variables and instruments (parental education), and retrieve the residuals. 2nd step: estimate the wage equation including the residuals as additional regressors. OLS estimates are consistent if the coefficient on the 1st stage residuals in not significant. The estimated coefficient on the residual in the augmented OLS earnings equation is negative (-0.011 and -0.016 for females and males respectively) and statistically significant indicating that schooling is potentially endogenous. Alternatively, one may use a Durbin-Wu-Hausman (1978) test [10]. Under the null hypothesis, schooling is exogenous and OLS estimates are unbiased and efficient. Whereas, IV estimates are consistent under both null and alternative hypothesis. The idea of the DWH test is to check whether the difference $\hat{\beta}_{IV} - \hat{\beta}_{OLS}$ is significantly different from zero in the available sample. The DWH test indicates that schooling is exogenous for females since the test statistic fails to reject the null hypothesis of exogeneity of schooling (5.896<23.69). The DWH test suggests that schooling is endogenous for males since the test statistic rejects the null hypothesis of exogeneity of schooling (38.103>23.69).

[10] Durbin-Wu-Hausman (1978) test is more powerful than the first exogeneity test, since it only tests against a single alternative, endogeneity of schooling. However, Davidson and MacKinnon (1993, Ch7) tests all possible endogenous regressors that might pollute our results.

Table 6: Parental education as instruments for females and males

Method	Female			Male		
Variables	*OLS log earnings (1)*	*Reduced-form schooling (2)*	*IV log earnings (3)*	*OLS log earnings (4)*	*Reduced-form schooling (5)*	*IV log earnings (6)*
Dependent Variable	LWAGE	EDU	LWAGE	LWAGE	EDU	LWAGE
Constant	0.470*** (0.022)	15.503*** (0.225)	0.307*** (0.054)	0.685*** (0.019)	15.534*** (0.246)	0.467*** (0.045)
PEXP10	0.106*** (0.012)	0.019 (0.175)	0.109*** (0.013)	0.106*** (0.011)	-0.123 (0.181)	0.112*** (0.011)
PEXP10SQ	-0.016*** (0.003)	-0.257*** (0.039)	-0.014*** (0.003)	-0.010***(0.002)	-0.252*** (0.039)	-0.008*** (0.003)
PEXP10TC	-0.002 (0.005)	-0.339*** (0.070)	-0.003 (0.006)	0.030*** (0.004)	-0.204*** (0.071)	0.022*** (0.005)
PEXP10SQTC	0.000 (0.002)	0.056** (0.028)	0.001 (0.002)	-0.008*** (0.002)	0.025 (0.024)	-0.006*** (0.002)
EDU	0.040*** (0.001)	---	0.049*** (0.003)	0.026*** (0.001)	---	0.039*** (0.003)
DISABLED	-0.029** (0.012)	0.003 (0.178)	-0.027** (0.012)	-0.068** (0.011)	-0.282* (0.156)	-0.060*** (0.011)
ENGLISH	0.001 (0.010)	0.014 (0.143)	0.000 (0.010)	0.033*** (0.010)	-0.529*** (0.172)	0.036*** (0.010)
FRENCH	-0.011 (0.013)	-0.041 (0.186)	-0.008 (0.013)	0.017 (0.013)	-0.356* (0.212)	0.023* (0.014)
ABORIGINAL	-0.015 (0.017)	-0.596*** (0.224)	-0.004 (0.017)	-0.007 (0.016)	-0.942***(0.241)	0.014 (0.016)
VISIBLEM	-0.019* (0.015)	0.059 (0.246)	-0.019 (0.015)	-0.075*** (0.014)	0.397 (0.260)	-0.083*** (0.014)
ATLANTIC	-0.090*** (0.008)	-0.666*** (0.095)	-0.082*** (0.008)	-0.076*** (0.007)	-0.705*** (0.108)	-0.065*** (0.007)
QUEBEC	-0.019* (0.011)	-0.035 (0.155)	-0.016 (0.011)	-0.017 (0.011)	-0.277* (0.166)	-0.005 (0.011)
PRAIRIES	-0.044*** (0.007)	-0.492*** (0.090)	-0.038*** (0.007)	-0.032*** (0.007)	-0.418*** (0.104)	-0.026*** (0.007)
BC	0.020** (0.009)	-0.219* (0.124)	0.022** (0.009)	0.037*** (0.009)	-0.359***(0.137)	0.040*** (0.009)
MARRIED	0.023*** (0.006)	0.238*** (0.078)	0.020*** (0.006)	0.050*** (0.006)	1.083*** (0.094)	0.033*** (0.007)
EDFATH1	---	-0.710*** (0.109)	---	---	-1.039*** (0.128)	---
EDFATH2	---	-0.293*** (0.110)	---	---	-0.519*** (0.127)	---
EDFATH4	---	0.377** (0.153)	---	---	0.192 (0.169)	---
EDFATH5	---	1.250*** (0.163)	---	---	1.223*** (0.179)	---
EDMOTH1	---	-0.620*** (0.108)	---	---	-0.608*** (0.125)	---
EDMOTH2	---	-0.379*** (0.102)	---	---	-0.131 (0.118)	---
EDMOTH4	---	0.383*** (0.127)	---	---	0.616*** (0.138)	---
EDMOTH5	---	0.810*** (0.194)	---	---	0.972*** (0.230)	---

Method	Female			Male		
Variables	*OLS log earnings (1)*	*Reduced-form schooling (2)*	*IV log earnings (3)*	*OLS log earnings (4)*	*Reduced-form schooling (5)*	*IV log earnings (6)*
Sample size	5039	5039	5039	5288	5288	5288
Adjusted R-squared	0.291	0.302	0.277	0.250	0.320	0.207
Exogeneity test on EDU RESID	-0.011*** (0.004)	---	---	-0.016*** (0.003)	---	---
Test for Exclusion restriction	---	[63.224] {0.000} $F_{5\%, 8, 5016} = 1.94$	---	---	[70.541] {0.000} $F_{5\%, 8, 5286} = 1.94$	---
Sargan's test	---	---	[1.708] $\chi^2_{5\%,7} = 14.067$	---	---	[10.676] $\chi^2_{5\%,7} = 14.067$
Durbin-Wu Hausman test	---	---	[5.896] $\chi^2_{5\%,14} = 23.69$	---	---	[38.103] $\chi^2_{5\%,14} = 23.69$

Notes:

(1) White's Heteroscedasticity corrected standard errors are used.

(2) The numbers in parenthesis are standard errors, the numbers in square brackets are test statistics and the numbers in large brackets are p-values.
(3) Significance: ***: 0.01, **: 0.05, *:0.1.

The interesting results from Davidson and MacKinnon (1993, Ch7) and Durbin-Wu-Hausman (1978) test suggest that, in contrast with males, the schooling decision is exogenous for Canadian females. A plausible explanation might be gender discrimination in the labour market against women to education choice. Female wages may differ from male wages because of productivity differences that arise from differences in the absenteeism and turnover of males and females and differences in the human capital endowment that includes acquired attributes such as education, labour market information and labour market experience, as well as innate characteristics such as intelligence, strength, personality and drive. These unobserved characteristics of men are rewarded in the labour market due to discrimination against women. Thus, even though unobserved characteristics are correlated with education choices for women, they are not correlated with wages, which gives exogeneity. Also, in reality, due to employers' own ingrained prejudices as well as to erroneous information fostered by male customers and co-workers, female labours' productivity and unobserved motivation (energy) are consistently underestimated.

Note that both exogeneity tests above assume that the instrument (parental education) is valid. Thus, next, the validity of instrumental variables is examined.

The results of F test and Sargan test shown in table 6 suggest that parental education is a valid instrument, since the F test rejects the hypothesis that parental education had no effect on their children's education at the 5% level (63.224>1.94, 70.541>1.94 for females and males respectively) and Sargan test [11] fails to reject the overidentifying restrictions at the 5% level (1.708<14.067, 10.676<14.067 for females and males respectively).

Nevertheless, people may argue that parental education can correlate with children's ability or affects children's earnings via structural errors[12]. Deschenes (2002) argued that unspecified combination of individual specific abilities, influences of familial environment, and inherited skills might cause the unobserved heterogeneity in log earnings through the absolute advantage and comparative advantage[13]. But the author argues that there might exist another possibility shown below in figure 3.

The top graph in figure 3 shows two possible routes that make parent's education correlate with Children's unobserved ability. The bottom left graph shows that children's education is correlated with both parent's education and children's unobserved ability, but parent's education and children's unobserved ability are uncorrelated, which makes the parental education a valid instrument. The bottom right graph demonstrates that parent's ability is not correlated with children's unobserved ability through parent's nature and nurture, which again makes the parental education a valid instrument, though parent's education may be correlated parent's ability. In

---

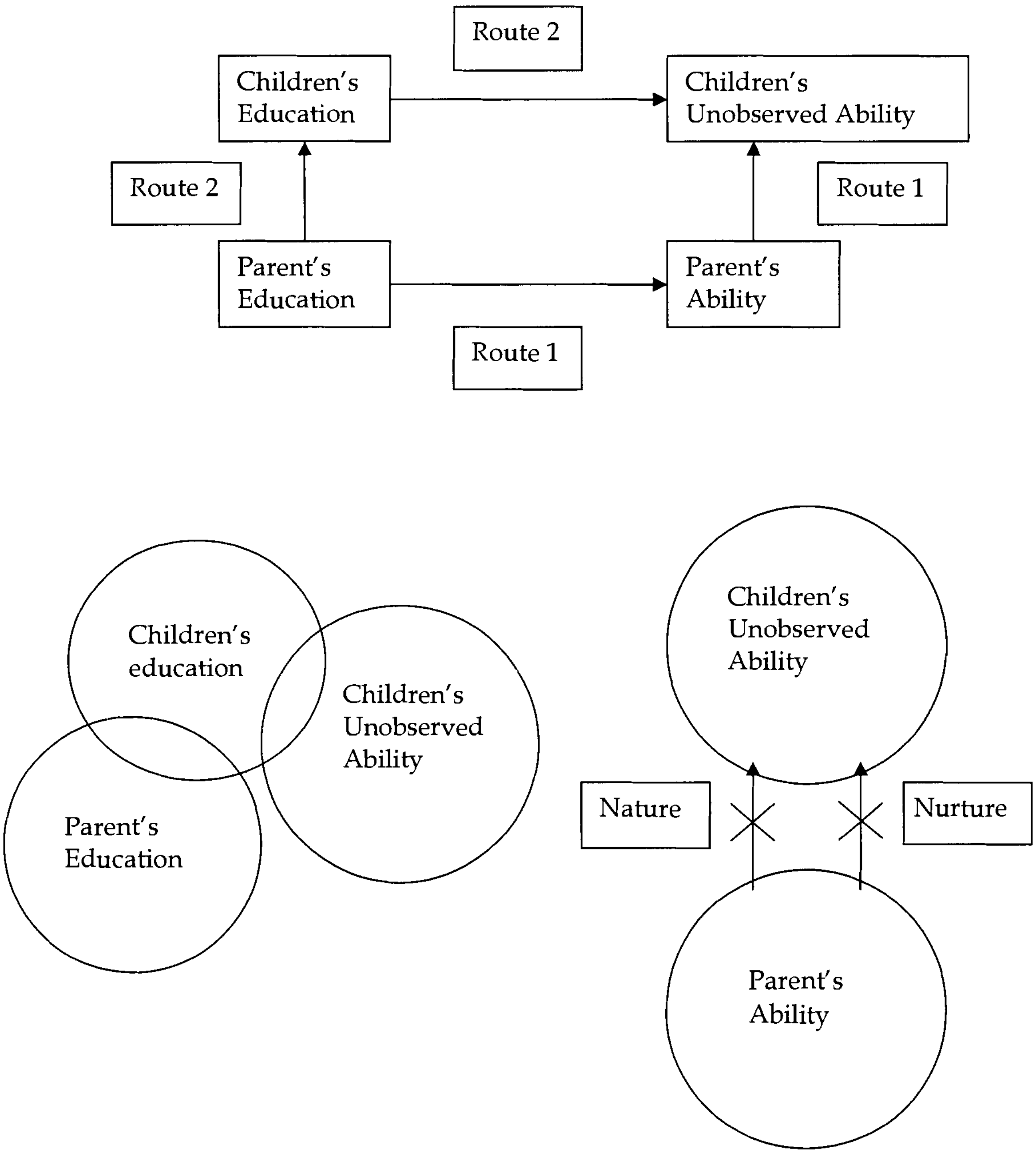
[11] The test statistic is obtained by taking a sample size multiplied by R squared of auxiliary regression of residuals from 2SLS upon the full set of instruments. The hypothesis being tested with the Sargan test is that the model is correctly specified and that the instruments used are valid; i.e., the imposed moment conditions are correct.

[12] Chung (2003) included parental income and individual innate skill (Armed Forces Qualification Test) in the wage equation. Parental education appears to be a valid instrument. Unfortunately, 94 SLID is lack of these control variables.

[13] The absolute advantage refers to the level of ability of individuals that does not interact with the level of schooling. The comparative advantage (endogeneity) refers to the other ability that is the heterogeneous component of the education slope interacting with the level of schooling and granting higher net returns to schooling to individuals with higher the other ability.

reality, parents who are smart do not necessary give birth to smart children and know how to foster their children and make them smart.

Figure 3: Correlation between parent's education and children's unobserved ability.



CHAPTER 4

CONCLUSIONS AND EXTENSIONS

Many researchers limit their samples to working males or married working females when estimating the returns to schooling. But in this paper, I examined the necessity for correcting for the problem of non-random sampling. Semiparametric approach is superior to parametric in terms of its better predictability in the range of low index people and its flexibility in assumptions. When there are endogenous regressors, without controlling for endogeneity (if technically infeasible), conventional semiparametric model provides biased estimates (See table 3). Since for females schooling is exogenous and parametric specifications are rejected through various normality tests and Horowitz test, we are pretty confident about our semiparametric results. We are not sure if there is a selectivity bias for working males (only Newey (1999) suggests the existence of a selectivity bias) because normality assumption is rejected; plus, schooling is endogenous. For males, 2SLS is the preferred specification judged in terms of the consistency criteria. The IV estimate for male wage workers is larger than OLS estimate, indicating a negative ability bias[14]. A positive self-selection bias for females is discovered. The returns to schooling for female wage workers in Canada is about 4.4~4.9%, which is higher than the returns for male wage workers of 3.9%.

---

[14] The direction that the IV estimation deviates from the OLS estimate depends on which group of individuals is influenced by the instrument. Because 94 SLID is a high quality dataset, it is unlikely that the effect of measurement error dominate the overall bias in OLS. Thus, the explanation for the large IV is a *negative* ability bias in the OLS estimates (See section 3.3.2.2).

There is no study in the literature that simultaneously disentangles sample selection bias and omitted ability bias, especially in a semiparametric framework. But pioneering work has been done by Blundell and Powell (2004)[15] in this area. They develop a control function approach to non-parametric estimation with endogenous regressors in a labour market participation model. The control function accounts for endogeneity in triangular and fully simultaneous binary response models. They find that the correction for endogeneity is important and the estimated effect of interest is shown to be strongly biased when inappropriate parametric distribution assumptions are imposed. Since semiparametric model is very sensitive to endogeneity (endogenous schooling) and it relies highly on exogeneity assumption, Newey (1999) and Klein & Spady (1993) might give unreliable results for male paid workers.

More studies solving the mixed problem of endogeneity, non-random sampling and any other violation of traditional parametric assumptions seem warranted. It would be interesting to have in-depth robustness checks by replacing logged hourly wage with logged annual labour earnings, studying the return to different schooling attainment, testing the nonlinearity of education and nonlabour income in, for instance, a semiparametric generalized partial linear model (See Hardle, Mammen and Muller (1998)) and instrumenting the potentially endogenous regressors such as nonlabour income and marital status on which our interest does not particularly centre.

---

[15] In the paper, they adopted three steps for estimation of the parameters of interest. The first step uses nonparametric regression methods to estimate the error term in the reduced form and the unrestricted conditional mean. The second step imposes the linear index assumptions on the unrestricted conditional mean to obtain the estimated eigenvector corresponding to the unique zero eigenvalue of variance covariance matrix of the augmented structural equation. The final step recovers an estimator of the marginal probability distribution function of the structural errors, which is defined to be the average structural function.

APPENDICES

Appendix A: Generalized Linear Model (GLM)

Econometric analysis of labour force participation by GLM was summarized by McCullagh and Nelder (1989) and takes a form of

$$E(D) = G(Z^{T}\gamma), \tag{A1}$$

where $E(D)$ denotes the expected value of the dependent variable D, Z is a vector of explanatory variables, γ is an unknown parameter vector and $G(\bullet)$ is a known link function. The expectation $\mu = E(D)$ depends on an index function $I = Z^{T}\gamma$. $Var(D) = \sigma^{2}Var(E(D)) = \sigma^{2}Var(\mu)$ and G (or G^{-1}) is the link function of I and μ. There are a number of ways to estimate the GLM.

In this paper, suppose $D = 1$ when one chooses to work and $D = 0$ when not. Then one could apply the linear probability model (LPM) that assumes dichotomous D is a linear function of Z. $E(D) = Z^{T}\gamma$ with an identity link function and can be interpreted as the conditional probability that one chooses to work given individuals' characteristics. However, the shortcomings of this approach are obvious. First, one cannot guarantee that the predicted probabilities from this model will lie between 0~1 interval. Second, the disturbances u cannot be normally distributed; it follows the binomial distribution. Third, even if $E(u_i) = 0$ and $E(u_i u_j) = 0$, for $i \neq j$, the disturbances u_i are not homoscedastic, since $Var(u \mid Z) = E(D)(1 - E(D)) = Z^{T}\gamma(1 - Z^{T}\gamma)$. Last, the computed

R^2 is likely to be low. Hence, more appropriate models should be introduced. Within the latent regression model framework (See equation (1) and (2)), we assume that u_i has mean zero and has either a standardized logistic with know variable $\pi^2/3$ or a standard normal distribution with variance 1. If the disturbances u are distributed as a logistic probability density function (PDF), Logit model that can be estimated by the GLM is given by

$$\Pr ob(D=1 \mid Z) = \Lambda(Z^{\mathrm{T}}\gamma) = \mu = \frac{e^{Z^{\mathrm{T}}\gamma}}{1+e^{Z^{\mathrm{T}}\gamma}}, \tag{A2}$$

where $\Lambda(\bullet)$ is the Logistic distribution function. If the disturbances u are distributed as a normal probability density function, Probit model is given by

$$\begin{aligned} &\Pr ob(D=1 \mid Z) = \Pr ob(D^* > 0 \mid Z) = \Pr ob(Z^{\mathrm{T}}\gamma + u > 0) = \Pr ob(u > -Z^{\mathrm{T}}\gamma \mid Z) \\ &= \Pr ob(u < Z^{\mathrm{T}}\gamma \mid Z) = \mu = \Phi(\frac{Z^{\mathrm{T}}\gamma}{\sigma_u}) = \int_{-\infty}^{T} \phi(t)dt \end{aligned}, \tag{A3}$$

where $\Phi(\bullet)$ is the standard normal distribution function and σ_u is the standard deviation of u. Two assumptions for identification are imposed. First is that variance of u is normalized to be 1; this is called scale normalization. Second is that the threshold of latent regression is normalized to be 0; this is named location normalization.

To estimate the Logit and Probit models, maximum likelihood method is implemented. Redenoting the distribution function of $(u \mid Z^{\mathrm{T}}\gamma)$ as $F(u)$ and its density as $f(u)$, a MLE would maximize

$$1/N\sum_{i=1}^{N} D_i \ln\{1 - F(-Z^{\mathrm{T}}\gamma)\} + (1-D_i)\ln\{F(-Z^{\mathrm{T}}\gamma)\} \tag{A4}$$

with respect to γ. Then one can obtain ML estimates of Probit or Logit coefficients. However, if the distribution is not normal or if the disturbance terms are not homoscedastic, the estimator of γ under certain specification will be inconsistent.

Appendix B: Klein and Spady (1993)

The weight function $w(\bullet)$ in equation (11) can be introduced for numerical or technical reasons. The probabilities are squared, for the probability estimates might be negative if a bias-reducing (higher order) kernel is employed for their estimation. The probability function $\Pr ob(D_i = 1 \mid Z_i) = \Pr ob(D_i = 1 \mid Z_i^{\mathrm{T}}\gamma) = \Pr ob(Z_i^{\mathrm{T}}\gamma > u_i \mid Z_i^{\mathrm{T}}\gamma)$ can be rewritten by Bayes rule as

$$\Pr ob(Z_i^{\mathrm{T}}\gamma > u_i \mid Z_i^{\mathrm{T}}\gamma) = \frac{\Pr ob(Z_i^{\mathrm{T}}\gamma > u_i) \cdot g_{\theta|D=1}(Z_i^{\mathrm{T}}\gamma \mid Z_i^{\mathrm{T}}\gamma > u_i)}{g_\theta(Z_i^{\mathrm{T}}\gamma)}, \tag{B1}$$

where g_θ is the density function of the index $Z^{\mathrm{T}}\gamma$ and $g_{\theta|D=1}$ the density of the index conditional on the selection status $D_i = 1$. This is equivalent to

$$\Pr ob(D_i = 1 \mid Z_i) = \frac{\Pr ob(D_i = 1) \cdot g_{\theta|D=1}(Z_i^{\mathrm{T}}\gamma \mid D_i = 1)}{g_\theta(Z_i^{\mathrm{T}}\gamma)} \tag{B2}$$

with the densities g_θ and $g_{\theta|D=1}$ nonparametrically estimable by univariate Kernels and $\Pr ob(D_i = 1)$ replaced by its sample average. The quasi-likelihood function can be computed for any coefficient vector γ. The vector γ that maximizes equation (11) is the Klein and Spady estimate $\hat{\gamma}_{Klein\&Spady}$. To computing the probability density functions Klein and Spady insert standard nonparametric density estimation algorithms to acquire

$$\Pr \hat{o} b(D_i = 1 \mid Z_i) = \frac{\frac{1}{N-1}\sum_{i \neq j}^{N} D_j \cdot \frac{1}{h} \cdot K\left(\frac{Z_i^{\mathrm{T}}\gamma - Z_j^{\mathrm{T}}\gamma}{h_n}\right)}{\frac{1}{N-1}\sum_{i \neq j}^{N} \frac{1}{h} \cdot K\left(\frac{Z_i^{\mathrm{T}}\gamma - Z_j^{\mathrm{T}}\gamma}{h_n}\right)}, \tag{B3}$$

where bandwidth h_n is a non-stochastic window satisfying condition

$n^{-1/6} < h_n < n^{-1/8}$ and $\int Z^2 K(Z) dZ = 0$. To achieve desired asymptotic properties the true probability distribution $\Pr ob(Z_i^{\mathrm{T}}\gamma > u_i \mid Z_i^{\mathrm{T}}\gamma)$ must be continuously differentiable with bounded derivatives. Moreover, small estimated densities that may disturb the outcomes should be adjusted or trimmed. The estimator is then both consistent and asymptotically normal, and achieves the semiparametric efficiency bound assuming that the errors and the regressors are independent,

$$\sqrt{n}(\hat{\gamma}_{Klein\&Spady} - \gamma) \xrightarrow{d} N(0, V_{Klein\&Spady}), \tag{B4}$$

where $V^{-1}_{Klein\&Spady} = E\left\{ \frac{\partial\Omega}{\partial\gamma} \frac{\partial\Omega}{\partial\gamma^{\mathrm{T}}} \frac{1}{\Omega\cdot(1-\Omega)} \right\}$

and $\Omega = \Pr ob(D_i = 1 \mid Z_i) = \Pr ob(Z_i^{\mathrm{T}}\gamma > u_i \mid Z_i^{\mathrm{T}}\gamma)$. Thus, if the estimated probability function is inserted into equation (11), the quasi-likelihood behaves like a likelihood function. Therefore, the variance matrix can be estimated by its parametric analog.

Identification[16] of single index models has been investigated by Ichimura (1993) and, for binary models, by Manski (1988). It is worthwhile to note that coefficient estimates for the models are not directly comparable and scale normalization is required.

---

[16] Horowitz and Lee (2002) clarified the importance of restrictions to insure identification before one can estimate coefficients and link function. The semiparametric conditional mean function takes the form $m(Z) = \varphi(Z^{\mathrm{T}}\gamma) = E(D \mid Z = z) = \varphi^*(A + B\cdot Z^{\mathrm{T}}\gamma)$, where A and $B \neq 0$ are arbitrary and φ^* is defined by the relation $m(Z) = \varphi(Z^{\mathrm{T}}\gamma) = E(D \mid Z = z) = \varphi^*(A + B\cdot Z_i) = \varphi(Z_i)$ for all Z_i in the support of $Z^{\mathrm{T}}\gamma$. Therefore, γ and φ are not identified unless restrictions are imposed that uniquely specify A and B. The restriction on A is called *location normalization* and can be imposed by requiring Z to contain no intercept. The restriction on B is called *scale normalization* that can be realized by

Appendix C: The Data Set

Variables and definition

Variable Names	Definition	Variable Names in 94 SLID
WAGE WORKER	Dichotomous indicator that equals 1 if an individual is a wage worker and 0 if not. The category of non-wage workers includes individuals who are either self-employed (farming or family business) or engaged in non-market home production	FPDWK28C FSEUI28C FSEIN28C FUNFW28C
LWAGE	The log of composite hourly wage rate: the average of all paid worker jobs held by the person (excluding any self employment) during the year	CMPHW28C
AGE10	Respondent's age in years divided by 10	EAGE26C
AGE10SQ	AGE10^2	---
N0004C	Number of children less or equal to 4 years old	N000427C
N0519C	Number of children between 5 and 19 including 5 and 19	N050827C N001427C N151927C
NLINCOME	Respondent's yearly non-labour income ((Economic family's annual income -Respondent's annual labour income)/10,000)	EARNG27C TTWGS28C
MARRIED	Dichotomous indicator that equals 1 if an individual is married in 1994 and 0 if an individual is in a common-law status, a separated, divorced, widowed or single (never married) status.	MARST26C
EDU	Respondent's years of schooling	YRSCH18C
HHSIZE	The number of family members in the household	HHSZ25C
DISABLED	Dichotomous indicator that equals 1 if an individual has a long-term condition that limits the person at home, at school or in other activities, or in the kind or amount of activity s/he can do at work and 0 if not	DISAB26C
MOTHER TONGUE	Three dichotomous variables (ENGLISH, FRENCH). Other is the left-out dummy variable	MOTN2G15
GROUP STATUS	Three dummy variables indicating group status (ABORIGINAL VISIBLEM, WHITE). WHITE is the left out dummy, which includes people who have Canadian, British, French or Other European ethnic	EOABOR15 EOVM15 EOCAN15 EOBRIT15

setting the γ coefficient of one component of Z equal to 1. A further identification is that Z must include at least one continuously distributed component whose γ coefficient is non-zero.

	background	EOFR15 EOEURO15
REGION OF RESIDENCE	Five dummy variables (ATLANTIC QUEBEC PRAIRIES BC) ONTARIO is the omitted category	REGRE25C
PEXP10	Potential labour market experience (Approximated by (AGE-EDU-6)/10)	---
PEXP10SQ	Potential labour market experience squared (PEXP10 ^2)	---
PARENTAL EDUCATION	Five levels of schooling as dummy variables indicating respondent's mother and father's level education respectively (EDMOTH1/EDFATH1: Elementary school (includes no schooling), EDMOTH2/EDFATH2: some high school, EDMOTH3/EDFATH3: Completed high school, EDMOTH4/EDFATH4: Non-university certificate or diploma (e.g. community college), EDMOTH5/EDFATH5: University degree (no level specified)). EDMOTH3 and EDFATH3 are the left-out dummies	EDMOTH21 EDFATH21

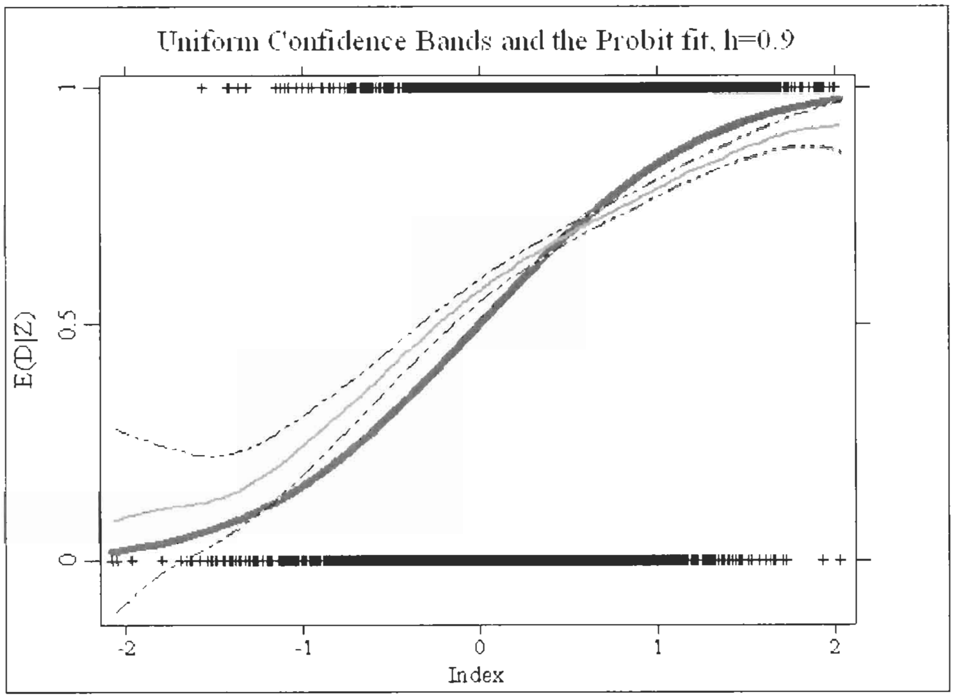
Means of the variables by gender and employment status

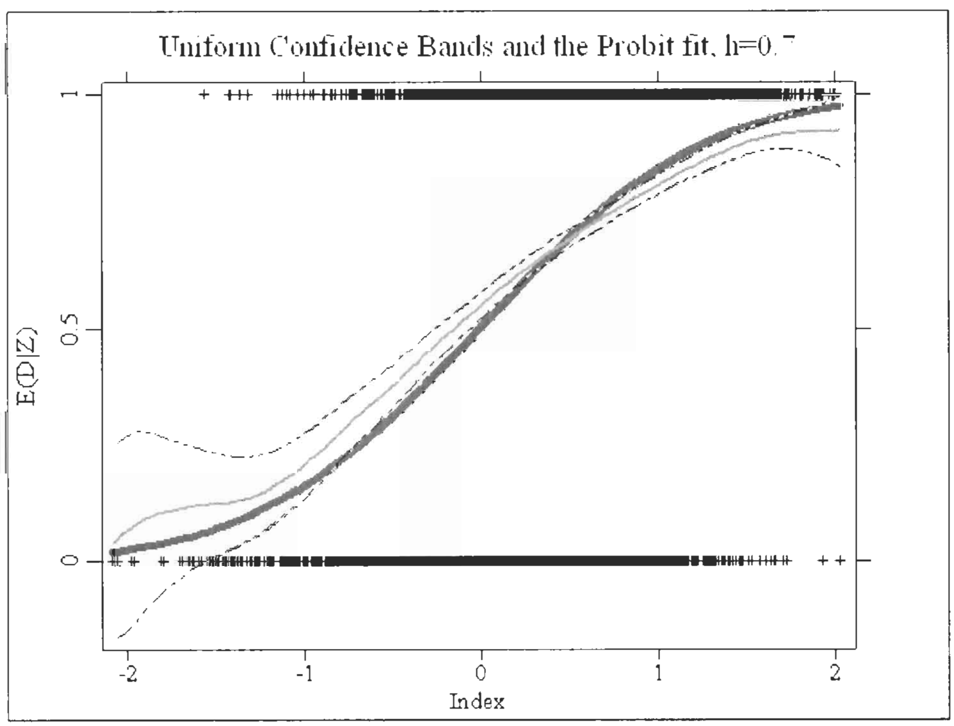
Variables	Female		Male	
	Paid	*Not paid*	*Paid*	*Not paid*
Number of observations	5039 67.23%	2456 32.77%	5288 75.94%	1675 24.06%
Endogenous variables				
WAGEWORKER	1.00	0.00	1.00	0.00
LWAGE	1.09 (0.21)	---	1.21 (0.19)	---
Exogenous identifying variables				
NLINCOME ('0000 CAN $)	2.94 (2.81)	3.09 (3.31)	1.69 (2.31)	3.10 (3.62)
HHSIZE	3.21 (1.28)	3.39 (1.38)	3.27 (1.34)	3.19 (1.43)
EDFATH1	0.48 (0.50)	0.60 (0.49)	0.49 (0.50)	0.58 (0.49)
EDFATH2	0.22 (0.42)	0.18 (0.38)	0.23 (0.42)	0.20 (0.40)
EDFATH4	0.08 (0.27)	0.07 (0.26)	0.07 (0.26)	0.05 (0.23)
EDFATH5	0.06 (0.24)	0.03 (0.16)	0.06 (0.23)	0.05 (0.22)
EDMOTH1	0.41 (0.49)	0.55 (0.50)	0.43 (0.49)	0.54 (0.50)
EDMOTH2	0.25 (0.43)	0.20 (0.40)	0.23 (0.42)	0.20 (0.40)
EDMOTH4	0.11 (0.31)	0.07 (0.26)	0.09 (0.29)	0.08 (0.27)
EDMOTH5	0.03 (0.17)	0.03 (0.16)	0.03 (0.18)	0.02 (0.14)
Remaining variables				
EDU	13.19 (2.84)	11.58 (3.15)	12.91 (3.25)	11.83 (3.70)
N0004C	0.19 (0.39)	0.22 (0.41)	0.21 (0.41)	0.15 (0.36)
N0519C	0.68 (0.79)	0.71 (0.81)	0.66 (0.80)	0.60 (0.79)
DISABLED	0.05 (0.21)	0.17 (0.38)	0.06 (0.24)	0.20 (0.40)
ENGLISH	0.69 (0.46)	0.63 (0.48)	0.66 (0.47)	0.65 (0.48)
FRENCH	0.22 (0.42)	0.26 (0.44)	0.26 (0.44)	0.23 (0.42)
ABORIGINAL	0.02 (0.15)	0.03 (0.18)	0.02 (0.14)	0.01 (0.10)
VISIBLEM	0.03 (0.17)	0.03 (0.18)	0.03 (0.18)	0.02 (0.15)
ATLANTIC	0.21 (0.40)	0.24 (0.43)	0.21 (0.41)	0.20 (0.40)
QUEBEC	0.18 (0.39)	0.23 (0.42)	0.21 (0.41)	0.20 (0.40)
PRAIRIES	0.24 (0.42)	0.20 (0.40)	0.22 (0.42)	0.25 (0.43)
BC	0.10 (0.30)	0.08 (0.28)	0.09 (0.28)	0.09 (0.28)
PEXP10	2.02 (0.97)	2.52 (1.09)	2.07 (1.00)	2.58 (1.07)
AGE10	3.94 (0.88)	4.28 (0.98)	3.96 (0.90)	4.37 (0.94)
MARRIED	0.69 (0.46)	0.75 (0.43)	0.71 (0.46)	0.68 (0.47)

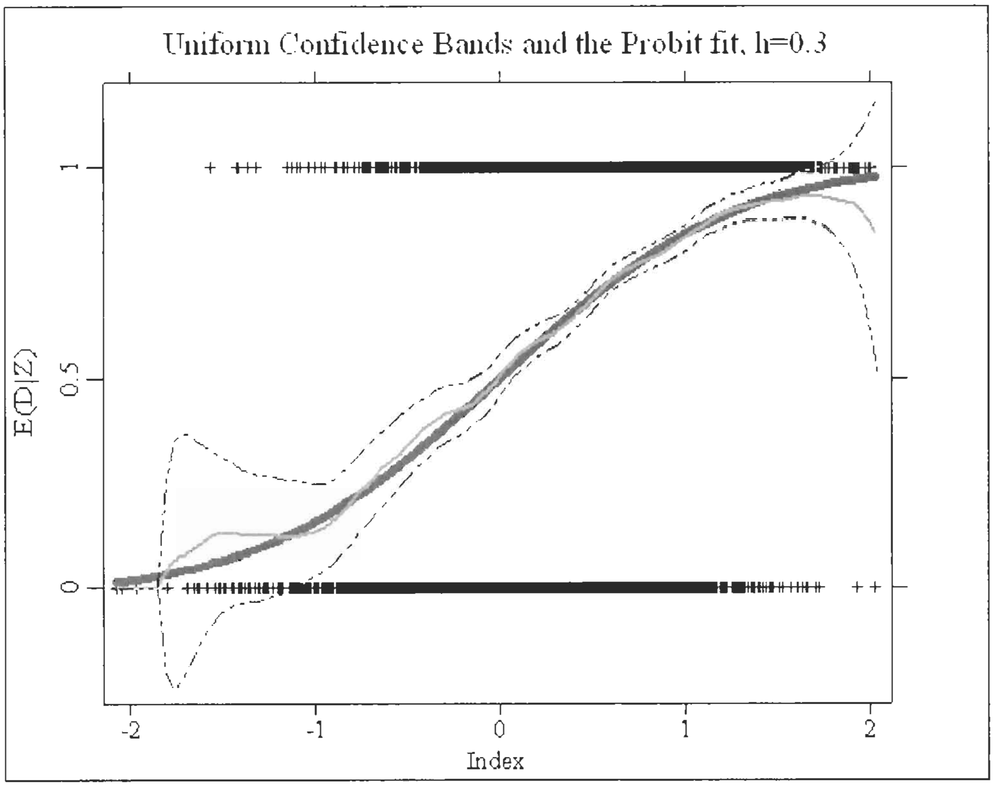
Notes: The numbers in parentheses are the standard deviations of the means.

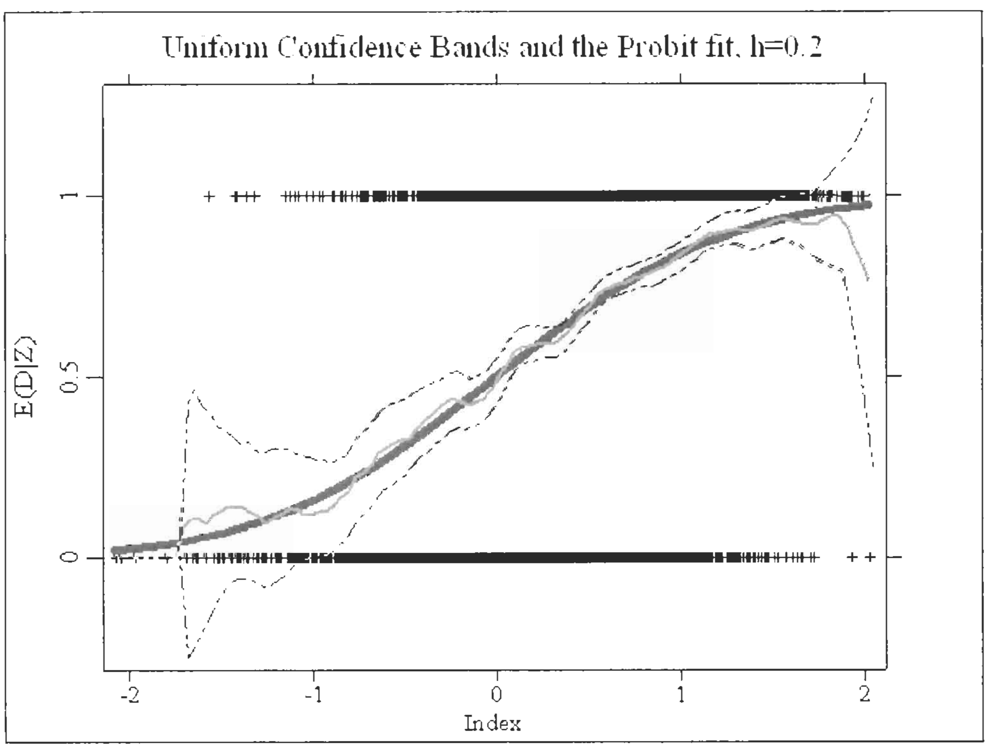
Appendix D: Result Figures

Horowitz (1993) test for females

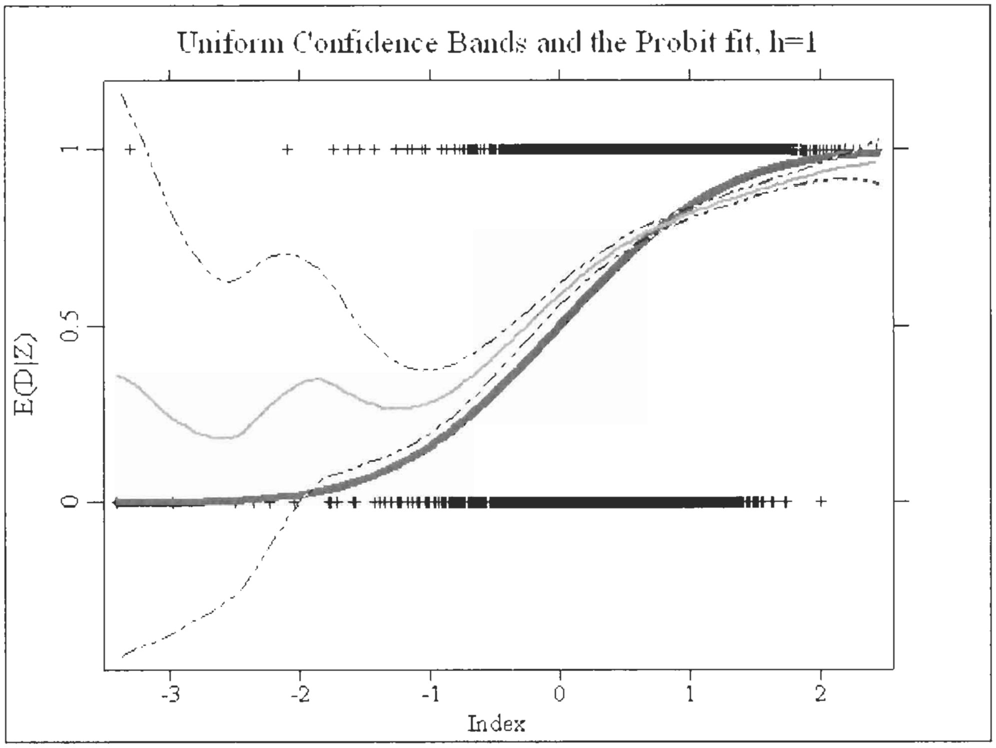


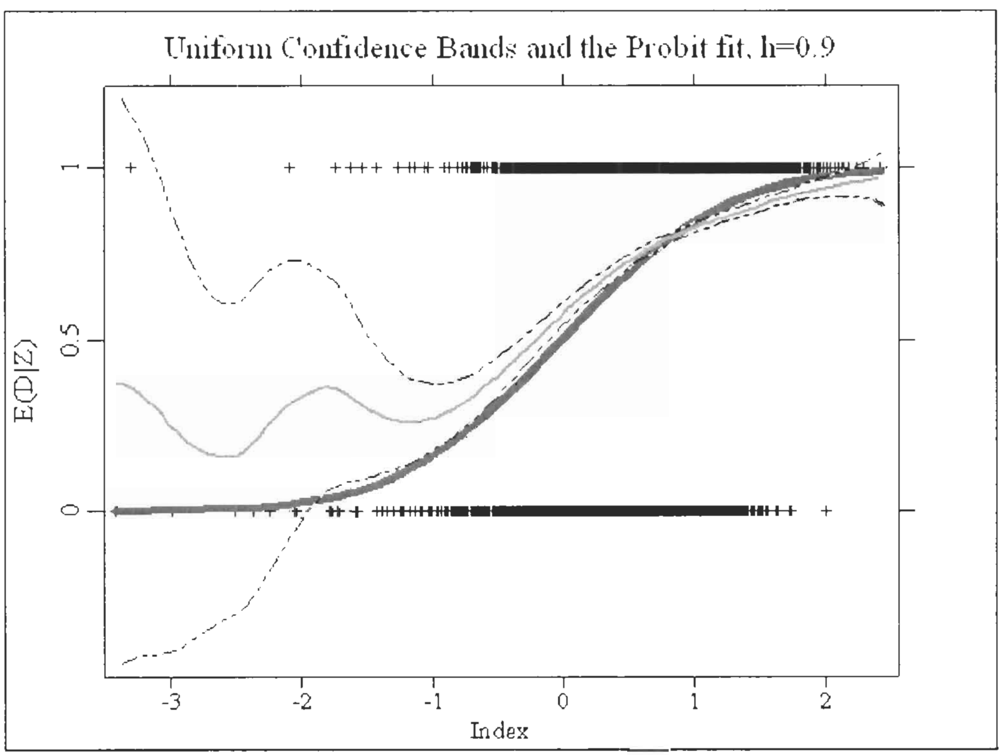


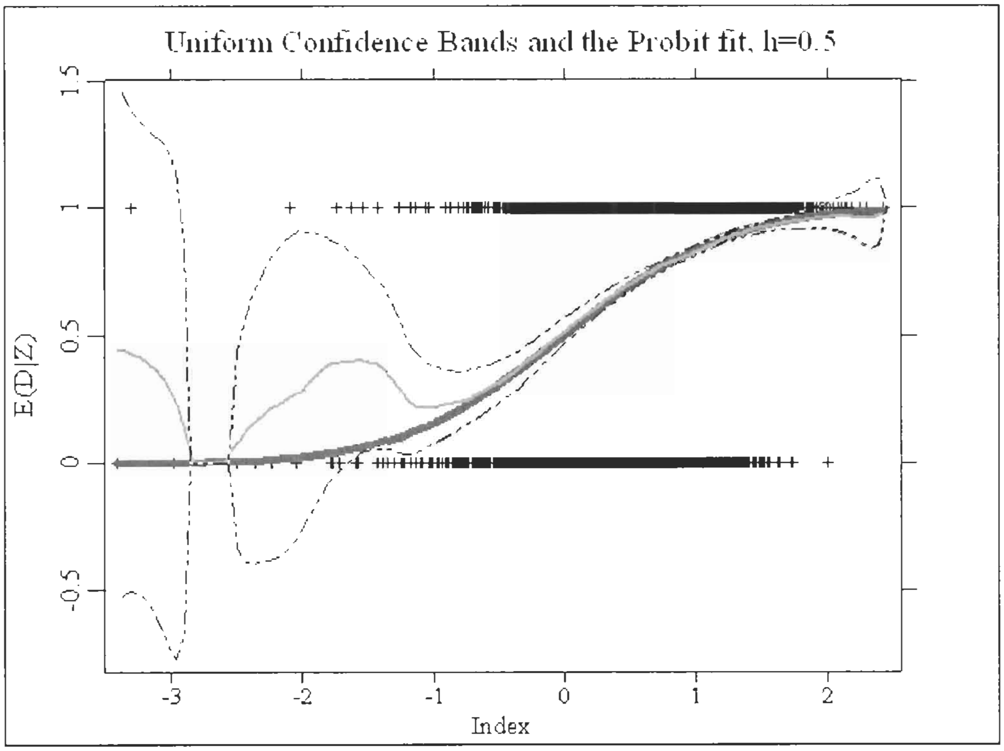

Uniform Confidence Bands and the Probit fit, h=0.3
E(D|Z)
1
0.5
0
-2
-1
0
1
2
Index

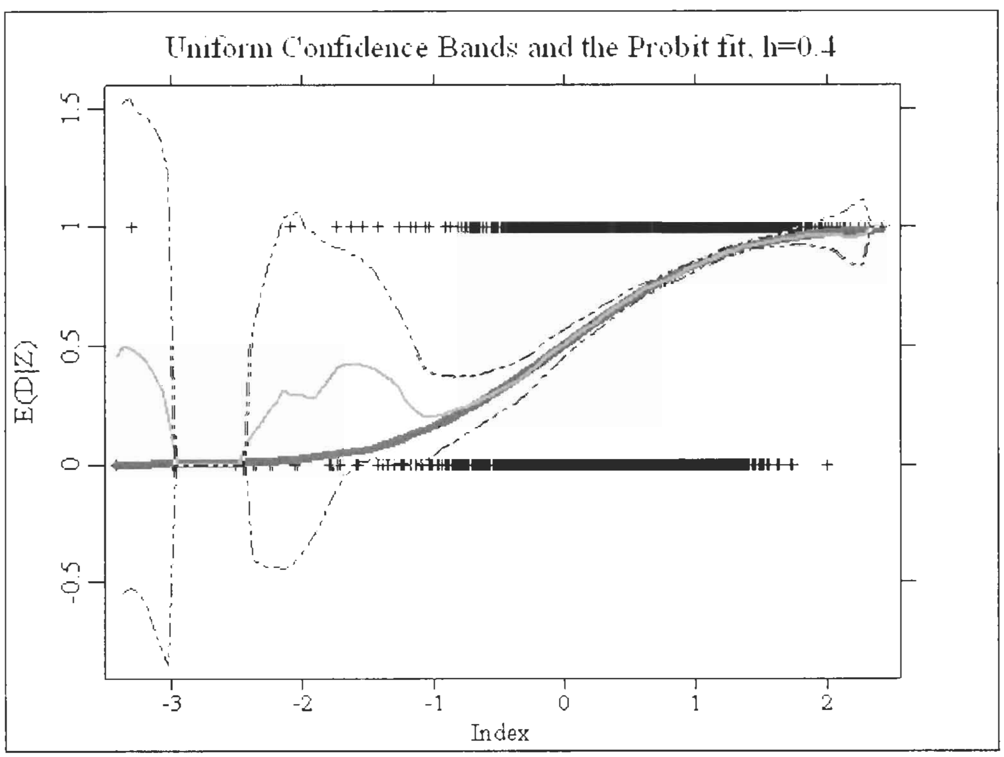

Uniform Confidence Bands and the Probit fit, h=0.2
E(D|Z)
1
0.5
0
-2
-1
0
1
2
Index

Horowitz (1993) test for males






Uniform Confidence Bands and the Probit fit, h=0.5
E(D|Z)
1.5
1
0.5
0
-0.5
-3
-2
-1
0
1
2
Index


Uniform Confidence Bands and the Probit fit, h=0.4
E(D|Z)
1.5
1
0.5
0
-0.5
-3
-2
-1
0
1
2
Index

Appendix E: Generalized Cross-Validation

To select the optimal bandwidth in nonparametric regression, Generalized Cross-Validation (GCV) is used. The Nadaraya-Watson estimator, defined as

$$\hat{m}_h(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) y_i}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)}, \tag{E1}$$

approaches y_i when the bandwidth h approaches zero. Mean Squared Error (MSE) is obtained by the cross-validation criterion, defined as

$$CV(h) = \sum_{i=1}^{n} \left\{ y_i - \hat{m}_{h,-i}(x_i) \right\}^2, \tag{E2}$$

where $\hat{m}_{h,-i}(x_i)$ denotes as $\hat{m}_{h,-i}(x_i) = \frac{\sum_{i \neq j}^{n} K_h(x_i - x_j) y_i}{\sum_{i \neq j}^{n} K_h(x_i - x_j)}$ the kernel regression estimate which is obtained without using the i th observation (x_i, y_i). Härdle (1990) showed that equation (E2) can be written as $CV(h) = \sum_{i=1}^{n} \left\{ y_i - \hat{m}_h(x_i) \right\}^2 \Xi(W_{hi}(x_i))$ with $\Xi(u) = (1-u)^{-2}$, the penalty function for generalized cross-validation criterion, and $W_{hi}(x_i) = \frac{K(0)}{\sum_{j=1}^{n} K(\frac{x_i - x_j}{h})}$. Here, small values for bandwidth chosen are penalized by the penalty function and the (leave-one-out) cross-validation is treated as a Sum of

Squared Residual (SSE). Since minimizing $CV(h)$ is on average equivalent to

minimizing $MSE(h)$, one could choose $\hat{h}_{Optimal}$ by minimizing $CV(h)$.

REFERENCES

[1] ANDREWS, DONALD W. K.; SCHAFGANS, MARCIA M. A., (1998), 'Semiparametric Estimation of the Intercept of a Sample Selection Model,' *Review of Economic Studies*, July 1998, v. 65, iss. 3, pp. 497-517.

[2] ANGRIST, JOSHUA D., AND ALAN B. KRUEGER, (1991): 'Does Compulsory School Attendance Affect Schooling and Earnings,' *Quarterly Journal of Economics*, 106, 979-1014.

[3] BERA, ANIL K.; JARQUE, CARLOS M.; LEE, LUNG-FEI, (1984), 'Testing the Normality Assumption in Limited Dependent Variable Models,'*International Economic Review*, October 1984, v. 25, iss. 3, pp. 563-78.

[4] BLUNDELL, RICHARD W.; POWELL, JAMES L., (2004), 'Endogeneity in Semiparametric Binary Response Models,' *Review of Economic Studies*, July 2004, v. 71, iss. 3, pp. 655-79

[5] BUCHINSKY, MOSHI, (1998), 'Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research,' *Journal of Human Resources*, Vol 13, iss. 1, pp. 1-30.

[6] BUTCHER, K.F. CASE, A., (1994). The effect of sibling sex compostion on women's education and earnings. *Quarterly Journal of Economics* 109, 531-563.

[7] CARD, DAVID, (1999) 'The causal effect of education on earnings,' in *Handbook of Labor Economics*, ed. Orley Ashenfelter and David Card (Amsterdam and New York: North-Holland).

[8] ---, (2001) 'Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems,' *Econometrica*, 69, 1127-1160.

[9] CHRISTOFIDES LOUIS N; LI QI, ZHENJUAN LIU; INSIK MIN, (2003), 'Recent Two-Stage Sample Selection Procedures with an Application to the Gender Wage Gap,' *Journal of Business and Economic Statistics*, 21, 3.

[10] CHUNG, T. -P., (2003), 'Returns to education: Updates for Malaysia,' *Applied Economics Letters*, V. 10, iss. 13, pp. 837-41.

[11] DAVIDSON, R. AND J. G. MACKINNON, (1993), *Estimation and Inference in Econometrics, chapter 7*, (New York: Oxford University Press).

[12] DESCHENES, OLIVIER, (2002), 'Estimating the Effects of Family Background on the Return to Schooling,' Department of Economics, UC Santa Barbara, University of California at Santa Barbara, *Economics Working Paper Series*: 1020.

[13] GRILICHES, Z., (1976), 'Wages of Very Young Men,' *Journal of Political Economy*, 84, S69-S85.

[14] HARDLE, WOLFGANG., (1990), *Applied Nonparametric Regression*, Econometric Society Monographs No. 19, Cambridge University Press.

[15] ---; OLIVER, LINTON, (1994), 'Applied Nonparametric Methods,' *Handbook of Econometrics, IV.*

[16] ---; ENNO MAMMEN; MARLENE MULLER, (1998), 'Testing parametric versus semiparametric modelling in generalized linear models,' *Journal of American Statistical Association*, 93, 444 page 1461.

[17] HAUSMAN, JERRY A, (1978), 'Specification Tests in Econometrics,' *Econometrica*, 46, iss. 6, pp. 1251-71.

[18] HECKMAN, JAMES, (1974), 'Shadow Prices, Market Wages and Labor Supply,' *Econometrica*, 42, 679-694.

[19] ---; (1979), 'Sample Selection Bias as a Specification Error,' *Econometrica*, 47, 153-161.

[20] HOROWITZ, JOEL L., (1993), 'Semiparametric Estimation of a Work-Trip Mode Choice Model,' *Journal of Econometrics* v. 58, iss. 1-2, pp. 49-70.

[21] ---; LEE SOKBAE, (2002), 'Semiparametric methods in applied econometrics: do the models fit the data?' *Statistical Modelling* 2:3-22.

[22] ICHIMURA, HIDEHIKO, (1993),'Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models,' *Journal of Econometrics*, July 1993, v. 58, iss. 1-2, pp. 71-120.

[23] KANE, THOMAS J., AND CECILIA E. ROUSE, (1993): 'Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?' *NBER Working Paper #4268.*

[24] KLEIN, ROGER W.; Spady, Richard H., (1993), 'An Efficient Semiparametric Estimator for Binary Response Models,' *Econometrica*, March 1993, v. 61, iss. 2, pp. 387-421.

[25] LEMIEUX, THOMAS, AND DAVID CARD, (2001): 'Education, Earnings, and the 'Canadian G.I.Bill,' *Journal of Canadian Economics*, #Vol. 34, No.2, 313-344.

[26] MALUCCIO, JOHN, (1997): 'Endogeneity of Schooling in the Wage Function,' *Unpublished Manuscript*, Yale University Department of Economics.

[27] MANSKI, CHARLES F, (1988), 'Identification of Binary Response Models,'*Journal of the American Statistical Association*, September 1988, v. 83, iss. 403, pp. 729-38.

[28] MARTINS, M.F.O., (2001), 'Parametric and Semiparametric Estimation of Sample Selection Models: An empirical Application to the Female Labour Force in Portugal', *Journal of Applied Econometrics*, 16: 23-39.

[29] McCULLAGH P. AND J. A. NELDER, (1989). *Generalized Linear Models*. (Chapman and Hall, New York, NY).

[30] NEWEY, WHITNEY, (1999), 'Two Step Series Estimation of Sample Selection Models,' Massachusetts Institute of Technology (MIT), Department of Economics, *Working papers*: 99-04.

[31] POWELL, J.L., J.H. STOCK AND T.M. STOKER, (1989), 'Semiparametric Estimation of Weighted Average Derivatives,' *Econometrica*, 57, 1403-143

[32] ROBINSON, PETER M., (1988), 'Root- N-Consistent Semiparametric Regression,' *Econometrica*, v. 56, iss. 4, pp. 931-54.

[33] SARGAN, J. D. , (1958), 'The estimation of economic relationships using instrumental variables,' *Econometrica*, v. 26, pp. 393-415.

[34] SCHAFGANS, MARCIA M. A., (1997), 'Semiparametric Estimation of a Sample Selection Model: A Simulation Study, Sticerd,' *Discussion paper* No. EM/97/326, London School of economics.

[35] ---, (1998), 'Ethnic Wage Differences in Malaysia: Parametric and Semiparametric Estimation of the Chinese-Malay Wage Gap,' *Journal of Applied Econometrics* V.13, iss.5, pp, 481-504.

[36] --- ,(2000), 'Gender Wage Differences in Malaysia: Parametric and Semiparametric Estimation,' *Journal of Development of Economics*, Vol. 63, iss. 2, pp. 351-78.

[37] --- ,(2004), 'Finite Sample Properties for the Semiparametric Estimation of the Intercept of a Censored Regression Model,' *Statistica Neerlandica*, Vol. 58, 1, 35-56.

[38] VELLA, FRANCIS, (1998), 'Estimating models with sample selection bias: A survey,' *Journal of Human Resources*, Vol 33, pp. 127-169.

[39] YATCHEW, ADONIS, (2003), '*Semiparametric regression for the applied econometrician*,' Chapter 3 and 4, (Cambridge University Press).