ANNALS OF ECONOMICS AND FINANCE **6**, 155–167 (2005)

A Simple Matching Method for Estimating Sample Selection Models Using Experimental Data

Songnian Chen

The Hong Kong University of Science and Technology

and

Yahong Zhou

The Shanghai University of Economics and Finance

In this paper estimation of sample selection models using experimental data is considered with some weak restriction imposed on the error distribution. Under a normality setting, the most popular approach is the two-step method proposed by Heckman (1979). But Heckman's approach relies on the nonlinearity of the probit function (i.e. the nonlinearity of the selection correction function) unless some exclusion restriction is imposed. Furthermore, Heckman's method is sensitive to the underlying distributional assumption. Following this two-step method, several semiparametric estimators have been proposed for sample selection models by explicitly imposing the exclusion restriction. Using experimental data, this paper proposes a simple semiparametric matching method. There are certain advantages of our estimator over Heckman's estimator and the existing semiparametric estimators under either the parametric setting and semiparametric setting. We do not rely on the nonlinearity of the selection correction function or the exclusion restriction. In addition, unlike other semiparametric methods, we can also estimate the intercept term in the equation of interest. The estimator is shown to be consistent and asymptotically normal under some regularity conditions. A small monte carlo study illustrates the usefulness of the new estimator.

Key Words: Matching method; Experimental data.
JEL Classification Numbers: C30, C40.

1. INTRODUCTION

In this paper estimation of a regression equation (the outcome equation) subject to a sample selection rule and random assignment is considered

1529-7373/2005

based on experimental data. The experimental data in question here is generated as follows. In the first stage the selection rule identifies the group of nonparticipants for whom the regression equation is observable. In the second stage, some randomization scheme is applied to the remaining individuals, among whom the regression equation is only observable for the randomized-out group. In the context of job training program evaluations (see e.g., Heckman et. al 1998), we are interested in estimating the earnings equation for nontrainees using experimental data (henceforth the baseline earnings equation). This is an important step in determining various aspects of the program benefits and characterizing the selection bias. In the first stage, a selection rule classifies the whole sample into nontrainees (or nonparticipants) and prospective trainees. In the second stage only a fraction of the prospective trainees receive training according to certain random assignment, while the remaining portion of the latter group are randomized out for the training program. Thus we observe the baseline earnings equation for the nonparticipants and the randomized-out control group.

Estimation of sample selection models in the context of evaluating various training programs has mainly been based on techniques developed for nonexperimental data sets. In recognizing deficiencies of the conventional methods and limitations of the nonexperimental data, many researchers have turned to the available experimental data to recover various aspects of training programs (see, e.g., Lelonde (1986), Fraker and Maynard(1987), Heckman et. al (1998)). There have been numerous social experiments, especially for the purpose of evaluating the impact of federal job training on earnings and employment. In this paper we propose a new approach to estimating sample selection models by taking advantage of unique features of experimental data.

For sample selection models, the least squares method would produce inconsistent estimates due to the presence of the selection correction term in the outcome equation. The usual approach is to specify the distribution of the underlying errors parametrically, normality in particular, and independent of the explanatory variables. Then the parameters of the model can be consistently estimated by maximum likelihood or other likelihood based methods. The two-step method proposed by Heckman (1974,1976) is by far the most popular approach by including a consistent estimate of the selection correction term as part of the regressors in the outcome equation. One important limitation of Heckman's approach is its reliance on the nonlinearity of the probit function (i.e. the nonlinearity of the selection correction function) unless some exclusion restriction is imposed. More significantly, misspecification of the error distribution in sample selection models will in general render likelihood-based estimators inconsistent. Since a parametric form of error distributing can not generally be

justified by economic theory, following Heckman's two-step approach, several semiparametric estimation methods have been proposed recently for sample selection models (see, e.g., Andrews, Newey (1988), Powell (1989), Heckman et. al (1998), among others), which only assume weak restriction on the error distribution to guard against possible misspecification. These semiparametric estimators, however, require that the exclusion be satisfied. In addition, the intercept in the outcome equation is absorbed into the selection correction term in these semiparametric approaches, thus can not be estimated along the slope parameters.

In the paper we propose a new semiparametric estimator by taking advantage of unique features of experiment data. The idea behind our estimator is based on the following observation. In the experimental data with mild conditions there exist pairs of individuals with offsetting selection biases. Simple matching of such pairs would eliminate the selection bias in a straightforward way. Our estimator does not rely on the nonlinearity of the selection correction function or exclusion restriction. Furthermore, unlike other semiparametric methods, we can also consistently estimate the intercept term in the outcome equation.

This paper is organized as follows. The next section describes the model and motivates the proposed estimator. Section 3 gives regularity conditions and investigates the large sample properties of the estimator. They are shown to be consistent and asymptotically normal. Section 4 reports a small monte Carlo study. The final section concludes.

2. THE MODEL AND ESTIMATORS

We consider estimation of the sample selection model with experimental data defined by

$$y = x\beta_0 + u \tag{1}$$

$$d = 1\{w\delta_0 - v > 0\} \tag{2}$$

where we wish to estimate $\beta_0 \in R^{K_2}$ and $\delta_0 \in R^{K_1}$ based on observations of $(d, (1-d)y, d(1-R)y, x, w))$, and R is an random variable independent of the other variables in the model that can take on values 0 and 1. Here y is the potential outcome equation. d is a discrete choice variable, (x, w) are vectors of exogenous variables which may have components in common and R is an randomization indicator. Let z be a vector consist of the distinct components in (x, w). In the first stage, the selection equation (2) determines the subsample of nonparticipants with $d = 0$ for whom the equation (1) is observable. For the remaining individuals, the potential outcome equation is observable according as the randomization indicator R is equal to 0 or not. Consequently the potential outcome equation is

observable if $d_1 = 1$ or $d_2 = 1$ where $d_1 = (1-d)$ and $d_2 = d(1-R)$. In the context of training programs evaluations with experiment data (see, e.g., Heckman et. al. 1998), here we are interested in estimating the baseline earnings equation for nontrainees. $d = 1$ indicates that a person applies and is provisionally accepted into the program before the act of randomization. $R = 1$ if a person for whom $d = 1$ is randomly assigned into the program, and $R = 0$ if the person is denied access to the program. Therefore we observe the baseline earnings equation for the individuals with $d_1 = 1$ or $d_2 = 1$.

When the error terms are normally distributed, the model can be estimated by maximum likelihood. But the usual approach is the computationally simpler two-step method first proposed by Heckman (1974,1979). Under normality we have

$$E(u|d_1 = 1, z) = E(u|d = 0, z) = \sigma_{12}\sigma_1^{-1}\lambda_1(w\delta_0/\sigma_1)$$

and

$$E(u|d_2 = 1, z) = E(u|d = 1, R = 0, z) = E(u|d = 1, z) = \sigma_{12}\sigma_1^{-1}\lambda_2(w\delta_0/\sigma_1)$$

where σ_{12}=cov(u, v), σ_1=var(v), $\lambda_1(t) = \phi(t)/\Phi(t)$, and $\lambda_2(t) = -\phi(t)/(1-\Phi(t))$ with $\phi(t)$ and $\Phi(t)$ denoting the density and distribution functions for the standard normal random variable. Define $d^* = d_1 + d_2$, so we have

$$E[u - d_1\lambda_1(w\delta_0) - d_2\lambda_2(w\delta_0)|d^* = 1, z] = 0$$

Heckman's two step estimator is based on the following moment equation for the subsample $d^* = 1$,

$$y = x\beta_0 - \sigma_{12}\sigma_1^{-1}(d_1\lambda_1(w\delta_0) + d_2\lambda_2(w\delta_0)) + \epsilon_1 \tag{3}$$

such that $E(\epsilon_1|d_1, d_2, z) = 0$. Given a first step estimator $\hat{\delta}$ for δ_0, β_0 can be consistently estimated by regressing y on $(x, d_1\lambda_1(w\hat{\delta}) + d_2\lambda_2(w\hat{\delta}))$ for the subsample d^*=1. Note that when $x = w$, this approach will depend on the nonlinearity of the $\lambda_1(w\delta_0)$ and $\lambda_2(w\delta_0)$. However, as pointed out by Leung and Yu (1996), Nawata (1994), and Vella (1995), among others, these functions can be close to be linear in certain ranges, which might lead to unreliable estimates for β_0.

Another potentially more serious drawback to this and other likelihood-based methods is their sensitivity to the assumed parametric distribution of the unobservable error terms in the model. Recently several semiparametric estimators (e.g., Andrews (1991), Newey (1988), Powell (1989), among others) have been proposed for sample selection models which do not impose

parametric forms on error distributions. In the context of experimental data, these semiparametric estimators are based on the following observation. Under the condition the error term (u, v) is independent of the regressors (or the slightly weaker assumption of the index restriction), we have the following partial linear setup for the subsample $d^* = 1$

$$y = x\beta_0 + d_1 K_1(w\delta_0) + d_2 K_2(w\delta_0) + \epsilon_2 \tag{4}$$

where $K_1(w\delta_0) = E(u|d_1 = 1, z)$, $K_2(w\delta_0) = E(u|d_2 = 1, z)$ are unknown selection correction terms, and $E(\epsilon_2|d_1, d_2, z) = 0$. The objective of these approaches is to eliminate the contaminating selection correction terms in (4). Notice, however, that under the setup of equation (4), it is necessary for w to have component not included in x for identification. In addition, an explicit intercept term is not allowed in β_0 since it would be absorbed in the selection correction terms.

Instead of taking the equation (4) as the departure point, our estimation approach will rely on the zero mean restriction $E(u) = 0$. The idea behind our estimator is based on the following observation. If β_0 were known, then we can observe u given $v > w\delta_0$ for individuals with $d_1 = 1$ and u given $v < w\delta_0$ for individuals with $d_2 = 1$. Under random sampling the combination of the error terms $u_i 1\{v_i > t\} + u_j 1\{v_j < t\}$ will have the same moments as u_i, for any constant t. If there exist pairs of observations i and j with $w_i\delta_0 = w_j\delta_0$, then the zero mean restriction and a matching of these two observations lead to the following zero mean condition

$$\begin{aligned} &E\{(1 - R_j)[d_{1i}u_i + d_{2j}u_j]|w_i\delta_0 = w_j\delta_0\} \\ &= E(1 - R_j)E[d_{1i}u_i + d_{2j}u_j]|w_i\delta_0 = w_j\delta_0] \\ &= 0 \end{aligned}$$

i.e.

$$E\{(1 - R_j)[(1 - d_i)(y_i - x_i\beta_0) + d_j(y_j - x_j\beta_0)]|w_i\delta_0 = w_j\delta_0\} = 0 \tag{5}$$

Therefore estimation of β_0 can be based on the moment equation (5). An instrumental variables approach is adopted here. This approach could be directly implemented if δ_0 were known and there exist pairs of observations in a random sample with exactly identical indices with positive probability. Nevertheless, given a consistent estimator $\hat{\delta}$ for δ_0, if the nuisance function $E[d_1 u|w\delta_0{=}t]$ is sufficiently smooth, the preceding matching arguments would hold approximately for pairs of observations with approximately similar pairs of $w_i\hat{\delta}$ and $w_j\hat{\delta}$, i.e.$w_i\hat{\delta} \approx w_j\hat{\delta}$.

As several methods exist in the econometric literature for semiparametric estimation of the binary choice model, (see, for example, Cosslett (1983),

Han (1987), Ichimura (1987), Klein and Spady (1993), and Powell, Stock and Stoker (1989)), in this article we assume a consistent estimator of δ_0 exists, and therefore, we will concentrate on the estimation of β_0. Consequently, the estimator $\widehat{\beta}$ of β_0 is defined as a weighted instrumental variables estimator

$$\widehat{\beta} = \widehat{S}_{xx}^{-1}\widehat{S}_{xy} \tag{6}$$

where

$$\widehat{S}_{xx} = \frac{2}{n(n-1)}\sum_{i<j}(1-R_j)(x_i+x_j)'((1-d_i)x_i+d_jx_j)\frac{1}{h}K(\frac{w_i\hat{\delta}-w_j\hat{\delta}}{h})$$

$$\widehat{S}_{xy} = \frac{2}{n(n-1)}\sum_{i<j}(1-R_j)(x_i+x_j)'((1-d_i)y_i+d_jy_j)\frac{1}{h}K(\frac{w_i\hat{\delta}-w_j\hat{\delta}}{h})$$

where the kernel weight $\frac{1}{h}K(\frac{w_i\hat{\delta}-w_j\hat{\delta}}{h})$ gives declining weight to pairs with large values of $\left|w_i\hat{\delta}-w_j\hat{\delta}\right|$.

3. LARGE SAMPLE PROPERTIES OF THE ESTIMATOR

In this subsection we derive the asymptotic properties of the proposed estimator. We begin by making the following assumptions.

ASSUMPTION 1. *The vectors (y_i, x_i, d_i, w_i) generated from (1) and (2) are independent and identically distributed across i, with finite sixth order moments for each component. The randomization indicator R is independent of the other variables. The error term (u, v) is independent of z, with $E(u) = 0$.*

ASSUMPTION 2. *The preliminary estimator $\widehat{\delta}$ of δ_0 is $\sqrt{n}$-consistent, and has the following asymptotic linear representation*

$$\widehat{\delta} = \delta_0 + \frac{1}{n}\sum\psi_i + o_p\left(n^{-1/2}\right)$$

for some $\psi_i = \psi(d_i, w_i)$, such that $E\psi(d_i, w_i) = 0$ and $E\|\psi(d_i, w_i)\|^2 < \infty$.

ASSUMPTION 3. *Define $g_x(u) = E(x_i|w_i\beta_0 = u)$, with $g_w(u)$ similarly defined, then each component of $g_x(u)$ and $g_w(u)$ are four times continuously differentiable.*

Define

$$S_{xx} = E(x_i + E(x_i|w_i\delta_0)'(x_i + E(x_i|w_i\delta_0)p(w_i\delta_0)E(1 - R_i)$$

ASSUMPTION 4. *The matrix S_{xx} is nonsingular.*

ASSUMPTION 5. *The kernel function $K(\cdot)$ has a bounded support. It is four times continuously differentiable and satisfies $\int K(u)\,du = 1$ and $\int u^l K(u)\,du = 0$ for $l = 1, 2, 3$.*

ASSUMPTION 6. *The bandwidth sequence h satisfies $nh^6/\ln(n) \to \infty$, and $nh^8 \to 0$ as $n \to \infty$.*

Assumption 1 describes the model and the data. The independence assumption between (u, v) and x can be relaxed to allow x to include endogenous variables, and the distribution of (u, v) can be allowed to depend on w through the linear index $w\delta_0$. Several estimators for δ_0 mentioned in the previous section (Han (1987), Ichimura (1993), Klein and Spady(1993), and Powell, Stock and Stoker(1989)) satisfy assumption 2. Assumption 4 is an identification condition. Notice that

$$S_{xx} = E(x_i'x_i + 3E(x_i'|w_i\delta_0)E(x_i|w_i\delta_0)p(w_i\delta_0)E(1 - R_i).$$

Assume that the support of $w_i\delta_0$ is the whole line, then the nonsingularity of S_{xx} is implied by the nonsingularity of $Ex_i'x_i$. Assumption 5 is a 'higher' order bias reduction kernel condition, which, together with the rate of convergence condition on the bandwidth sequence in Assumption 6, ensures that the estimator proposed is asymptotically unbiased. Assumption 3 is a boundedness and smoothness conditions, which can be justified by some primitive conditions on the distributions of the variables in the model (see Lee (1994) and Sherman (1994) for some discussions on similar conditions).

Rewriting (6) as

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \widehat{S}_{xx}^{-1}\widehat{S}_{xu}$$

where

$$\widehat{S}_{xu} = \frac{2}{\sqrt{n}(n-1)}\sum_{i<j}(1 - R_j)(x_i + x_j)((1 - d_i)u_i + d_ju_j)\frac{1}{h}K(\frac{w_i\hat{\delta} - w_j\hat{\delta}}{h})$$

we will establish the asymptotic results for $\widehat{\beta}$ in two steps; first, we show that $\widehat{S}_{xx}$ converges in probability to a nonsingular matrix; second, we establish an asymptotic linear representation for $\widehat{S}_{xu}$.

LEMMA 1. *Under Assumptions 1 through 6 above, as* $n \to \infty$, $\widehat{S}_{xx} \xrightarrow{p} S_{xx}$ *where* S_{xx} *is defined as in assumption 3.*

Proof. mimic the proof of Lemma 5.1 of Powell (1989). ∎

By assumption 3, Lemma 1 implies that the matrix inverse in the definition of $\widehat{\beta}$ is well defined in large samples.

Next we consider $\widehat{S}_{xu}$. The following lemma establishes an asymptotic linear representation for $\widehat{S}_{xu}$, Similar to the proof of Theorem 5.1 of Powell (1989) we can establish the follow lemma.

LEMMA 2. *Under conditions 1 through 6, we have*

$$\sqrt{n}\widehat{S}_{xu} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\zeta_i - \Omega\psi_i] + o_p(1)$$

where

$$\zeta_i = [x_i + E(x_i|w_i\delta_0)]u_i^*$$

with $u_i^* = (E(1-R_i))((1-d_i)u_i - \lambda(w_i\delta_0)) + (1-R_i)(d_iu_i + \lambda(w_i\delta_0))p(w_i\delta_0)$, *and*

$$\Omega = 2E(1-R_i)E[\lambda'(w_i\delta_0)E(x|w_i\delta_0)'(w_i - E(w|w_i\delta_0)p(w_i\delta_0)].$$

Combining lemmas 1 and 2, we obtain the main theorem by the central limit theorem.

THEOREM 1. *Under conditions 1-6, the estimator* $\widehat{\beta}$ *is consistent for* β_0, *and asymptotically normal,*

$$\sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma = S_{qq}^{-1}[C_{\zeta\zeta} + \Omega C_{\psi\zeta} + C_{\zeta\psi}\Omega' + \Omega C_{\psi\psi}\Omega']S_{qq}^{-1}$$

for $C_{\zeta\zeta} = E[\zeta_i\zeta_i'], C_{\psi\psi} = E[\psi_i\psi_i']$ *and* $C_{\psi\zeta} = E[\psi_i\zeta_i'] = C_{\zeta\psi}'$.

In order for large-sample inference on β_0 to be carried out using the estimator $\widehat{\beta}$, a consistent estimator of Σ needs to be constructed. In the following one such estimator is proposed following Powell (1989).

Lemma 1 shows $\widehat{S}_{xx}$ is a consistent estimator for S_{xx}. An analogous estimator for Ω is

$$\widehat{\Omega} = \frac{2}{n(n-1)} \sum_{i<j} (1-R_j)((1-d_i)\widehat{u}_i + d_j\widehat{u}_j)\frac{1}{h^2}K'(\frac{w_i\hat{\delta} - w_j\hat{\delta}}{h})(x_i+x_j)'(w_i-w_j)$$

where $\widehat{u}_i = y_i - x_i\hat{\beta}$.

To estimate $C_{\psi\psi}$ and $C_{\psi\zeta}$, it is useful to assume that a sequence of $\{\widehat{\psi}_i\}$ exists which satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \|\widehat{\psi}_i - \psi_i\|^2 = o_p(1).$$

This sequence, of course, depends upon the particular first-step estimator of δ_0; an example of an appropriate sequence $\{\widehat{\psi}_i\}$ for a particular preliminary estimator $\widehat{\delta}$ is given by Powell, Stock and Stoker (1989). Similar sequences can be constructed for the estimators proposed by Ichimura (1993) and Klein and Spady (1993).

As for the sequence $\{\zeta_i\}$, let

$$\widehat{\zeta}_i = \frac{1}{(n-1)} \sum_{j\neq i} (1-R_j)(x_i+x_j)'((1-d_i)\widehat{u}_i + d_j\widehat{u}_j)\frac{1}{h}K(\frac{w_i\widehat{\delta} - w_j\widehat{\delta}}{h}).$$

LEMMA 3. *: Under Assumptions 1-7 above, as* $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} \|\widehat{\zeta}_i - \zeta_i\|^2 = o_p(1).$$

Proof. mimic Lemma 6.2 of Powell (1989). ∎

4. A MONTE CARLO STUDY

In the section we present a small Monte Carlo study to illustrate the usefulness of the proposed estimator. The data is generated according to the following model

$$y = x_1 + x_2 + u$$

$$d = 1\{w_1 + w_2 + v > 0\}$$

and $R = 1\{r > 0\}$, where v has the standard normal distribution. The regressors x_1 and x_2 are draw from a normal $N(0,1)$ distribution and a uniform $U(-2,2)$ distribution, respectively. Different designs are constructed by varying the distributions of the error terms and the structure of (w_1, w_2). In all cases, r is draw from $U(-0.5, 0.5)$ independent of the rest of the variables in the model. $u{=}\sqrt{0.5} * v + \sqrt{0.5} * v^*$ with v^* draw from $N(1,0)$ independent of the other variables in the model. We consider two different designs for the regressors with $(w_1, w_2){=}(x_1, x_2)$ and $w_1 = x_1$and $w_2 = x_2/2{+}x_3/2$ respectively, with x_3 is drawn from a uniform $U(-2,2)$ distribution independent of (x_1, x_2). Data on v are also generated from three different distributions, namely, normality, nonnormality and heteroscedasticity. Consequently we have six designs from these variations of the regressors and error terms.

Here we consider the finite sample performance of our estimator, along with the Heckman's two step estimator; the estimators proposed by Newey (1988) and Powell (1989) are also considered when the exclusion restriction applies. The first-step estimator is chosen to be the probit maximum likelihood estimator. The results from 300 replications from each design are presented with sample size of 100. For each estimator under consideration, we report the mean value (Mean), the standard Deviation (SD), and the root mean square error (RMSE). For Powell's estimator, we use the standard normal density as the kernel function. For the Newey's estimator, the approximating series is . The bandwidth and the number of series are chosen by generalized cross-validation (G. Wahba 1979). We also the standard normal density as the kernel function for our estimator which the bandwidth is chosen by minimizing MM_h where

$$MM_h = |\sum_{i<j}(1{-}R_j)(x_i{+}x_j)((1{-}d_i)(y_i{-}x_i\beta){+}d_j(y_j{-}x_j\beta))\frac{1}{h}K(\frac{w_i\hat{\delta} - w_j\hat{\delta}}{h})|.$$

since our estimated is based on a related moment condition.

Table 1 reports the simulation results for our estimator and Heckman's two-step estimator for the first design where v is a standard normal $N(0,1)$. In the case where $(w_1, w_2){=}(x_1, x_2)$, even though normality is correctly specified, our estimator is superior due to "weak" nonlinearity of the probit function in the relevant range. When there is an exclusion restriction, i.e., $w_1 = x_1$and $w_2 = x_2/2 + x_3/2$, Heckman's estimator is slightly better.

In Table 2 we consider the same design the error term departing from joint normality; $v{=}2v_*^3 + v_*^2 - 1$ where v_* is drawn from a standard normal $N(0,1)$ independent of the other variables. In this case, joint normality is misspecified. Heckman's method produces inconsistent estimates for both designs of the regressors. Our approach still provides reasonably good

TABLE 1.

Results of simulation with normal errors

		$x \neq w$			$x = w$		
Estimator	Coeff.	Mean	SD	RMSE	Mean	SD	RMSE
Heckman	α	0.999	0.172	0.171	0.974	0.307	0.308
	β_1	−1.006	0.075	0.075	−0.981	0.145	0.146
	β_2	0.996	0.137	0.137	0.984	0.218	0.219
Matching	α	0.997	0.093	0.093	0.994	0.097	0.097
	β_1	−1.002	0.099	0.099	−0.987	0.121	0.121
	β_2	1.004	0.112	0.112	1.005	0.115	0.115

TABLE 2.

Results of simulation with nonnormal errors

		$x \neq w$			$x = w$		
Estimator	Coeff.	Mean	SD	RMSE	Mean	SD	RMSE
Heckman	α	0.904	0.597	0.603	0.908	1.212	1.213
	β_1	−1.005	0.118	0.118	−0.972	0.292	0.293
	β_2	0.928	0.373	0.379	0.899	0.749	0.755
Matching	α	0.999	0.154	0.153	1.002	0.132	0.132
	β_1	−0.998	0.152	0.151	−0.999	0.163	0.163
	β_2	0.990	0.170	0.170	0.976	0.204	0.205

estimates. Existence of an exclusion restriction improves the performance of both estimators.

The results with a heteroscedastic error are reported in Table 3. The error term v=exp$(w\delta_0)v_*$, where v_* is a standard normal $N(0,1)$ independent of the other variables. It is obvious Heckman's estimate is biased, while the other three estimators are still consistent. As expected, all the estimators perform better when there is an exclusion restriction.

5. CONCLUSION

In this paper we consider semiparametric estimation of sample selection models using experimental data. We propose a new estimator by matching pairs of observations with offsetting selection bias. which does not rely on the nonlinearity of the selection correction function or some exclusion restrictions. We improve upon Heckman's two-step under parametric setting in that our estimator does not rely on the nonlinearity of the selection correction function when there is no exclusion restriction. We also improve upon the existing semiparametric estimators in that exclusion restriction is not needed for our procedure. The estimator is shown to be consistent and

asymptotically normal under some regularity conditions. A small monte carlo study illustrates the usefulness of the new estimator.

Our estimator is based on access to random samples. Frequently, available data are in the form of choice based samples. It might be possible to modify the current approach to still consistently estimate the outcome equation using choice based samples. It is a topic for future research.

REFERENCES

Andrews, D.W.K. 1991, Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* **59**, 307-345.

Cosslett, S.R., 1983, Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* **51**, 765-782.

Cosslett,S.R.,1991, Seimparametric estimation of a regression model with sample selectivity. In: W.A. Barnett, J.L. Powell, and G.Tauchen, eds, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. (Cambridge University Press, Cambridge).

Donald, S.G. 1995, Two-step estimation of heteroskedastic sample selection models. *Journal of Econometrics* **65**, 347-380.

Fraker, T. and R. Maynard, 1984, *An Assessment of Alternative Comparison Group Methodologies for Evaluating Employment and Training Programs*. Princeton, NJ: MPR, Inc.

Han, A.K., 1987, Nonparametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics* **35**, 303-316.

Heckman, J.J., 1974, Shadow prices, market wages, and labor supply. *Econometrica* **42**, 679-693.

Heckman, J.J., 1976, The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**, 475-492

Heckman, J.J. 1990, Varieties of selection bias. *American Economic Review* **80, 2**, 313-318.

Heckman, J.J., H., Ichimura, J., Smith, and P. Todd, 1998, Characterizing selection bias using experimental data. *Econometrica* **66**, 1017-1098.

Horowitz, J.L., 1992, A smooth maximum score estimator for the binary response model. *Econometrica* **60**, 505-531.

Ichimura, H., 1993, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models 58, 71-120.

Klein, R.W. and R.S. Spady, 1993, An efficient semiparametric estimator of the binary response model. *Econometrica* **61**, 387-421.

Lee, L.F. 1994, Semiparametric instrumental variable estimation of simultaneous equation sample selection models. *Journal of Econometrics* **63**, 341-388.

Leung, S.F. and S.Yu, 1996, On the choice between sample selection and two-part models. *Journal of Econometrics* **72**, 197-229.

LaLonde, R., 1986, Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* **76**, 604-620.

Manski, C.F., 1985, Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. *Journal of Econometrics* **27**, 313-333.

Nawata, K., 1993, A note on the estimation of models with sample-selection biases. *Economics Letters* **42**, 15-24.

Newey, W.K., 1988, Two-step series estimation of sample selection models. *Manuscript*. Department of Economics, Princeton University, Princeton, N.J.

Powell, J.L. 1989, Semiparametric estimation of bivariate latent variable models. *Manuscript*. Social Research Institute, University of Wisconsin, Madison, WI.

Powell, J.L., J.H. Stock, and T.M. Stoker, 1989, Semiparametric estimation of weighted average derivatives. *Econometrica* **57**, 1403-1430.

Sherman, R.P. 1994, U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* **11**, 372-395.

Vella, F. 1995, Estimating models with sample selection bias: A survey. *Manuscript*. Rice University