# SEMIPARAMETRIC ESTIMATION OF HETEROSCEDASTIC BINARY CHOICE SAMPLE SELECTION MODELS UNDER SYMMETRY

By Songnian Chen*
Department of Economics
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong

July 1999

Abstract

Binary choice sample selection models are widely used in applied economics with large cross-sectional data where heteroscedaticity is typically a serious concern. Existing parametric and semiparametric estimators for the binary selection equation and the outcome equation in such models suffer from serious drawbacks in the presence of heteroscedasticity of unknown form in the latent errors. In this paper we propose some new estimators to overcome these drawbacks under a symmetry condition, robust to both nonnormality and general heterscedasticity. The estimators are shown to be $\sqrt{n}$-consistent and asymptotically normal. We also indicate that our approaches may be extended to other important models.

Keywords: Heteroscedasticity, Binary Choice, Sample Selection, Symmetry

* I would like to thank Bo Honoré, Roger Klein, Lung-Fei Lee, Jim Powell, a co-editor, and two anonymous referees for helpful comments.

# 1 Introduction

Sample selection models have received a great deal of attention since the seminal work of Gronau (1974) and Heckman (1974) on female labor supply. They have also found wide application in modelling the impact of unions, occupational choice, schooling, the choice of region of residence and choice of industry, among others. In the female labor supply model, a binary choice selection equation determines whether or not someone works, and then conditional on her working we observe the hours worked. A typical binary choice sample selection model has the form

$$d_i = 1 \quad {}^{*}$$

$$ {}^{*}_{2i} \qquad (1.2)$$

$i = 1, 2, \ldots, n$, where the latent variables $y^{*}_{1i}$ and $y^{*}_{2i}$ are defined by

$$y^{*}_{1i} = x_i^{\prime}\gamma_0 + v_{1i} \qquad (1.3)$$

$$y^{*}_{2i} = x_i^{\prime}\beta_0 + v_{2i} \qquad (1.4)$$

Equation (1.1) is the binary choice selection equation, and (1.2) corresponds to the outcome equation. In this model, $d$ and $y$ are observable dependent variables, $x \in R^q$ is a vector of exogenous variables, and $(v_1, v_2)$ is a vector of latent error terms. The parameters of interest are $\beta_0$ and $\gamma_0$. If the distribution of $(v_1, v_2)$ conditional on $x$ is known up to a set of finite parameters, $\beta_0$ and $\gamma_0$ can then be estimated by maximum likelihood (Amemiya (1985)), and $\gamma_0$ can also be estimated by a computationally simpler two-step approach by Heckman (1974). However, these likelihood-based approaches typically yield inconsistent estimators if either the parametric form of the error distribution is misspecified or if conditional heteroscedasticity of the error terms given the exogenous variables is not correctly modelled parametrically. Such parametric specifications cannot, in general, be justified by economic theory. This fact has motivated the recent interest in semiparametric methods, which do not require parametric specification of error distribution and/or functional form of heteroscedasticity. While departure from normality has serious consequences for commonly used parametric estimators, there is evidence suggesting that these estimators are more severely affected by heteroscedasticity of unknown form than by nonnormality, and furthermore, semiparametric estimators requiring homoscedasticity also behave badly in the presence of unknown form of heteroscedasticity (see, e.g., Donald (1995), Horowitz (1992), Klein and Spady (1993), and Powell (1986)). Therefore, it is extremely desirable to develop semiparametric estimators that are not only robust to nonnormality, but also to general heteroscedasticity, because the binary choice sample selection model is widely used with large cross sectional data, and thus is often plagued with heteroscedasticity. In the past two decades, a large number of semiparametric approaches have been developed for the binary choice sample selection model; there are, however, serious drawbacks associated with the existing approaches. In this paper, we propose some new semiparametric estimators to overcome these drawbacks. Specifically, we consider $\sqrt{n}$-consistent estimation of both the binary choice selection equation and the outcome equation under a conditional symmetry restriction, allowing for general form of unknown heteroscedasticity and nonnormality.

Many semiparametric estimators for $\gamma_0$ in the binary choice model have been proposed in the literature under various weak distributional restrictions. The most common weak restrictions are the independence (and index) restriction, the conditional mean, median and symmetry restrictions. The approaches by Cosslett (1983), Han (1987), and Sherman (1993)) under the independence restriction require homoscedasticity. The index (or monotonic index) restriction needed for the estimators by Ahn et al. (1996), Cavanagh and Sherman (1998), Härdle and Stoker (1989), Horowitz and Härdle (1996), Klein and Spady (1993), and Powell et al. (1989) only allows for very limited form of unknown heteroscedasticity. Since no location restriction is imposed under the independence and index restrictions, the intercept term is not estimated. Under a conditional median restriction, where very general form of unknown heteroscedasticity is allowed, Manski (1985) and Horowitz (1992) proposed maximum score and smoothed maximum score estimators, respectively. However, these two estimators converge at rates slower than $\sqrt{n}$. In fact, Chamberlain (1986) showed that no $\sqrt{n}$-consistent estimator exists under the their assumptions. By extending Chamberlain (1986), Zheng (1995) showed that $\sqrt{n}$-consistent estimation is not possible even if the conditional median restriction is strengthened to the conditional symmetry restriction. More recently, Chen and Khan (1999) showed that Chamberlain's result still holds even under normality when arbitrary form of heteroscedasticity is allowed, which, in turn, suggests that certain restrictions on the type of unknown heteroscedasticity is necessary for $\sqrt{n}$-consistent estimation of the binary choice model. While popular in linear and nonlinear regression analysis, the conditional mean restriction has rarely been used in the analysis of discrete choice models; Horowitz (1993) and Manski (1988) illustrated the difficulty for identification under the conditional mean restriction, and presented a nonidentification result. Recently, however, based on an integration-by-parts argument, Lewbel (1998b) proposed a $\sqrt{n}$-consistent estimator for $\gamma_0$ under a conditional mean restriction and a mild exclusion restriction on heteroscedasticity; like other existing estimators, however, there are several serious drawbacks (to be discussed in detail below) with Lewbel's approach. In this paper we propose a $\sqrt{n}$-consistent estimator for $\gamma_0$ to overcome these drawbacks by strengthening the conditional mean and median restrictions to a conditional symmetry restriction, while allowing for more general form of heteroscedasticity than that of Lewbel (1998b).

Following Heckman's two-step approach in a parametric setting, several semiparametric two-step estimators have been proposed for the estimation of the outcome equation. The approaches by Andrews (1991), Cosslett (1991), Newey (1988b), and Powell (1989) require the independence or index restriction, thus only allowing for very limited form of unknown heteroscedasticity. Another major drawback with these estimators is that the cross-equation exclusion restriction that some regressor in the selection equation is excluded from the outcome equation, is required for model identification. In general, this type of cross-equation exclusion restriction cannot be justified by economic theory. In addition, the intercept term in the outcome equation, a parameter of great importance itself (see, e.g., Andrews and Schafgans (1998) and Heckman (1990)), is not estimated in these approaches. By relying on "identification at infinity" (see, e.g., Chamberlain (1986) and Heckman (1990)), Andrews and Schafgans (1998) considered estimating the intercept term, but their estimator converges at a rate slower than $\sqrt{n}$. Recently, by imposing an index and symmetry restriction, Chen (1999b) considered $\sqrt{n}$-consistent estimation of both the intercept and slope parameters without the cross-equation exclusion restriction, but the approach rules out general form of unknown heteroscedasticiy. Maintaining the normality assumption on the error distribution, Donald (1995) recently proposed a two-step estimator allowing for general form of heteroscedasticity; consequently, Donald's approach is susceptible to inconsistency due to nonnormality. In addition, his method does not take into account the available parametric structure in the binary selection equation in estimating the outcome equation, which, in turn, would adversely affect the performance of the resulting estimator. In this paper we propose a $\sqrt{n}$-consistent estimator for both the intercept and slope parameters by only imposing a joint symmetry assumption, which relaxes the normality assumption of Donald (1995). In addition, our approach allows for even more general form of heteroscedasticity. Furthermore, unlike Donald (1995), the parametric structure in the binary selection equation will be explicitly accounted for in estimating the outcome equation; exploitation of such parametric structure would be particularly important when heteroscedasticity is only related to a small set of exogenous variables compared with the total number of exogenous variables in the selection equation, in which case our approach will be much less susceptible to the "curse of dimensionality". Like Chen (1999b) and Donald (1995), no cross-equation exclusion restr

$$\gamma_0 + v_{1i} > 0\} \tag{2.1}$$

$$y_i = d_i(x_i'\beta_0 + v_{2i}) \tag{2.2}$$

$i = 1, 2, \ldots, n$. For the binary choice model, it is by now well known that some scale normalization is needed to identify $\gamma_0$. Let $x = (x_0, \tilde{x}')'$, where $x_0$ is the first component of $x$. We require that the conditional on $\tilde{x}$, $x_0$ has everywhere positive density with respect to Lebesgue measure. The scale normalization is achieved by setting the first component of $\gamma_0$ to one, thus $\gamma_0 = (1, \tilde{\theta}_0')'$ (see, e.g., Cosslett (1983), Horowitz (1992), Ichimura (1993), and Manski (1985) ). In this paper we consider $\sqrt{n}$-consistent estimation of both the binary choice selection equation and the outcome equation under a joint symmetry restriction, allowing for a general form of unknown heteroscedasticity. Specifically, we assume that the distribution of the error term $(v_1, v_2)$ depends on $x$ only through $(z^2, x_2)$ and symmetric around the origin; namely, $f(v_1, v_2|x) = f(v_1, v_2|z^2, x_2)$ and $f(v_1, v_2|x) = f(-v_1, -v_2|x)$, where $z = x'\gamma_0$, and $x_2$ is subvector of $\tilde{x} = (x_1', x_2')'$ such that $\quad{}^1$ and $x_2 \in R^{q_2}$ with $q_1 + q_2 + 1 = q$.

The heteroscedasticity assumption made here is quite general. It allows an index restriction that the error distribution depends on $x$ only through $(x'\gamma_0)^2$, as in Prais (1953), Prais and Houthakker (1955), and Theil (1951),; more significantly, arbitrary form of dependence on $x_2$ can be accommodated. An important special case is when $f(v_1, v_2|x) = f(v_1, v_2|x_2)$, which amounts to a mild exclusion restriction on heteroscedasticity. Chamberlain (1992), Donald (1995), Fishe et al. (1979), Goldfeld and Quandt (1965), Greene (1994), Lewbel (1998a,1998b), Maddala and Nelson (1975), and Powell (1994), among others, have adopted a similar exclusion restriction. This type of exclusion restriction often arises in many economic applications; for example, in studies of firm profits, the dominant variable affecting heteroscedasticity is typically assumed to be firm size, while in the studies of family expenditures, heteroscedasticity is often related to family income only. As suggested in Lewbel (1998a,1998b), the error distribution in consumer demand model should be independent of those variables determined from the supply side of the economy, such as prices, thus such variables would be excluded from heteroscedasticity. A popular form of heteroscedasticity is commonly introduced by relating the conditional distribution of error term to a vector of exogenous variables $x_h$ (see, e.g., Amemiya (1977), Breusch and Pagan (1979), Davidson and MacKinnon (1984), Goldfeld and Quandt (1972), Harvey (1976), Kmenta (1971), and Rutemiller and Bowers (1968)), among others, and the exclusion restriction on heteroscedasticity follows readily when $x_h$ is a proper subset[1] vector of $x$. In addition, models in which $x_2$ has random coefficients are also included in our setting.

Write the binary choice selection equation as

$$d_i = 1\{x_i'\gamma_0 + v_{1i} > 0\} = 1\{x_{0i} + x_{1i}'\theta_{10} + x_{2i}'\theta_{20} + v_{1i}\}$$

Thus, given our heteroscedasticity assumption, it is obvious that certain location restriction on

[1]Our approach can be easily modified to deal with the more general case in which $x_h$ also contains components excluded from $x$.

the error distribution is required to identify $\nu_{20}$. Manski (1985) and Horowitz (1992) proposed maximum score and smoothed maximum score estimators under a conditional median restriction, respectively; their estimators, however, converge at rates slower than $\sqrt{n}$. Recently, based on an integration-by-parts argument, Lewbel (1998b) considered $\sqrt{n}$-consistent estimation under a conditional mean restriction by imposing a mild exclusion restriction on the form of heteroscedasticity; however, Lewbel's approach suffers from several serious drawbacks. First, his approach relies on a fragile identification condition related to the tail behavior of regressors that requires very strong boundary conditions on the regressors relative to that of the error term $v_1$; in particular, his procedure rules out the probit and logit models with bounded regressors. Second, Lewbel (1998b) deals with the case with $q_1 = 0$ and involves $(q - 1)$-dimensional nonparametric smoothing. In contrast, our approach below deals with the more general case with $q_1 \geq 0$ and only needs $q_2$-dimensional nonparametric smoothing. Therefore our procedure is less susceptible to the curse of dimensionality, especially when $q_2$ is small, as in Goldfeld and Quandt (1965), Greene (1994), Kmenta (1971), Maddala and Nelson (1975), and Park (1966), among others. Here we strengthen the conditional median and mean restrictions to the conditional symmetry restriction to overcome the shortcomings mentioned above. In fact, we impose a symmetry restriction on the joint conditional distribution of $(v_1; v_2)$ to consider $\sqrt{n}$-consistent estimation of the binary choice equation as well as the outcome equation. This symmetry restriction relaxes the normality assumption imposed by Donald (1995). It is worth pointing out that we allow for a more general form of heteroscedasticity than that of Donald (1995) and Lewbel (1998b). In addition, by taking into account the linear structure in the selection equation, our estimator for the outcome equation would be much less susceptible to the "curse of dimensionality" than that of Donald (1995) when $q_2$ is much smaller than $q$. As a central tendency measure, the symmetry restriction has been widely used as a common shape restriction on the error distribution. (see, e.g, Chen (1998b, 1999a,b,c), Cosslett (1987), Honoré et.al (1997), Lee (1996), Linton (1993), Manski (1988), Newey (1988a, 1991), and Powell (1986)). As indicated below, the full symmetry can be relaxed to some extent. Also, there is some evidence (see, e.g., Powell (1986), Honoré et al. (1997)) that symmetry-based estimators possess certain robustness to violations of the symmetry assumption.

To motivate our estimator for $\gamma_0$, we first consider the case with homoscedasticity. Under the condition that $v_1$ is independent of $x$, for a pair of observations $(i; j)$, $i \neq j$, Han (1987) established the following rank condition

$$E(d_i - d_j | x_i; x_j) > 0 \quad \text{if and only if} \quad (x_i - x_j)'\gamma_0 > 0 \tag{2.3}$$

to estimate the slope parameter, and Chen (1998b) used the following rank condition

$$E(d_i + d_j | x_i; x_j) > 1 \quad \text{if and only if} \quad (x_i + x_j)'\gamma_0 > 0 \tag{2.4}$$

to estimate the intercept term ce and symmetry restriction. Notice that these two rank conditions have their own advantages and disadvantages. Equation (2.3) can only identify the slope parameter, whereas Equation (2.4) can be used to identify both the slope and intercept terms. On the other hand, reasonably accurate estimation of the relevant parameter n (2.3) and (2.4) requires that there are a large portion of observations for which $(x_i - x_j)'\gamma_0$ and $(x_i + x_j)'\gamma_0$ lie in the neighborhood of 0, respectively; in general, the former is easier to satisfy than the latter. We now extend these rank conditions to the heteroscedastic case. We assume the symmetry and heteroscedasticity assumption made above. In addition, we assume that $F(z|z^2; x_2) = E(d|x)$ is a strictly increasing function[2] of $z$ for every $x_2$. Consider a pair of observation $(i; j)$, $i \neq j$, such that $x_{2i} = x_{2j}$. Then similar to (2.3) and (2.4), we can show that

$$E(d_i - d_j | x_i; x_j; x_{2i} = x_{2j}) = F(z_i | z_i^2; x_{2i}) - F(z_j | z_j^2; x_{2j}) > 0 \quad (2.5)$$

if and only if $x_{0i} + x_{1i}'\gamma_{10} > x_{0j} + x_{1j}'\gamma_{10}$, and

$$E(d_i + d_j | x_i; x_j; x_{2i} = x_{2j}) = F(z_i | z_i^2; x_{2i}) + F(z_j | z_j^2; x_{2j}) > 1 \quad (2.6)$$

if and only if $(x_i + x_j)\gamma_0 > 0$. Similar to the comparison between (2.3) and (2.4), (2.5) and (2.6) each has its own weakness and strength; (2.5) can only be used to identify and estimate $\gamma_{10}$, whereas (2.6) can be used to identify and estimate the whole vector $\gamma_0$. On the other hand, for accurate estimation based on (2.5) and (2.6), it is essential to have a large portion of observations for which $x_{0i} - x_{0j} + (x_{1i} - x_{1j})'\gamma_{10}$ and $(x_i + x_j)\gamma_0$ lie in the neighborhood of 0, respectively. Typically the latter is more difficult to satisfy. Our estimator is defined by combining both (2.5) and (2.6) in order to exploit the strength in each rank condition.

For the special case when $x_2$ is discrete, and $P(x_{2i} = x_{2j}) > 0$, then following Abrevaya (1999), Cavanagh and Sherman (1998), Chen (1998a,b), Han (1987), Horowitz (1992), Manski (1985), and Sherman (1993), among other, we can estimate $\gamma_0$ by maximizing $H_n^*(\gamma)$ with respect to $\gamma$, where

$$H_n^*(\gamma) = H_{1n}^*(\gamma_1) + H_{2n}^*(\gamma)$$

$$H_{1n}^*(\gamma_1) = \frac{1}{n(n-1)} \sum_{i \neq j} h_1^*(w_i; w_j; \gamma_1)$$

and.

$$H_{2n}^*(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} 1\{x_{2i} = x_{2j}\} h_2^*(w_i; w_j; \gamma)]$$

with

$$h_1^*(w_i; w_j; \gamma_1) = [(d_i - d_j)[2 \cdot 1\{(x_{i0} - x_{j0}) + (x_{i1} - x_{j1})'\gamma_1 > 0\} - 1]$$

and

$$h_2^*(w_i; w_j; \gamma) = [(d_i + d_j - 1)[2 \cdot 1\{(x_i + x_j)'\gamma > 0\} - 1]$$

[2] This monotonicity condition can be relaxed if a semiparametric likelihood approach, such as that of Chen (1999c) and Klein and Spady (1993), is adopted; it is currently being investigated separately.

for $w_i = (d_i, x_i)$, $w_j = (d_j, x_j)$, $i, j = 1, 2, \ldots, n$. Let $x = (x^{d\prime}, x^{c\prime})'$ where $x^d$ and $x^c$ represent the vectors of discrete and continuous components respectively. Obviously, the above approach does not work when $x$ contains continuous components. To allow for continuous elements in $x_2$, similar to Honoré and Kyriazidou (1998), we modify the objective function by replacing the indicator function $1\{x^c_{2i} = x^c_{2j}\}$ with kernel weights, which give increasingly large weights to pairs of observations for which $x^c_{2i}$ and $x^c_{2j}$ are close. Specifically, we have the following modified objective function

$$H_n^{**}(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} 1\{x^d_{2i} = x^d_{2j}\} K_1\left(\frac{x^c_{2i} - x^c_{2j}}{a_1}\right)\{h_1^*(w_i, w_j; \gamma_1) + h_2^*(w_i, w_j; \gamma)\}$$

where $K_1(\cdot)$ is a kernel function specified below, and $a_1$ is a bandwidth sequence converging to zero as $n$ increases. Since heteroscedasticity related to discrete components can be treated as groupwise heteroscedasticity, and is much easier to deal with than its continuous counterpart, we assume $x_2$ only contains continuous components, for notational simplicity. Furthermore, analogous to Horowitz (1992), we consider smoothed versions of the indicator functions $1\{(x_{i0} - x_{j0}) + (x_{i1} - x_{j1})'\gamma_1 > 0\}$ and $1\{(x_i + x_j)'\gamma > 0\}$ for both computational and technical reasons[3]. Finally, we propose to estimate $\gamma_0$, by $\gamma_n = (1, \hat{\nu}_n)$, as a solution to

$$\arg\max_{\gamma \in G} H_n(\gamma) = H_{1n}(\gamma_1) + H_{2n}(\gamma)$$

where $G$ is a subset of $R^q$ specified below,

$$H_{1n}(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} K_1\left(\frac{x_{2i} - x_{2j}}{a_1}\right) h_1(w_i, w_j; \gamma_1)$$

and

$$H_{2n}(\gamma) = \frac{1}{n(n-1)} \sum_{i \neq j} K_1\left(\frac{x_{2i} - x_{2j}}{a_1}\right) h_2(w_i, w_j; \gamma)$$

with

$$h_1(w_i, w_j; \gamma_1) = (d_i - d_j)\left[2L\left(\frac{(x_{i0} - x_{j0}) + (x_{i1} - x_{j1})'\gamma_1}{a_2}\right) - 1\right]$$

and

$$h_2(w_i, w_j; \gamma) = (d_i + d_j - 1)\left[2L\left(\frac{(x_i + x_j)'\gamma}{a_2}\right) - 1\right]$$

$L(\cdot)$ is a cumulative distribution function, and $a_2$ is a bandwidth sequence converging to zero as $n$ increases.

[3] By employing this smoothing scheme, one requires less stringent assumptions on the smoothness of the distribution of $x_2$.

We now turn to the estimation of the outcome equation. To motivate our estimator for $\beta_0$, we write the outcome equation as

$$y_i = d_i x_i'\beta_0 + \lambda(z_i, x_{2i}) + \xi_i \tag{2.7}$$

where $\lambda(z, x_2) = E(dv_2|x) = E(dv_2|z, x_2)$ is the selection bias term, and $E(\xi|x) = 0$ by construction. Equation (2.7) has a partial linear structure as in Engle et al. (1986), Powell (1989) and Robinson (1988). Taking conditional expectation of both sides of Equation (2.7) leads to

$$E(y|z_i, x_{2i}) = E(dx'|z_i, x_{2i})\beta_0 + \lambda(z_i, x_{2i}) \tag{2.8}$$

Substracting (2.7) from (2.8) yields

$$y_i - E(y|z_i, x_{2i}) = (d_i x_i' - E(dx'|z_i, x_{2i}))\beta_0 + \xi_i \tag{2.9}$$

It might appear that (2.9) can be used to estimate $\beta_0$ as in Powell (1989) and Robinson (1988). There are, however, three major drawbacks in estimate $\beta_0$ based on (2.9). First, the components in $\beta_0$ corresponding $x_2$ cannot be identified. Second, the intercept term is not identified either. Third, a cross-equation exclusion restriction is necessary. Instead, we will exploit the symmetry restriction to overcome these drawbacks. Chen (1999b) recently has shown that under homoscedasticity and symmetry $\lambda_{hm}(z) = E(dv_2|x) = E(dv_2|z) = \lambda_{hm}(-z)$; in particular, $\lambda_{hm}(z) = c\phi(z/\sigma)$ under normality, where $\phi(\cdot)$ is the standard normal density function, $\sigma$ is the standard deviation of $v_1$, and $c$ is a constant. We can easily show that $\lambda(-z, x_2) = \lambda(z, x_2)$ in our current heteroscedastic setting. Thus, we obtain

$$E(y_i|-z_i, x_{2i}) = E(d_i x_i'|-z_i, x_{2i})\beta_0 + \lambda(z_i, x_{2i}) \tag{2.10}$$

Substracting (2.10) from (2.7) yields

$$y_i - E(y_i|-z_i, x_{2i}) = [d_i x_i' - E(d_i x_i'|-z_i, x_{2i})]\beta_0 + \xi_i \tag{2.11}$$

Notice that

$$E(d_i x_i'|x_i) - E(d_i x_i'|-z_i, x_{2i}) = F(z_i|z_i^2, x_{2i})x_i' - F(-z_i|z_i^2, x_{2i})E(x_i'|-z_i, x_{2i})$$

is, in general, of full rank. In particular,

$$[E(d_i x_i'|x_i) - E(d_i x_i'|-z_i, x_{2i})]\gamma_0 = [F(z_i|z_i^2, x_{2i}) + F(-z_i|z_i^2, x_{2i})]z_i$$

which is nonzero with positive pr ) suggests an instrumental variables approach to estimating $\beta_0$ if the expectation terms were known. An appropriate set of instrumental variables would be

$$\begin{aligned} & E(d_i x_i|x_i) - E(d_i x_i|-z_i, x_{2i}) \\ = \; & E(d_i|z_i, x_{2i})x_i - E(d_i x_i|-z_i, x_{2i}) \\ = \; & E(1 - d_i|-z_i, x_{2i})x_i - E(d_i x_i|-z_i, x_{2i}) \end{aligned} \tag{2.12}$$

We will replace the expectation terms in (2.11) and (2.12) by nonparametric kernel estimates. For technical convenience, we adopt a density weighted version, in the spirit of Powell (1989). Let $p(z; x_2)$ denote the joint density function of $(z_i; x_{2i})$. Define

$$p_n(-\hat{z}_i; x_{2i}) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{a_3 a_4^{q_2}} K_3\left(\frac{\hat{z}_i + \hat{z}_j}{a_3}\right) K_4\left(\frac{x_{2i} - x_{2j}}{a_4}\right)$$

$$E_n(d_i x_i | -\hat{z}_i; x_{2i}) = \frac{\sum_{j \neq i} d_j x_j K_3\left(\frac{\hat{z}_i + \hat{z}_j}{a_3}\right) K_4\left(\frac{x_{2i} - x_{2j}}{a_4}\right)}{\sum_{j \neq i} K_3\left(\frac{\hat{z}_i + \hat{z}_j}{a_3}\right) K_4\left(\frac{x_{2i} - x_{2j}}{a_1}\right)}$$

and

$$E_n(1 - d_i | -\hat{z}_i; x_{2i}) = \frac{\sum_{j \neq i} (1 - d_j) K_3\left(\frac{\hat{z}_i + \hat{z}_j}{a_3}\right) K_4\left(\frac{x_{2i} - x_{2j}}{a_4}\right)}{\sum_{j \neq i} K_3\left(\frac{\hat{z}_i + \hat{z}_j}{a_3}\right) K_4\left(\frac{x_{2i} - x_{2j}}{a_4}\right)}$$

where $K_3(\cdot)$ is a kernel function, and $a_3$ is a bandwidth sequence converging to zero as $n$ increases; $p_n(-\hat{z}_i; x_{2i})$, $E_n(d_i x_i | -\hat{z}_i; x_{2i})$ and $E_n(1 - d_i | -\hat{z}_i; x_{2i})$ are nonparametric estimates of $p(-z_i; x_{2i})$, $E(d_i x_i | -z_i; x_{2i})$, and $E(1 - d_i | -z_i; x$ espectively, with $\hat{z}_i = x_i' \gamma_n$ for $i = 1, 2, \ldots, n$. We are now ready to propose the following estimator for $\beta_0$;

$$\beta_n = \hat{S}_{nxx}^{-1} \hat{S}_{nxy} \tag{2.13}$$

where

$$\hat{S}_{nxx} = \frac{1}{n} \sum_{i=1} [E_n(1 - d_i | -\hat{z}_i; x_{2i}) x_i - E_n(d_i x_i | -\hat{z}_i; x_{2i})][d_i x_i' - E_n(d_i x_i' | -\hat{z}_i; x_{2i})] p_n^2(-\hat{z}_i; x_{2i})$$

and

$$\hat{S}_{nxx} = \frac{1}{n} \sum_{i=1} [E_n(1 - d_i | -\hat{z}_i; x_{2i}) x_i - E_n(d_i x_i | -\hat{z}_i; x_{2i})][y_i - E_n(y_i | -\hat{z}_i; x_{2i})] p_n^2(-\hat{z}_i; x_{2i})$$

Notice that unlike Donald (1995), the linear structure in the latent regression in the selection equation has been taken into account in our approach to estimating the outcome equation.

Remark 1: The proposed estimator for $\gamma_0$ is based on two rank conditions with equal weighting. It is possible to use different weights with possible efficiency gains. Also, notice that the estimation of $\gamma_0$ involves maximizing over a $(q_1 + q_2)$-dimensional parameter space. We could use a computationally simpler two-step method; specifically, we can estimate $\gamma_{10}$ by $\gamma_{a1n}$ which maximizes $H_n(\gamma_1)$ with respect to $\gamma_1$; In the second step, $\gamma_{20}$ can be estimated by maximizing

$$H_{bn}(\gamma_2) = \sum_{i \neq j} K_1\left(\frac{x_{2i} - x_{2j}}{a_1}\right) h_2(w_i, w_j; (1; \gamma_{a1n}'; \gamma_2')')$$

with respect to $\gamma_2$. Consequently, we only need to maximize over $q_1$ and $q_2$ dimensional parameter spaces separately instead of a $(q_1 + q_2)$-dimensional parameter space.

Remark 2: As discussed earlier, we have focused on heteroscedastictiy associated with continuous exogenous variables for notational simplicity. For more general cases, we can use mixed kernels as in Bierens (1987) to deal with heteroscedasticity associated both discrete and continuous variables, and the details will be similar to the continuous case presented here. For the important special case of groupwise heteroscedasticity, however, there exists a computationally more efficient alternative. Suppose heteroscedasticity is related to $x_{hd}$, which has finite support $x_{hd1}; x_{hd2}; \ldots; x_{hdK}$. Let the observations in the sample for which $x_{hd} = x_{hdk}$ are $x_{k_i}$, $i = 1; 2; \ldots; n_k$. For this subsample, define an augmented subsample of size $2n_k$:

$$x^*_{k_i} = x_{k_i} \quad \text{and} \quad d^*_{k_i} = d_{k_i} \quad \text{for} \quad i = 1; 2; \ldots; n_k$$

and

$$x^*_{k_i} = - x_{k_i} \quad \text{and} \quad d^*_{k_i} = 1 - d_{k_i} \quad \text{for} \quad i = n \qquad 2n_k$$

Then we define an estimator $\gamma_{nd}$, which maximizes

$$H_{dn}(\gamma) = \sum_{k=1}^{K} \sum_{k_i < k_j} [(d^*_{k_i} > d^*_{k_j})(x^*_{k_i}\gamma > x^*_{k_j}\gamma) + (d^*_{k_i} < d^*_{k_j})(x^*_{k_i}\gamma < x^*_{k_j}\gamma)] \tag{2.14}$$

Some algebraic manipulation will show that $\gamma_{nd}$ also maximizes $H^*_n(\gamma)$. Direct implementation of the maximization problem (2.14) requires $O(n^2)$ evaluations in each iteration step. However, as pointed out by Cavanagh and Sherman (1998), this maximization problem can be implemented with only $O(n \ln n)$ evaluations in each iteration step, which is computationally much more efficient.

Remark 3: One widely used model in applied economics and statistics is the transformation model in the form

$$\Lambda_0(y) = x'\gamma_0 + v$$

where $y$ is the dependent variable and $x$ is the independent variable, $v$ is the unobservable disturbance term, and $\Lambda_0$ is a strictly increasing unknown transformation function (see Horowitz (1996) for details). Recently, with a first-step estimator for $\gamma_0$ available, various estimators for $\Lambda_0$ have been proposed (Chen (1998a), Horowitz (1996), Klein and Sherman (1998)) without parametric specification for the transformation function or the error distribution since misspecification of either function could lead to inconsistent estimates and invalid inferences. One major drawback associated with these approaches is that they require the error distribution to be independent of $x$. Here we consider the estimation of the transformation under heteroscedasticity. Specifically, we assume that the conditional density of $v$ depends on $x$ only through $x_2$, a proper subset of $x$, and symmetric around the origin; namely, $f(v|x) = f(v|x_2)$ and $f(v|x) = f(-v|x)$. As in the binary choice model above, scale and location normalization is needed for identification. We adopt the same scale normalization as above; location normalization is achieved by setting the intercept to zero.

To estimate $\Lambda_0(y_0)$, $\Lambda_0(\cdot)$ evaluated at $y_0$, define

$$d_{iy_0} = 1\{y_i > y_0\} = \{x_i'\gamma_0 - \Lambda_0(y_0) + v_i > 0\} \tag{2.15}$$

Therefore $-\Lambda_0(y_0)$ becomes the intercept term for the resulting binary choice model in (2.15). Thus $\Lambda_0(y_0)$ can be estimated by the method proposed above. Alternatively, we can adopt a two-step approach; in the first step $\gamma_0$ is estimated by other methods, such as the approach for the ordered response model proposed in Section 4.1 or the estimator by Chen (1999d); in the second step, $\Lambda_0(y_0)$ can be estimated by the method for binary choice model except that the slope parameters are replaced by the estimator in the first stage.

Remark 4: The proposed estimator for $\beta_0$ is based on the whole sample, including the selected subsample with $d = 1$ and the censored subsample with $d = 0$. Estimation could also be based on the selected subsample, as in Andrews (1991), Donald (1995), Heckman (1974), Newey (1988), and Powell (1989). Similar to (2.7) and (2.8), we have, conditional on $d_i = 1$,

$$y_iF(z_i|z_i^2, x_{2i}) = F(z_i|z_i^2, x_{2i})x_i'\beta_0 + \lambda(z_i, x_{2i}) + \xi^* \tag{2.16}$$

with $E(\xi^*|x, d = 1) = 0$, and

$$E(y_i| -z_i, x_{2i}, d_i = 1)F(-z_i|z_i^2, x_{2i}) = F(-z_i|z_i^2, x_{2i})E(x_i'| -z_i, x_{2i}, d_i = 1)\beta_0 + \lambda(z_i, x_{2i}) \tag{2.17}$$

Thus, conditional on $d_i = 1$,

$$\begin{aligned} & y_iF(z_i|z_i^2, x_{2i}) - E(y_i| -z_i, x_{2i}, d_i = 1)F(-z_i|z_i^2, x_{2i}) \\ &= [F(z_i|z_i^2, x_{2i})x_i' - F(-z_i|z_i^2, x_{2i})E(x_i'| -z_i, x_{2i}, d_i = 1)]\beta_0 + \xi_i^* \end{aligned} \tag{2.18}$$

Consequently, $\beta_0$ can also be estimated by an instrumental variables approach based on (2.18).

Remark 5: Powell (1989) suggested that his pairwise difference estimation approach is basically equivalent to that of Robinson (1988). Similarly, we can show that our estimator for $\beta_0$, $\beta_n$, essentially, can be motivated by the following moment condition based on pairwise difference

$$\begin{aligned} & E(y_i - d_ix_i\beta_0 - y_j - d_jx_j\beta_0)|z_i + z_j = 0, x_i, x_j) \\ &= E(d_iv_{2i} - d_jv_{2j}|z_i + z_j = 0, x_i, x_j) \\ &= 0 \end{aligned}$$

since conditional on $(x_i, x_j)$, $d_iv_{2i} - d_jv_{2j}$ is symmetrically distributed given $z_i + z_j = 0$. However, the symmetry restriction actually implies that

$$E[\Xi(y_i - d_ix_i\beta_0 - y_j - d_jx_j\beta_0)|z_i + z_j = 0, x_i, x_j] = 0 \tag{2.19}$$

holds for any even function $\Xi$. Therefore it is possible to improve efficiency by exploiting more moment conditions as in Newey (1988).

# 3 Large Sample Properties

In this section, we consider large sample properties of the estimators proposed in the previous section. We first make the following assumptions.

finite eighth-order moments for each component.

**Assumption 2:** The conditional density of $(v_1; v_2)$ given $x$, $f(v_1; v_2|x)$ depends on $x$ only through $(z^2; x_2)$, and symmetric around the origin; that is $f(v_1; v_2|x) = f(v_1; v_2|z^2; x_2)$, and $f(-v_1; -v_2|x) = f(v_1; v_2|x)$. In addition, $F(z|z^2; x_2)$ is a strictly increasing function of $z$ for every $x_2$.

**Assumption 3:** (a) The support of the distribution of $x$ is not contained in any proper linear subspace of $R^q$. (b) $0 < P(d = 1|x) < 1$ for almost all $x$. (c) The distribution of $x_0$ conditional on $\tilde{x}$ has everywhere positive density with respect to Lebesgue measure.

**Assumption 4:** The first component of $\gamma_0$ is set to one, and $\gamma_0$ is an interior point of a compact set $G$.

**Assumption 5:** The functions $E(x_1|z; x_2)$, $E(x_1 x_1'|z; x_2)$, $p(z; x_2)$, and $F(z|z^2; x_2)$ are $s_1$ times continuously differentiable with respect to $x_2$ and twice continuously differentiable with respect to $z$, these functions and their partial derivatives are dominated by $M_1(z; x_2)$ with $EM_1^4(z; x_2) < \infty$; in addition, $|M_1(z + b_1; x_2 + b_2) - M_1(z; x_2)| < M_2(z; x_2)(|b_1| + \|b_2\|)$ for some function $M_2(z; x_2)$ and $(b_1; b_2')'$ in a small neighborhood of the origin, with $EM_2^4(z; x_2) < \infty$.

**Assumption 6:** The kernel function $K_1(\cdot)$ is of bounded variation with a bounded support, $s_1$ times continuously differentiable and is a $s_1$-th order bias-reducing kernel: $\int K(u)du = 1$, and $\int u_1^{i_1} u_2^{i_2} \cdots u_{q_2}^{i_{q_2}} du_1 du_2 \cdots du_{q_2} = 0$ if $0 < i_1 + i_2 + \cdots + i_{q_2} < s_1$.

**Assumption 7:** The kernel function $L(\cdot)$ a cumulative distribution function, and $l(\cdot) = L'(\cdot)$ is a twice continuously differential symmetric density function.

**Assumption 8:** The bandwidth sequences satisfy $na_1^{q_2/2+\epsilon} a_2^{3/2+\epsilon} \to \infty$; $na_1^{q_2} a_2 \to \infty$, $na_1^{2s_1} \to 0$, and $na_2^4 \to 0$ for a small positive constant $\epsilon$.

Define $Q = Q_1 + Q_2$, where

$$Q_1 = 4E\frac{\partial F(z|z^2; x)}{\partial z} p(z; x_2) S_1$$

and

$$Q_2 = 2E\frac{\partial F(z|z^2; x)}{\partial z} p(-z; x_2)(S_{21} + S_{22})$$

with

$$S_1 = \begin{pmatrix} (x_1 - E(x_1|z; x_2))(x_1 - E(x_1|z; x_2))' & 0 \\ 0 & 0 \end{pmatrix}$$

$$S_{21} = (\tilde{x} + E(\tilde{x}|-z;x_2)(\tilde{x} + E(\tilde{x}|-z;x_2)'$$

and

$$S_{22} = E[(\tilde{x} - E(\tilde{x}|-z;x_2)(\tilde{x}' - E(\tilde{x}'|-z;x_2)|-z;x_2]$$

**Assumption 9**: The matrix Q is positive definite.

Assumption 1 describe the model and the data. The independence assumption could be relaxed as in Andrews (1994) and Whang and Andrews (1993). The existence of higher order moments of x is made mainly to apply Theorem 3 of Sherman (1994). (For details, See the discussion following Theorem 3 in Sherman (1994)).

For the purpose of estimating both the selection equation and the outcome equation, we state a joint symmetry condition in Assumption 2, although only a marginal symmetry condition is needed for estimating the selection equation, and the error term in the outcome equation can be of the form $v_2^* = v_2 + v_2^{**}$, such that $(v_1; v_2)$ satisfies Assumption 2 and $E(v_2^{**}|v_1) = 0$. As pointed out earlier, the monotonicity is needed for the rank-based estimation approach, and may be relaxed for approaches based on semiparametric likelihood (such as Chen (1999c) and Klein and Spady (1993)) or semiparametric least squares (Ichimura and Lee (1991)). Unlike Donald (1995), normality is not required. In addition, our approach allows for more general heteroscedasticity than Donald (1995) and Lewbel (1998a, 1998b).

Assumption 3 is an identification condition. (See Manski (1985), Ichimura (1993), and Horowitz (1992) for related discussions). It implies that x has at least one continuously distributed component, and that this component has unbounded support. However, this assumption of unbounded support can be relaxed following the arguments in Horowitz (1998).

Assumption 4 is standard in the literature. Assumption 5 is a boundedness and smoothness condition. Assumptions 6, 7, and 8 place restrictions on the kernel functions and bandwidth sequences. Notice that the kernel function used for controlling for heteroscedasticity is $q_2$-dimensional. Thus our approach is particularly useful when $q_2$ is small when the problem of the 'curse of dimensionality' is not serious. The matrix Q in Assumption 9 is analogous to the Hessian form of the information matrix in maximum likelihood estimation.

**Theorem 1** Under Assumptions 1-9, $\nu_n$ is consistent and asymptotically normal,

$$\sqrt{n}(\nu_n - \nu_0) \xrightarrow{d} N(0; \Sigma_1)$$

where

$$\Sigma_1 = Q^{-1}\Omega_1 Q^{-1}$$

with

$$\psi_{1i} = \begin{pmatrix} \psi_{11i} \\ 0 \end{pmatrix} + \psi_{12i}$$

$$\psi_{11i} = (d_i - F(z_i|z_i^2; x_{2i}))(x_{1i} - E(x_1|z_i; x_{2i}))p(z_i; x_{2i})$$

$$\psi_{12i} = (d_i - F(z_i|z_i^2; x_{2i}))(\bar{x}_i + E(\bar{x}|-z_i; x_{2i}))p(-z_i; x_{2i})$$

and $\Omega_1 = E\psi_{1i}\psi_{1i}'$

We now turn to the estimation of the outcome equation. The following additional assumptions are made.

Assumption 10: The functions $\lambda(z; x_2)$; $E(x|z; x_2)$, $E(xx'|z; x_2)$, $p(z; x_2)$ and $F(z|z^2; x_2)$ are $s_3$ times continuously differentiable with respect to $x_2$ and twice continuously differentiable with respect to $z$, and these functions and their partial derivatives are dominated by $M_3(z; x_2)$ with $EM_3^6(z; x_2) < \infty$; in addition, $|M_4(z + b_1; x_2 + b_2) - M_1(z; x_2)| < M_4(z; x_2)(|b_1| + \|b_2\|)$ for some function $M_2(z; x_2)$ and $(b_1; b_2')'$ in a small neighborhood of the origin with $EM_2^6(z; x_2) < \infty$.

Assumption 11: The kernel function $K_3(\cdot)$ is $s_3$ times and $K_4(\cdot)$ is twice continuously differentiable with bounded supports; $K_3(\cdot)$ a $s_3$-th order bias-reducing kernel, and $K_4(\cdot)$ a second order bias-reducing kernel.

Assumption 12: The bandwidth sequences satisfy $na_3^{q_2+\varepsilon}a_4^{1+\varepsilon} \to \infty$, $na_4^2 \to \infty$, $na_3^{2s_3} \to 0$ and $na_4^4 \to 0$ for a small positive constant $\varepsilon$ as $n \to \infty$.

Let

$$S_{xx} = E\{[(F(z|z^2; x_2)x - F(-z|z^2; x_2)E(x|-z; x_2))(F(z|z^2; x_2)x - F(-z|z^2; x_2)E(x|-z; x_2))']p^2(-z; x_2)\}$$

Assumption 13: The matrix $S_{xx}$ is positive definite.

Assumption 10 contains some boundedness and smoothness conditions. Assumptions 11 and 12 place restrictions on the kernel functions and bandwidth sequences. Assumption 13 is the main identification condition. The presence of $p(-z; x_2)$ in the definition of $S_{xx}$ implies that accurate estimation would require significant portion of individuals with selection probabilities above as well as below 0.5. This identification condition holds quite generally, even without the cross equation exclusion restriction (see Chen (1999b) for some related discussions). In contrast, the approaches by Andrews (1991), Cosslett (1991), Newey (1988b) and Powell (1989) rely crucially on this exclusion restriction, even though it typically cannot be justified by economic theory. In addition, the identification condition allows $x$ to contain a constant term, thus the intercept term in the outcome equation can be treated equally as the slope parameter, estimatible at the usual parametric rate.

Theorem 2 Under Assumptions 1-13, $\beta_n$ is consistent and asymptotically normal,

$$\sqrt{n}(\beta_n - \beta_0) \xrightarrow{d} N(0; \Sigma_2)$$

where

$$\Sigma_2 = S_{xx}^{-1}\Omega_2 S_{xx}^{-1}$$

$\Omega_2 = E[(\psi_{2i} + S_{x1}\psi_{1i})(\psi_{2i} + S_{x1}\psi_{1i})']$ with

$$S_{x1} = E\Big(\frac{\partial\lambda(-z;x_2)}{\partial z}p^2(-z;x)[F(z|z^2;x_2)x - F(-z|z^2;x_2)E(x|-z;x_2)][\tilde{x}' + E(\tilde{x}'|-z;x_2)]\Big)$$

and

$$\psi_{2i} = (d_i v_{2i} - \lambda(z_i;x_{2i}))F(z_i|z^2;x_{2i})(x_i - E(x|-z_i;x_{2i}))p^2(-z_i;x_{2i})$$

For the purpose of carrying out large sample statistical inferences, consistent estimators of $\Sigma_1$ and $\Sigma_2$ need to be constructed. From the proofs of the theorems, we can see that $-\frac{\partial^2 H_n(\gamma_n)}{\partial\theta\partial\theta'}$ and $S_{nxx}(\gamma_n)$ are consistent for $Q$ and $S_{xx}$ respectively. Define $\hat{v}_{2i} = y_i - d_i x_i \beta_n$ for $i = 1,2,\ldots,n$. Then

$$\hat{S}_{1nxv_2} = \frac{1}{n^2(n-1)}\sum_{j,l\neq i}(d_i\hat{v}_{2i} - d_j\hat{v}_{2j})[(1-d_l)x_i - d_l x_l] \frac{1}{a_3^{2q_2}a_4^2}K_3\left(\frac{x_{2i}-x_{2j}}{a_3}\right)K_3\left(\frac{x_{2i}-x_{2l}}{a_3}\right)K_5(\gamma_n)$$

can be shown to be consistent for $S_{x1}$, where $K_5(\cdot)$ is defined in (A.7) in the Appendix. Next, define

$$\hat{\psi}_{1i} = \begin{pmatrix}\hat{\psi}_{11i}\\ 0\end{pmatrix} + \hat{\psi}_{12i}$$

$$\hat{\psi}_{11i} = \frac{1}{n-1}\sum_{j\neq i}\frac{1}{a_1^{q_2}a_2}K_1\Big(\frac{x_{2i}-x_{2j}}{a_1}\Big)(d_i - d_j)I\Big(\frac{\hat{z}_i+\hat{z}_j}{a_2}\Big)(x_{1i}-x_{1j})$$

$$\hat{\psi}_{12i} = \frac{1}{n-1}\sum_{j\neq i}\frac{1}{a_1^{q_2}a_2}K_1\Big(\frac{x_{2i}-x_{2j}}{a_1}\Big)(d_i + d_j - 1)I\Big(\frac{\hat{z}_i+\hat{z}_j}{a_2}\Big)(\tilde{x}_i+\tilde{x}_j)$$

$$\hat{\psi}_{2i} = \frac{1}{n(n-1)}\sum_{jl}(d_i\hat{v}_{2i} - d_j\hat{v}_{2j})\frac{1}{a_3^{2q_2}a_4^2}K_3\left(\frac{x_{2i}-x_{2j}}{a_3}\right)K_4\left(\frac{\hat{z}_i+\hat{z}_j}{a_4}\right) ((1-d_l)x_i - d_l x_l)K_3\left(\frac{x_{2i}-x_{2l}}{a_3}\right)K_4\left(\frac{\hat{z}_i+\hat{z}_l}{a_4}\right)$$

$$\hat{\Sigma}_1 = \left[-\frac{\partial^2 H_n(\gamma_n)}{\partial\theta\partial\theta'}\right]^{-1}\hat{\Omega}_1\left[-\frac{\partial^2 H_n(\gamma_n)}{\partial\theta\partial\theta'}\right]^{-1}$$

and

$$\hat{\Sigma}_2 = S_{nxx}^{-1}(\gamma_n)\hat{\Omega}_2 S_{nxx}^{-1}(\gamma_n)$$

where

$$\hat{\Omega}_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{\psi}_{1i}\hat{\psi}_{2i}'$$

and

$$\hat{\Omega}_2 = \frac{1}{n}\sum_{i=1}^{n}[(\hat{\chi}_{2i} + \hat{S}_{1nxv_2}\hat{\chi}_{1i})(\hat{\chi}_{2i} + \hat{S}_{1nxv_2}\hat{\chi}_{1i})']$$

Then, following the arguments in the proof of Theorems 1 and 2, we can show that $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are consistent for $\Sigma_1$ and $\Sigma_2$, respectively.

# 4 Extensions

In this section we indicate that our previous approaches can be extended to estimate some other important models, including an ordered response model, a sample selection model with endogenous regressors, a censored nonparametric regression model, and a panel data sample selection model. Full details and regularity conditions will not be given. Similar notations to those in the previous sections will be used without explanation if no confusion arises.

## 4.1 Estimating an Ordered Response Model

The ordered response model has been widely used in applied economics (see Amemiya (1985) and Maddala (1983) for a review). An ordered response model with $K+1$ choices is commonly defined as

$$d_{ik} = 1\{x_i'\gamma_0 + v_i > \alpha_{ok})$$

for $k = 1, 2, \ldots, K$ and $i = 1, 2, \ldots, n$. We assume that the distribution of $v$ given $x$ is symmetric around the origin, and depends on $x$ only through $x_2$, a subset of $x$. The same scale normalization is adopted as in the binary case above. The location normalization is achieved by setting the intercept term in $\gamma_0$ to zero; thus $-\alpha_{ok}$ becomes the new intercept. We consider the estimation of the slope parameter $\gamma_0$ as well as the threshold values $\alpha_{ok}$, $k = 1, 2, \ldots, K$.

To motivate our approach, we first consider some rank conditions related to choices $k_1$ and $k_2$. For a pair of observation $(i, j)$, $i \neq j$, such that $x_{2i} = x_{2j}$, analogous to (2.5) and (2.6), we have

$$E(d_{ik_1} - d_{jk_2}|x_i, x_j, x_{2i} = x_{2j}) = F(x_i'\gamma_0 - \alpha_{ok_1}|x_{2i}) - F(x_j'\gamma_0 - \alpha_{ok_2}|x_{2j}) > 0 \tag{4.1}$$

if and only if $x_{0i} - x_{0j} + (x_{1i} - x_{1j})'\gamma_{10} - (\alpha_{ok_1} - \alpha_{ok_2}) > 0$, and

$$E(d_{ik_1} + d_{jk_2}|x_i, x_j, x_{2i} = x_{2j}) = F(x_i'\gamma_0 - \alpha_{ok_1}|x_{2i}) + F(x_j'\gamma_0 - \alpha_{ok_2}|x_{2j}) > 1 \tag{4.2}$$

if and only if $(x_i + x_j)'\gamma_0 - (\alpha_{k_1} + \alpha_{k_2}) > 0$. Let $\alpha_o = (\alpha_{o1}, \alpha_{o2}, \ldots, \alpha_{oK})'$. Then following the discussions in the binary case, we can estimate $(\gamma_0, \alpha_o)$ by $(\hat{\gamma}, \hat{\alpha}_o)$, which maximizes

$$H_{on}(\gamma, \alpha) = \sum_{k_1, k_2}\sum_{i \neq j} K_1\left(\frac{x_{2i} - x_{2j}}{a_1}\right)[h_{o1}(w_i, w_j, \gamma_1, \alpha_{k_1}, \alpha_{k_2}) + h_{o2}(w_i, w_j, \gamma, \alpha_{k_1}, \alpha_{k_2})]$$

with respect to $(\gamma, \alpha)$, where

$$h_{o1}(w_i, w_j, \gamma_1, \alpha_{k_1}, \alpha_{k_2}) = (d_i - d_j)[2L(\frac{(x_{i0} - x_{j0}) + (x_{i1} - x_{j1})\gamma_1 - (\alpha_{k_1} - \alpha_{k_2})}{a_2}) - 1]$$

and

$$h_{o2}(w_i, w_j, \gamma, \alpha_{k_1}, \alpha_{k_2}) = (d_i + d_j - 1)[2L(\frac{(x_i + x_j)'\gamma - (\alpha_{k_1} + \alpha_{k_2})}{a_2}) - 1]$$

## 4.2 Sample Selection Models with Endogenous Regressors and Heteroscedasticity

For sample selection models we here only focus on endogenous regressors in the binary selection equation since endogenous regressors in the outcome equation can be dealt with by the usual instrumental variables approach to the linear regression model,

To fix ideas, let $x_2$ denote the endogenous regressors with a reduced form $x_2 = \tau(x_u) + e$; where $x_u$ is a vector of exogenous variables, the disturbance term $e$ is allowed to be correlated with $(v_1, v_2)$. Assume that conditional on $x_{hu}$, $(x_0, x_1', x_u')'$ is independent of $(e, v_1, v_2)$. Furthermore, the conditional distribution of $(e, v_1, v_2)$ given $x_{hu}$ is symmetrically distributed around origin. An special case is when $x_{hu}$ is a subvector of $(x_1', x_u')'$. Let $x^* = (x_0, x_1', x_u', x_{hu}')'$. Similar to (2.5) and (2.6), we can show that

$$\begin{aligned} & E(d_i - d_j | x_i^*, x_j^*, x_{hui} = x_{huj}) \\ & = F(x_{i0} + x_{i1}'\gamma_{10} + \tau'(x_{ui})\gamma_{20} | x_{hui}) - F(x_{j0} + x_{j1}'\gamma_{10} + \tau'(x_{uj})\gamma_{20} | x_{huj}) > 0 \end{aligned} \tag{4.3}$$

if and only if

$$x_{i0} + x_{i1}'\gamma_{10} + \tau'(x_{ui})\gamma_{20} > x_{j0} + x_{j1}'\gamma_{10} + \tau'(x_{uj})\gamma_{20}$$

and

$$\begin{aligned} & E(d_i + d_j | x_i^*, x_j^*, x_{hui} = x_{huj}) \\ & = F(x_{i0} + x_{i1}'\gamma_{10} + \tau'(x_{ui})\gamma_{20} | x_{hui}) + F(x_{j0} + x_{j1}'\gamma_{10} + \tau'(x_{uj})\gamma_{20} | x_{huj}) > 1 \end{aligned} \tag{4.4}$$

if and only if

$$x_{i0} + x_{i1}'\gamma_{10} + \tau'(x_{ui})\gamma_{20}) + x_{j0} + x_{j1}'\gamma_{10} + \tau'(x_{uj})\gamma_{20} > 0$$

Then the unknown parameters in the selection equation can be estimated by a solution maximizing

$$\begin{aligned} \sum_{i \neq j} K(\frac{x_{hui} - x_{huj}}{a_1}) f(d_i - d_j)[2L\left(\frac{x_{i0} - x_{j0} + (x_{1i} - x_{1j})'\gamma_1 + (\hat{\tau}(x_{ui}) - \hat{\tau}(x_{uj}))'\gamma_2}{a_2}\right) - 1] \\ +(d_i + d_j - 1)[2L\left(\frac{x_{i0} + x_{j0} + (x_{1i} + x_{1j})'\gamma_1 + (\hat{\tau}(x_{ui}) + \hat{\tau}(x_{uj}))'\gamma_2}{a_2}\right) - 1]\} \end{aligned}$$

with respect to $\gamma$, where $\hat{\tau}(x_u)$ is a first-step nonparametric estimator for $\tau(x_u)$.

For the estimation of the outcome equation, similar to (2.11) we have

$$y_i - E(y_i|z_{ui}; x_{hui}) = [d_i x_i' - E(d_i x_i'|z_{ui}; x_{hui})]\beta_0 + \xi_i \tag{4.5}$$

where $z_{ui} = x_{0i} + x_{1i}'\gamma_{10} + \tau'(x_{ui})\gamma_{20}$, and $E(\xi_i|x^*; x_{hu}) = 0$. Consequently, Equation (4.5) suggests an instrumental variables estimation approach to estimating $\beta_0$ as in Section 2.

## 4.3 Sample Selection Models under Symmetry with a Nonparametric Selection Mechanism

In the literature of semiparametric estimation of sample selection models, most attention has been focused on estimating the parameters in the outcome equation while maintaining a parametric index structure on the binary selection equation. Recognizing that misspecification of the parametric form of the index function results in general in inconsistent estimators for the coefficients in the outcome equation, Ahn and Powell (1993) considered estimation of a sample selection model subject to a nonparametric selection mechanism. However, their approach suffers from the following drawbacks; first, estimation of the intercept term in the outcome equation is not considered in their approach; second, without the cross equation exclusion restriction, identification for the outcome equation will completely rely on the extent of the nonlinearity of the latent regression function in the binary selection equation; furthermore, only limited form of unknown heteroscedasticity is allowed. To overcome these drawbacks, in this section we extend our approach in the previous sections to the case with a nonparametric selection mechanism.

Consider the following sample selection model

$$d_i = 1\{m(x_i) + v_{1i} > 0\} \tag{4.6}$$

$$y_i = d_i x_i \beta + d_i v_{2i} \tag{4.7}$$

where the distribution of $(v_1; v_2)$ given $x$ is symmetric around the origin and depends on $x$ only through a subset $x_2$. Similar to (2.7), we have

$$y_i = d_i x_i' \beta_0 + \lambda(m(x_i); x_{2i}) + \xi_i$$

with

$$\lambda(m(x_i); x_{2i}) = \lambda(-m(x_i); x_{2i})$$

where $\lambda(m(x_i); x_{2i}) = E(d_i v_{2i}|m(x_i); x_{2i})$ and $E(\xi_i|x_i) = 0$. Let

$$P(x_i) = E(d|x_i) = F(m_n(x_i); x_{2i}) = \int_{-\infty}^{m(x_i)} f(v_1|x_{2i})dv_1 \tag{4.8}$$

where $f(v_1|x_{2i})$ is the conditional density of $v_1$ given $x$. Suppose that $f(v_1|x_{2i})$ is always positive, then Equation (4.8) implies that $F(\cdot; x_{2i})$ is invertible for every $x_2$. Thus $m(x_i)$ can be written as $\delta_1(P(x_i); x_{2i})$, which, in fact, has the property

$$-m(x_i) = \delta_1(1 - P(x_i); x_{2i})$$

Hence we have

$$\lambda(m(x_i); x_{2i}) = \lambda(\delta_1(P(x_i); x_{2i}); x_{2i}) = \delta(P(x_i); x_{2i})$$

with $\delta(P(x_i); x_{2i}) = \delta(1 - P(x_i); x_{2i})$. Consequently, we have

$$y_i = d_i x_i' \beta_0 + \delta(P(x_i); x_{2i}) + \xi_i$$

and

$$E(y_i|1 - P(x_i); x_{2i}) = E(d_i x_i'|1 - P(x_i); x_{2i})\beta_0 + \delta(P(x_i); x_{2i})$$

Thus

$$y_i - E(y|1 - P(x_i); x_{2i}) = (d_i x_i' - E(d_i x_i'|1 - P(x_i); x_{2i}))\beta_0 + \xi_i \tag{4.9}$$

Therefore, similar to the approach in Section 2 an instrumental variables estimator for $\beta_0$ can be proposed based on Equation (4.9) by replacing the expectation terms and the selection probabilities by some nonparametric estimates.

## 4.4 Nonparametric estimation of a censored regression model

The binary choice sample selection model reduces to the censored regression model when the two latent regression functions coincide. Parametric and semiparametric estimation of the censored regression model has received a great deal of attention in the literature. Due to the sensitivity of the existing parametric and semiparametric estimators to misspecification of the functional form of the latent regression function, it is of interest to consider nonparametric estimation of the censored regression model.

Consider the censored regression model

$$y = \max\{m(x) + v; 0\}$$

Nonparametric estimation of the censored regression model has been considered by Fan and Gijbels (1996) and Chaudhuri (1991) based on nonparametric quantile regression (typically, median regression). The median regression, however, can only estimate $m(x)$ at points where the censoring is less than fifty percent. Recently, Lewbel and Linton (1998) considered the estimation of the derivatives of the regression function through solving a partial differential equation system. In this section, we consider nonparametric estimation of the latent regression function under the condition that the conditional distribution of $v$ given $x$, depends on $x$ only through $x_2$, a subset of $x$, and symmetric around the origin.

To motivate our approach, Let $d = 1\{y > 0\}$. Following the discussions in Section 4.3, we have

$$\begin{aligned} E(dy|x) &= P(x)m(x) + \lambda(m(x); x_2) \\ &= P(x)m(x) + \delta(P(x); x_2) \end{aligned}$$

where $P(x) = E(d|x)$, $\lambda(m(x); x_2) = \lambda(-m(x); x)$ and $\delta(P(x); x_2) = \delta(1 - P(x); x_2)$. Consequently, we have

$$E(dy|P(x); x_2) = P(x)m(x) + \delta(P(x); x_2) \tag{4.10}$$

and

$$\begin{aligned} E(dy|1 - P(x); x_2) &= (1 - P(x))(-m(x)) + \delta(1 - P(x); x_2) \\ &= (1 - P(x))(-m(x)) + \delta(P(x); x_2) \end{aligned} \tag{4.11}$$

Then subtracting (4.10) from (4.11) gives

$$m(x) = E(dy|P(x); x_2) - E(dy|1 - P(x); x_2) \tag{4.12}$$

Consequently, estimation of $m(x)$ can be based on Equation (4.12) by replacing the unknown expectation terms by nonparametric estimates.

## 4.5 A Panel Data Sample Selection Model under Symmetry

Consider the following panel data sample selection model

$$y_{it} = d_{it}(x_{it}\beta_p + \alpha_{fi} + u_{it}) \tag{4.13}$$

$$d_{it} = 1\{g^*(x_{it}) + \eta_i + v^*_{it} > 0\} \tag{4.14}$$

where

$$\eta_i = \rho(x_{fi}) + v^{**}_i \tag{4.15}$$

for $i = 1, ..., n$ and $t = 1, 2$, $\beta_p$ is the parameter of interest, $\alpha_{fi}$ and $\eta_i$ are unobservable time invariant individual specific effects. We assume that the individual specific effect $\eta_i$ in the selection equation has the weak functional restriction $\eta_i = \rho(x_{fi}) + v^{**}_i$, where $\rho(\cdot)$ is an unknown function, $x_f$ represents a set of variables entering into individual effects, as determined by decision process of individual economic agents. Let $x_i = (x'_{i1}, x'_{i2}, x'_{fi}, x_{hfi})'$. We assume that the conditional distribution of the error term $(v^*_{it}, u_{it}, v^{**}_i)$ depends on $x_i$ only through $x_{hfi}$, and symmetric around the origin. Thus we allow for heteroscedasticity over time for each panel member and across individuals.

By specifying $g^*(x_{it}) = x'_{it}\gamma_p$, Chamberlain (1993) has shown that $\sqrt{n}$-consistent estimation of $\gamma_p$ is not possible if no restriction is imposed on the fixed effects $\eta_i$ unless $v^*_{it}$ has logistic distribution given $(x_{it}, \eta_i)$. Newey (1994) considered $\sqrt{n}$-consistent estimation of $\gamma_p$ by imposing (4.15) and normality. We extend Newey's model by relaxing the normality and the linear specification for $g^*(x_{it})$; in addition, we allow for heteroscedasticity across individuals. Note that Newey's focus is on the binary selection equation, whereas we are concerned with the estimation of the outcome equation. The individual effect $\alpha_i^*$ in the outcome equation is left unspecified.

Semiparametric estimation of panel data sample selection models has been considered by Kyriazidou (1997). By leaving the fixed effects terms unspecified, Kyirazidou (1997) considered estimating the parameter vector $\beta_p$ under a conditional exchangeability condition. However, Kyriazidou's (1997) approach requires a linear specification for $g^*(x_{it})$, a cross-equation exclusion restriction and homoscedasticity over time for each panel member; furthermore, her estimator converges at a rate slower than $\sqrt{n}$.

We now extend our estimation approach to the cross sectional case to the panel data model .To fix ideas, let $g_{it} = \rho(x_{fi}) + g^*(x_{it})$; for $i = 1, 2, \ldots, n$, $t = 1, 2$. Let $D_i = d_{1i}d_{2i}$; $\Delta u_i = u_{i1} - u_{i2}$; $\Delta y_i = y_{i1} - y_{i2}$ $\Delta x_i = x_{i1} - x_{i2}$ and $\delta_{\Delta u}(g_{i1}, g_{i2}, x_{hfi}) = E(D\Delta u|x_i)$. First differencing yields

$$D_i\Delta y_i = D_i\Delta x_i\beta_p + \delta_{\Delta u}(g_{i1}, g_{i2}, x_{hfi}) + \Delta u_i^*$$

where $\Delta u_i^* = D_i\Delta u_i - \delta_{\Delta u}(g_{i1}, g_{i2}, x_{hfi})$ such that $E(\Delta u_i^*|x_i) = 0$. Let $v_{it} = v_{it}^* + v_i^{**}$. Note that in our setting $(\Delta u_i, v_{i1}, v_{i2})$ is jointly symmetrically distributed conditional on $x_{hf}$, then we have

$$\begin{aligned} & \delta_{\Delta u}(g_{i1}, g_{i2}, x_{hf}) + \delta_{\Delta u}(-g_{i1}, -g_{i2}, x_{hf}) - \delta_{\Delta u}(-g_{i1}, g_{i2}, x_{hf}) - \delta_{\Delta u}(g_{i1}, -g_{i2}, x_{hfi}) \\ = \; & E[\Delta u_i 1\{-|g_{i1}| < v_{i1} < |g_{i1}|, -|g_{i2}| < v_{i2} < |g_{i2}||x_i] \\ = \; & 0 \end{aligned}$$

Thus,

$$\delta_{\Delta u}(g_{i1}, g_{i2}, x_{hfi}) = \delta_{\Delta u}(-g_{i1}, g_{i2}, x_{hfi}) + \delta_{\Delta u}(g_{i1}, -g_{i2}, x_{hfi}) - \delta_{\Delta u}(-g_{i1}, -g_{i2}, x_{hfi}) \tag{4.16}$$

Therefore, analogous to Equation (4.9), we obtain

$$D_i\Delta y_i - \delta_{\Delta y}^*(g_{i1}, g_{i2}, x_{hfi}) = (D_i\Delta x_i - \delta_{\Delta x}^*(g_{i1}, g_{i2}, x_{hfi}))\beta_p + \Delta u_i^* \tag{4.17}$$

where

$$\delta_{\Delta y}^*(g_{i1}, g_{i2}, x_{hfi}) = \delta_{\Delta y}(-g_{i1}, g_{i2}, x_{hfi}) + \delta_{\Delta y}(g_{i1}, -g_{i2}, x_{hfi}) - \delta_{\Delta y}(-g_{i1}, -g_{i2}, x_{hfi})$$

and

$$\delta_{\Delta x}^*(g_{i1}, g_{i2}, x_{hfi}) = \delta_{\Delta x}(-g_{i1}, g_{i2}, x_{hfi}) + \delta_{\Delta x}(g_{i1}, -g_{i2}, x_{hfi}) - \delta_{\Delta x}(-g_{i1}, -g_{i2}, x_{hfi})$$

with $\delta_{\Delta x}(g_{i1}, g_{i2}, x_{hfi}) = E(D\Delta x|g_{i1}, g_{i2}, x_{hfi})$ and $\delta_{\Delta y}(g_{i1}, g_{i2}, x_{hfi}) = E(D\Delta y|g_{i1}, g_{i2}, x_{hfi})$. For the selection probabilities $p_1 = E(d_1|x)$ and $p_2 = E(d_2|x)$, let $\lambda_{\Delta y}(p_1, p_2, x_{hf}) = E(D\Delta y|p_1, p_2, x_{hf})$,

$\lambda_{\cdot x}(p_1, p_2, x_{\langle f\rangle}) = E(D \cdot x | p_1, p_2, x_{\langle f\rangle})$. Then, with a similar invertibilitiy condition, arguing as in the cross sectional case, we have

$$\lambda_{\cdot u}(p_{i1}, p_{i2}, x_{\langle f\rangle}) = \lambda_{\cdot u}(1 - p_{i1}, p_{i2}, x_{\langle f\rangle}) + \lambda_{\cdot u}(p_{i1}, 1 - p_{i2}, x_{\langle f\rangle}) - \lambda_{\cdot u}(1 - p_{i1}, 1 - p_{i2}, x_{\langle f\rangle}) \quad (4.18)$$

Equation (4.17) can then be reformulated as

$$D_i \cdot y_i - \lambda^*_{\cdot y}(p_{i1}, p_{i2}, x_{\langle f\rangle}) = (D_i \cdot x_i - \lambda^*_{\cdot x}(p_{i1}, p_{i2}, x_{\langle f\rangle}))\beta_p + \cdot u_i^* \quad (4.19)$$

where

$$\lambda^*_{\cdot y}(p_{i1}, p_{i2}, x_{\langle f\rangle}) = \lambda_{\cdot y}(1 - p_{i1}, p_{i2}, x_{\langle f\rangle}) + \lambda_{\cdot y}(p_{i1}, 1 - p_{i2}, x_{\langle f\rangle}) - \lambda_{\cdot y}(1 - p_{i1}, 1 - p_{i2}, x_{\langle f\rangle})$$

and

$$\lambda^*_{\cdot x}(p_{i1}, p_{i2}, x_{\langle f\rangle}) = \lambda_{\cdot x}(1 - p_{i1}, p_{i2}, x_{\langle f\rangle}) + \lambda_{\cdot x}(p_{i1}, 1 - p_{i2}, x_{\langle f\rangle}) - \lambda_{\cdot x}(1 - p_{i1}, 1 - p_{i2}, x_{\langle f\rangle})$$

Finally, Equation (4.19) suggests an instrumental variables approach to estimating $\beta_0$.

## 5 Conclusion

In this paper we have considered semiparametric estimation of the binary choice sample selection model under a symmetry restriction, allowing for a very general form of unknown heterocedasticity. Our procedure estimates the intercept and slope parameters in the binary choice selection equation and the outcome regression equation. The estimators are $\sqrt{n}$-consistent and asymptotically normal. Our approach overcomes various serious drawbacks associated with existing estimators for the binary choice selection equation and the outcome equation. As indicated earlier, the full symmetry assumption used here could be relaxed to some extent. Also, we could test the validity of the symmetry by following the arguments of Zheng (1998).

From both theoretical and practical point of view, it is desirable to have efficient estimators for parameters of interest. Like most existing procedures, our method is a two-step procedure. Typically efficient estimation calls for a joint estimation of the binary selection equation and the outcome equation. Recently, Ai (1997) and Chen and Lee (1998) proposed joint estimation procedures, and their estimators achieve Chamberlain's (1986) semiparametric efficient bound under homoscedasticity. It is possible to derive the relevant semiparametric efficient bound in our heteroscedastic setting by following Cosslett (1987) and Chamberlain (1986). Furthermore, it is likely that the efficient procedures by Ai (1997) and Chen and Lee (1998) can be extended to the heteroscedastic case. This is an important topic for future research.

# Appendix

Proof of Theorem 1: Recall $\gamma_n$ maximizes

$$H_n(\gamma) = H_{1n}(\gamma_1) + H_{2n}(\gamma)$$

We prove the consistency by showing that (a) there exists a function $H(\gamma)$ such that $|H_n(\gamma) - H(\gamma)| \to 0$ in probability uniformly in $\gamma \in G$; and (b) $H(\gamma)$ is continuous and has a unique maximum at $\gamma_0$ (Amemiya, (1985)).

Since the treatment of $H_{1n}(\gamma_1)$ will be similar to that of $H_{2n}(\gamma)$, we will only provide detailed analysis for the latter. Notice that $\{H_{2n}(\gamma); \gamma \in G\}$ is a second order U-process. For the random sample $\{w_1, w_2, \ldots, w_n\}$, let $P_n$ denote the empirical measure that places $1/n$ on each $w_i$ and $U_n$ the random measure putting mass $1/n(n-1)$ on each ordered pair $(w_i, w_j)$. Then as in Arcones and Giné (1993) and Sherman (1993), we have the following decomposition

$$H_{2n}(\gamma) = Eh_2(w_i, w_j; \gamma) + P_n h_{21}(\cdot, \gamma) + U_n h_{22}(\cdot, \cdot, \gamma) \tag{A.1}$$

with

$$h_{21}(w; \gamma) = 2(Eh_2(w, w_i; \gamma) - Eh_2(w_i, w_j; \gamma))$$

and

$$h_{22}(w^1, w^2; \gamma) = h_2(w^1, w^2; \gamma) - Eh_2(w^1, w_i; \gamma) - Eh_2(w_i, w^2; \gamma) + Eh_2(w_i, w_j; \gamma)$$

where $h_{21}(w; \gamma)$ and $h_{22}(w^1, w^2; \gamma)$ are first and second order degenerate U-statistics respectively (see, e.g., Arcones and Giné (1993) and Sherman (1993)).

We now analyze the individual terms in (A.1). Define classes of functions $\mathcal{F}_{n1} = \{h_{21}(\cdot, \gamma): \gamma \in G\}$ and $\mathcal{F}_{n2} = \{\tilde{h}_{22}(\cdot, \gamma): \gamma \in G\}$ where $\tilde{h}_{22}(w^1, w^2; \gamma) = a_1^{q_2} h_{22}(w^1, w^2; \gamma)$. Then similar to Lemma 10A in Sherman (1994), we can show that $\mathcal{F}_{n1}$ and $\mathcal{F}_{n2}$ are Euclidean with a square integrable envelop function. Then by Theorem 3 of Sherman (1993)

$$P_n h_{21}(\cdot, \gamma) = O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$U_n h_{22}(\cdot, \cdot, \gamma) = \frac{1}{a_1^{q_2}} U_n \tilde{h}_{22}(\cdot, \cdot, \gamma) = O_p\left(\frac{1}{n a_1^{q_2/2+\epsilon}}\right)$$

uniformly in $\gamma \in G$ for any small positive $\epsilon$. Therefore

$$H_{2n}(\gamma) = Eh_2(w_i, w_j; \gamma) + o_p(1)$$

uniformly in $\gamma \in G$.

We now analyze $Eh_2(w_i, w_j; \gamma)$ For notational simplicity, we only treat the case $E(d|x) = F(z|z^2, x_2) = F(z|x_2)$,

$$Eh_2(w_i, w_j; \gamma) = \int\int \frac{1}{a_1^{q_2}} K_1\left(\frac{x_{2i} - x_{2j}}{a_1}\right) (F(x_i'\gamma_0|x_{2i}) + F(x_j'\gamma_0|x_{2j}) - 1)$$

$$
\begin{aligned}
&\left[2L\left(\frac{(x_i + x_j)'\gamma}{a_2}\right) - 1\right] p(x_{i0}, x_{i1}, x_{i2}) p(x_{j0}, x_{j1}, x_{j2}) dx_i dx_j \\
= &\int\int K_1(u)(F(x_i'\gamma_0 | x_{2i}) + F(x_{j0} + x_{j1}'\gamma_{10} + (x_{i2} - a_1 u)'\gamma_{20} | x_{i2} - a_1 u) - 1) \\
&\left[2L\left(\frac{x_{i0} + x_{j0} + (x_{i1} + x_{j1})'\gamma_1 + (2x_{i2} - a_1 u)'\gamma_2}{a_2}\right) - 1\right] \\
&p(x_{i0}, x_{i1}, x_{i2}) p(x_{j0}, x_{j1}, x_{i2} - a_1 u) dx_i dx_{j0} dx_{j1} du \\
= &H_2(\gamma) + o_p(1) \qquad \text{(A.2)}
\end{aligned}
$$

by the change of variable $x_{2j} = x_{2i} - a_1 u$ and the dominated convergence theorem, where

$$
H_2(\gamma) = \int F^{ij}[2 \cdot 1\{x_{i0} + x_{j0} + (x_{i1} + x_{j1})'\gamma_1 + 2x_{i2}'\gamma_2 > 0\} - 1] p(x_{i0}, x_{i1}, x_{i2}) p(x_{j0}, x_{j1}, x_{i2}) dx_i dx_{j0} dx_{j1}
$$

with

$$
F^{ij} = (F(x_i\gamma_0 | x_{2i}) + F(x_{j0} + x_{j1}'\gamma_{10} + x_{i2}'\gamma_{20} | x_{i2}) - 1)
$$

Notice that $F^{ij} > 0$ if and only if

$$
x_{i0} + x_{j0} + (x_{i1} + x_{j1})'\gamma_{10} + 2x_{i2}'\gamma_{20} > 0
$$

So

$$
H_2(\gamma_0) = \int\int\int |F^{ij}| p(x_{i0}, x_{i1}, x_{i2}) p(x_{j0}, x_{j1}, x_{i2}) dx_i dx_{j0} dx_{j1}
$$

Thus $H_2(\gamma)$ reaches maximum at $\gamma_0$.

If fact we can show that $\gamma_0$ is the unique maximum of $H_2(\gamma)$. Following the arguments in Lemma 3 of Manski (1985), we have for any $\gamma$,

$$
H_2(\gamma_0) - H_2(\gamma) > 0
$$

if

$$
\int\int\int 1\{(x_{i1} + x_{j1})'(\gamma_1 - \gamma_{10}) + 2x_{i2}'(\gamma_2 - \gamma_{20}) \neq 0\} p(x_{i0}, x_{i1}, x_{i2}) p(x_{j0}, x_{j1}, x_{i2}) dx_i dx_{j0} dx_{j1} > 0
$$

which, in turn, holds if

$$
P\{(x_{i1} + x_1)'(\gamma_1 - \gamma_{10}) + 2x_{i2}'(\gamma_2 - \gamma_{20}) = 0\} < 1 \qquad \text{(A.3)}
$$

where, conditional on $x_{i2}$, $(x_{i0}, x_{i1})$ and $(x_0, x_1)$ are independent and identically distributed with the conditional distribution being $p(x_{i0}, x_{i1} | x_{i2})$. By assumption 3, there exist a positive $\epsilon_1$ such that either

$$
P\{x_{i1}'(\gamma_1 - \gamma_{10}) + x_{i2}'(\gamma_2 - \gamma_{20}) > \epsilon_1\} > 0
$$

and

$$P\{x_{i1}'(\gamma_1 - \gamma_{10}) + x_{i2}'(\gamma_2 - \gamma_{20}) > \epsilon_1\} > 0$$

or

$$P\{x_{i1}'(\gamma_1 - \gamma_{10}) + x_{i2}'(\gamma_2 - \gamma_{20}) < -\epsilon_1\} > 0$$

and

$$P\{x_{i1}'(\gamma_1 - \gamma_{10}) + x_{i2}'(\gamma_2 - \gamma_{20}) < -\epsilon_1\} > 0$$

Thus, (A.3) follows readily. As a result, we have shown that (1) $|H_{2n}(\gamma) - H_2(\gamma)| \to 0$ in probability uniformly in $\gamma \in G$; and (2) $H_2(\gamma)$ is continuous and has a unique maximum at $\gamma_0$. Similarly, we can show that there exists a function $H_1(\gamma_1)$ such that (1) $|H_{1n}(\gamma_1) - H_1(\gamma_1)| \to 0$ in probability uniformly in $(1, \gamma_1', \gamma_{20}')' \in G$; and (2) $H_1(\gamma_1)$ is continuous and has a unique maximum at $\gamma_{10}$. Consequently, the consistency of $\hat{\gamma}_n$ follows by combining the above results.

We now prove the asymptotic normality. Since $\gamma_0$ is an interior point of the compact set G, thus $\gamma_n$ satisfies

$$\frac{\partial H_n(\gamma_n)}{\partial \hat{\gamma}} = 0$$

with probability close to one when n increases. A Taylor expansion yields

$$0 = \frac{\partial H_n(\gamma_0)}{\partial \hat{\gamma}} + \frac{\partial^2 H_n(\bar{\gamma}_n)}{\partial \hat{\gamma} \partial \hat{\gamma}'}(\hat{\gamma}_n - \hat{\gamma}_0)$$

where $\bar{\gamma}_n = (1, \bar{\gamma}_{1n}', \bar{\gamma}_{2n}')'$ lies between $\gamma_n$ and $\gamma_0$. Therefore

$$\sqrt{n}(\hat{\gamma}_n - \hat{\gamma}_0) = \left[ - \frac{\partial^2 H_n(\bar{\gamma}_n)}{\partial \hat{\gamma} \partial \hat{\gamma}'} \right]^{-1} \sqrt{n} \frac{\partial H_n(\gamma_0)}{\partial \hat{\gamma}}$$

We first consider

$$- \frac{\partial^2 H_n(\bar{\gamma}_n)}{\partial \hat{\gamma} \partial \hat{\gamma}'} = \begin{pmatrix} - \frac{\partial^2 H_{1n}(\bar{\gamma}_{1n})}{\partial \hat{\gamma}_1 \partial \hat{\gamma}_1'} & 0 \\ 0 & 0 \end{pmatrix} + \left[ - \frac{\partial^2 H_{2n}(\bar{\gamma}_n)}{\partial \hat{\gamma} \partial \hat{\gamma}'} \right]$$

Note that

$$- \frac{\partial^2 H_{2n}(\gamma)}{\partial \hat{\gamma} \partial \hat{\gamma}'} = \frac{1}{n(n-1)} \sum_{i \neq j} h_3(w_i, w_j; \gamma)$$

where

$$h_3(w_i, w_j; \gamma) = - \frac{2}{a_1^2 a_2^2} K_1\left(\frac{x_{2i} - x_{2j}}{a_1}\right)(d_i + d_j - 1) l'\left(\frac{(x_i + x_j)'\gamma}{a_2}\right)(\tilde{x}_i + \tilde{x}_j)(\tilde{x}_i + \tilde{x}_j)'$$

Similar to the proof of (A.2), we can show that

$$- \frac{\partial^2 H_{2n}(\gamma)}{\partial \hat{\gamma} \partial \hat{\gamma}'} = E h_3(w_i, w_j; \gamma) + o_p(1) = E h_3(w_i, w_j; \gamma_0) + o_p(1)$$

uniformly for $\gamma$ in a $o(1)$ neighborhood of $\gamma_0$ if $na_1^{q_2/2+\varepsilon}a_2^{3/2+\varepsilon} \to \infty$ for a small positive number $\varepsilon$. Thus, from the consistency of $\hat\theta_n$, we obtain

$$-\frac{\partial^2 H_{2n}(\hat\theta_n)}{\partial\theta\partial\theta'} = Eh_3(w_i; w_j; \gamma_0) + o_p(1)$$

With some algebraic manipulation, we can show that

$$\begin{aligned}
& Eh_3(w_i; w_j; \gamma_0) \\
= \;& 2\int\int [E(\xi\xi'|-z; x_2) + E(\xi|-z; x_2)E(\xi'|z; x_2) + E(\xi|z; x_2)E(\xi'|-z; x_2) + E(\xi\xi'|z; x_2)] \\
& \frac{\partial F(z|z^2; x_2)}{\partial z} p(-z; x_2)p(z; x_2)dz dx_2 + o(1) \\
= \;& 2E[\frac{\partial F(z|z^2; x_2)}{\partial z} p(-z; x_2)[E(\xi\xi'|-z; x_2) + E(\xi|-z; x_2)\xi' + \xi E(\xi'|-z; x_2) + \xi\xi'] + o(1) \\
= \;& Q_2 + o(1) \qquad\qquad (A.4)
\end{aligned}$$

Similarly, we can show that

$$\begin{pmatrix} -\frac{\partial^2 H_{1n}(\hat\theta_{n1})}{\partial\theta_1\partial\theta_1'} & 0 \\ 0 & 0 \end{pmatrix} = Q_1 + o_p(1)$$

Therefore we have shown

$$-\frac{\partial^2 H_n(\hat\theta_n)}{\partial\theta\partial\theta'} = Q + o(1) \qquad\qquad (A.5)$$

Next, we consider $\sqrt{n}\frac{\partial H_n(\gamma_0)}{\partial\gamma}$. Notice that

$$\sqrt{n}\frac{\partial H_{2n}(\gamma_0)}{\partial\theta} = \frac{2}{\sqrt{n}(n-1)}\sum_{i\neq j}\frac{1}{a_1^{q_2}a_2}K_1\left(\frac{x_{2i}-x_{2j}}{a_1}\right)(d_i + d_j - 1)I\left(\frac{(x_i + x_j)'\gamma_0}{a_2}\right)(\xi_i + \xi_j)$$

Similar to Powell et. al (1989), we can show that

$$\begin{aligned}
\sqrt{n}\frac{\partial H_{2n}(\gamma_0)}{\partial\theta} &= \frac{4}{\sqrt{n}}\sum_{i=1}(d_i + F(-z_i|z_i^2; x_{2i}) - 1)(\xi_i + E(\xi|-z_i; x_{2i}))p(-z_i; x_{2i}) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}\psi_{12i} + o_p(1)
\end{aligned}$$

Analogously, we have

$$\sqrt{n}\frac{\partial H_{1n}(\gamma_0)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}\psi_{11i} + o_p(1)$$

Hence

$$\sqrt{n}\frac{\partial H_n(\gamma_0)}{\partial \theta} = \frac{1}{\sqrt{n}}\sum_{i=1} \psi_{1i} + o_p(1) \tag{A.6}$$

Theorem 1 follows readily from the asymptotic linear representation in (A.6).

Proof of Theorem 2: Write $\beta_n$ as

$$\beta_n = S_{nxx}^{-1}(\gamma_n)S_{nxy}(\gamma_n)$$

where

$$S_{nxx}(\gamma) = \frac{1}{n^2(n-1)}\sum_{j,l\neq i} h_{xx}(w_i^*, w_j^*, w_l^*; \gamma)$$

and

$$S_{nxy}(\gamma) = \frac{1}{n^2(n-1)}\sum_{j,l\neq i} h_{xy}(w_i^*, w_j^*, w_l^*; \gamma)$$

with

$$\begin{aligned} h_{xx}(w_i^*, w_j^*, w_l^*; \gamma) &= ((1-d_l)x_i - d_l x_l)\frac{1}{a_3^{2q_2}a_4^2}K_3\left(\frac{x_{2i}-x_{2j}}{a_3}\right)K_4\left(\frac{x_i\gamma + x_j\gamma}{a_4}\right) \\ &\quad (d_i x_i - d_j x_j)' K_3\left(\frac{x_{2i}-x_{2l}}{a_3}\right)K_4\left(\frac{x_i\gamma + x_l\gamma}{a_4}\right) \end{aligned}$$

and

$$\begin{aligned} h_{xy}(w_i^*, w_j^*, w_l^*; \gamma) &= (y_i - y_j)\frac{1}{a_3^{2q_2}a_4^2}K_3\left(\frac{x_{2i}-x_{2j}}{a_3}\right)K_4\left(\frac{x_i\gamma + x_j\gamma}{a_4}\right) \\ &\quad ((1-d_l)x_i - d_l x_l)K_3\left(\frac{x_{2i}-x_{2l}}{a_3}\right)K_4\left(\frac{x_i\gamma + x_l\gamma}{a_4}\right) \end{aligned}$$

for $w_i^* = (d_i, x_i, y_i)$, $w_j^* = (d_j, x_j, y_j)$ and $w_l^* = (d_l, x_l, y_l)$, $i, j, l = 1, 2, \ldots, n$. Therefore, we have

$$\sqrt{n}(\beta_n - \beta_0) = S_{nxx}^{-1}(\gamma_n)\sqrt{n}S_{nxv_2}(\gamma_n)$$

where

$$S_{nxv_2}(\gamma) = \frac{1}{n^2(n-1)}\sum_{j,l\neq i} h_{xv_2}(w_i^*, w_j^*, w_l^*; \gamma)$$

with

$$\begin{aligned} h_{xv_2}(w_i^*, w_j^*, w_l^*; \gamma) &= (d_i v_{2i} - d_j v_{2j})\frac{1}{a_3^{2q_2}a_4^2}K_3\left(\frac{x_{2i}-x_{2j}}{a_3}\right)K_4\left(\frac{x_i\gamma + x_j\gamma}{a_4}\right) \\ &\quad ((1-d_l)x_i - d_l x_l)K_3\left(\frac{x_{2i}-x_{2l}}{a_3}\right)K_4\left(\frac{x_i\gamma + x_l\gamma}{a_4}\right) \end{aligned}$$

For $S_{nxx}(\gamma_n)$, similar to the proof of (A.5), we can show that

$$S_{nxx}(\gamma_n) = S_{xx} + o_p(1)$$

For $S_{nxv_2}(\gamma_n)$, a Taylor expansion yields

$$S_{nxv_2}(\gamma_n) = S_{nxv_2}(\gamma_0) + S_{1nxv_2}(\bar{\gamma}^*_n)(\gamma_n - \gamma_0)$$

where $\bar{\gamma}^*_n$ lies between $\gamma_n$ and $\gamma_0$,

$$\begin{aligned} S_{1nxv_2}(\gamma) &= \frac{1}{n^2(n-1)} \sum_{j,l\neq i} (d_i v_{2i} - d_j v_{2j})[(1-d_l)x_i - d_l x_l] \\ &\quad \frac{1}{a_3^{2q_2} a_4^2} K_3\left(\frac{x_{2i} - x_{2j}}{a_3}\right) K_3\left(\frac{x_{2i} - x_{2l}}{a_3}\right) K_5(\gamma) \end{aligned}$$

with

$$\begin{aligned} K_5(\gamma) &= K_4'\left(\frac{x_i\gamma + x_j\gamma}{a_4}\right) K_4\left(\frac{x_i\gamma + x_l\gamma}{a_4}\right)\left(\frac{x_i + x_j}{a_4}\right)' \\ &\quad + K_4\left(\frac{x_i\gamma + x_j\gamma}{a_4}\right) K_4'\left(\frac{x_i\gamma + x_l\gamma}{a_4}\right)\left(\frac{x_i + x_l}{a_4}\right)' \end{aligned} \qquad \text{(A.7)}$$

Again, similar to (A.5), we can show that

$$\begin{aligned} & S_{nx1}(\bar{\gamma}^*_n) \\ =\ & -E\left(\frac{\partial\lambda(z;x_2)}{\partial z} p^2(-z;x)[F(z|z^2;x_2)x - F(-z|z^2;x_2)E(x|-z;x_2)][x' + E(x'|-z;x_2)]\right) + o_p(1) \\ =\ & S_{x1} + o_p(1) \end{aligned}$$

Finally, we consider $S_{nxv_2}(\gamma_0)$. Following Arcones and Giné (1993) and Sherman (1993), we have the following decomposition for $S_{nxv_2}(\gamma_0)$,

$$\begin{aligned} \sqrt{n} S_{nxv_2}(\gamma_0) &= \frac{\sqrt{n}}{n(n-1)(n-2)} \sum_{i\neq j\neq l} h_{xv_2}(w_i^*, w_j^*, w_l^*; \gamma_0) + o_p(1) \\ &= \sqrt{n} E h_{xv_2}(w_i^*, w_j^*, w_l^*; \gamma_0) + S_{n1} + U_{n2} + U_{n3} + o_p(1) \\ &= S_{n1} + U_{n2} + U_{n3} + o_p(1) \end{aligned}$$

where

$$S_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1} E[h_{xv_2}(w_i^*, w_j^*, w_l^*; \gamma_0)|w_i^*]$$

and $U_{n2}$ and $U_{n3}$ are second and third order degenerate U-statistics, respectively. Similar to the proof of Powell et. al (1989), we can show that

$$S_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1} \psi_{2i} + o_p(1)$$

$$EU_{n2}^2 = o(1)$$

and

$$EU_{n3}^2 = o(1)$$

From the above results, we obtain

$$\sqrt{n}(\beta_n - \beta_0) = S_{xx}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}(\psi_{2i} + S_{x1}\psi_{1i}) + o_p(1)$$

Then, Theorem 2 follows readily from the asymptotic linear representation by applying the Lingerbeger-Levy Central Limit Theorem.

# References


[1] Abrevaya, J. (1999): \Semiparametric Estimation of the Generalized Linear Panel Model," g in Journal of Econometrics.

[2] Ai, C. (1997): \A Semiparametric Maximum Likelihood Estimator," Econometrica, 65, 933-963.

[3] Ahn, H., H. Ichimura and J.L. Powell (1996): \A Simple uscript.

[4] Ahn H. and J.L. Powell. (1993): \Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," Journal of Econometrics, 58, 3-29.

[5] Amemiya, T. (1977): \A note on Heteroscedastic Model," Journal of Econometrics, 6, 365-370.

[6] Amemiya, T. (1985): Advanced Econometrics, Cambridge, MA: Harvard University Press.

[7] Andrews, D.W.K. (1991): \Asymptotic Normality for Series Estimator for Various Nonparametric and Semiparametric Models," Econometrica 59, 307-345.

[8] Andrews, D.W.K. (1994): \Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," Econometrica 62, 43-72.

[9] Andrews, D.W.K. and M. Schafgans (1998): \Semiparametric Estimation of the Intercept of a Sample Selection Model," Review of Economics Studies 65, 497-517.

[10] Ancones, M.A. and E. Gin¶ (1993): \Limit Theorems for U-Process," Annals of Probability, 21, 1494-1592.

[11] Bierens, H.J. (1987): \Kernel Estimators of Regression Function," in Advances in Econometrics, Cambridge University Press.

[12] Breusch, T.S., and A.R. Pagan (1979): \A Simple Test for Heteroscedasticity and Random Coe± cient Variation," Econometrica 47, 1287-1294.

[13] Cavanagh, C. and R.P. Sherman (1998): \Rank Estimators for Monotonic Index Models", Journal of Econometrics, 86, 351-382.

[14] Chamberlai E± ciency s, 32, 189-218.

[15] Chamberlain, G. (1992): \E± ency Bounds for Semiparametric Regression," Econometrica, 60, 567-596.

[16] Chamberlain, G. (1993): \Feedback in Panel Data Models," Manuscript.

[17] Chaudhuri, P. (1991): \Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," Annals of Statistics, 19, 760-777.

[18] Chen, S. (1998a): \Rank Estimation of Transformation Models", Manuscript.

[19] Chen, S. (1998b): \Rank Estimation of a Location Parameter in a Binary Choice Model", Manuscript

[20] Chen, S. and L.F. Lee (1998): \E± cient Semiparametric Scoring Estimation of Sample Selection Models", Econometric Theory, 14, 423-462.

[21] Chen, S. (1999a): \Semiparametric Estimation of a Location Parameter in a Binary Choice Model," Econometric Theory, 15, 79-98.

[22] Chen, S. (1999b): \Distribution-Free Estimation of the Random Coe± cient Dummy Endogenous Variable Model," Journal of Econometrics, 91, 171-199.

[23] Chen, S. (1999c): \E± cient Estimation of Binary Choice Models under Symmetry", forthcoming in Journal of Econometrics.

[24] Chen, S. (1999d): \Semiparametric Estimation of Transformation Models," Work in Progress.

[25] Chen, S. and S. Khan (1999): \$\sqrt{n}$-Consistent Estimation of Order Response Models under Multiplicative Heteroscedasticity.," Manuscript.

[26] Cosslett, S.R. (1983): \Distribution ator of the Binary Choice Model," Econometrica, 51, 765-782.

[27] Cosslett, S.R. (1987): \E± ciency Bounds for Distribution-Free Estimators of the BInary Choice and Censored Regression Models," Econometrica, 55, 559-586.

[28] Cosslett, S.R. (1991): \Distribution-Free Estimator of a Regression Model with Sample Selectivity," in Nonparametric and Semiparametric Methods in Econometrics and Statistics, ed. by W.A. Barnett, J.L. Powell and G. Tauchen. Cambridge: Cambridge University Press.

[29] Davidson R., and J. MacKinnon (1984): \Convenient Speci¯cation Tests for Logit and Probit Models," Journal of Econometrics, 25, 241-262.

[30] Donald, S.G. (1995): \Two-step Estimation of Heteroskedastic Sample Selection Models," Journal of Econometrics, 65, 347-380.

[31] Engle, R.F., C.W.J. Granger, J. Rice, and A. Weiss (1986): \Semiparametric Estimates of the Relation between Weather and Electricity Demand," Journal of the American Statistical Association, 76, 817-823.

[32] Fan, J. and I. Gijbels (1996): Local Polynomial Modelling and Its Applications, Chapman and Hall.

[33] Fishe, R.P.H., G.S. Maddala, and R.P. Trost (1979): \Estimation of a Heteroscedastic Tobit Model," Manuscript.

[34] Goldf nd R.E. Quandt (1965): \Some Tests for Homoscedasticity," Journal of the American Statistical Association, 60, 539-547.

[35] Goldfeld, S.M., and R.E. Quandt (1972), Nonlinear Methods in Econometrics, Amsterdam: North-Holland.

[36] Greene, W.H. (1994), Econometric Analysis, rentice-Hall, Inc.

[37] Gronau, R. (1974): \Wage Comparisons | A Selectivity Bias," Journal of Political Economy, 82, 1119-1144.

[38] Han, A.K. (1987): \Nonparametric Analysis of a Generalized Regression Model: the Maximum Rank Correlation Estimator," Journal of Econometrics, 35, 303-316.

[39] Härdle, W. and T.M. Stoker (1989): \Investigating Smooth Multiple Regression by the Method of Average Derivative," Journal of the American Statistical Association, 84, 986-995.

[40] Harvey, A.C. (1976): \Estimating icity," Econometrica, 44, 461-465.

[41] Heckman, J.J. (1974): \Shadow Prices, Market Wages and Labor Supply," Econometrica, 42, 679-693.

[42] Heckman, J.J. (1990): \Varieties of selection bias," American Economic Review, 80, 2, 313-318.

[43] Honoré, B., E. Kyriazidou, and C. Udry (1997): \Estimation of Type 3 Tobit Models Using Symmetric and Pairwise Comparisons," Journal of Econometrics, 64, 241-278.

[44] Honoré, B. and E. Kyriazidou (1998): \Panel Data Discrete Choice Models with Lagged dependent Variables," Manuscript.

[45] Horowitz, J.L. (1992): \A Smoothed Maximum Score Estimator for the Binary Response Model," Econometrica, 60, 505-531.

[46] Horowitz, J.L. (1993): \Semiparametric and Nonparametric Estimation of Quantal Response Models," in Handbook of Statistics, Vol 11, G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., Amsterdam: North-Holland Publishing Company.

[47] Horowitz, J.L. (1996): \Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable ," Econometrica, 64, 103-137.

[48] Horowitz, J.L. and W. Härdle (1996): \Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates," Journal of the American Statistical Association, 91, 1632-1640.

[49] Horowitz, J.L. (1998): \Semiparametric Methods in Econometrics," Spinger-Verlag, New York.

[50] Ichimura, H. (1993): \Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Model," Journal of Econometrics, 58, 71-120.

[51] Ichimura, H. and L.F. Lee (1991): \Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in Nonparametric and Semiparametric Methods in Econometrics and Statistics, ed. by W.A. Barnett, J.L. Powell and G. Tauchen. Cambridge: Cambridge University Press.

[52] Klein, R.W. and R.S. Spady (1993): \An E± cient Semiparametric Estimator of the Binary Response Model," Econometrica, 61, 387-421.

[53] Klein, R.W. and R.P. Sherman (1998): \Shift Restrictions and Semiparametric Estimation in a Generalized Transformation Model," Manuscript.

[54] Kmenta, J. (1971), Elements of Econometrics, New York: Macmillan.

[55] Kyriazidou, E. (1997): \Estimation of A Panel Data Sample Selection Model," Econometrica, 1335-1364.

[56] Lee, M.J. (1996): \Nonparametric Two Stage Estimation of Simultaneous Equations with Limited Endogenous Regress

98a): \Semiparametric L us or Mismeasured Regressors," Econometrica, 66, 105-121.

[58] Lewbel, metric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," Manuscript.

[59] Lewbel A. and O.B. Linton (1998): \Nonparametric Censored Regression," Manuscript.

[60] Linton, O.B. (1993): \Adaptive Estimation in ARCH models," Econometric Theory, 9, 539-569.

[61] Maddala, G.S. (1983): Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press.

[62] Maddala, G.S., and F. Nelson (1974): \Switching Regression Models with Exogenous and Endogenous Switching," Proceedings of the American Statistical Association, 423-426.

[63] Manski, C.F. (1985): \Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," Journal of Econometrics, 27, 313-333.

[64] Manski, C.F. (1988): \Identi¯cation of Binary Response Models," Journal of the American Statistical Association, 83, 729-738.

[65] Newey, W.K. (1988a): \Adaptive Estimation of Regression Models Via Moment Restriction ," Journal of Econometrics, 38, 301-339.

[66] Newey, W.K. (1988b): \Two-Step Series Estimation of Sample Selection Models," Manuscript.

[67] Newey, W.K. (1991): \E± cient Estimation of Tobit Models under Symmetry," in Nonparametric and Semiparametric Methods in Econometrics and Statistics, ed. by W.A. Barnett, J.L. Powell and G. Tauchen. Cambridge: Cambridge University Press.

[68] Newey, W.K. (1994): \The Asymptotic Variance of Semiparametric Estimators," Econometrica, 62, 1349-1382.

[69] Park, R.E. (1966): \Estimation with Heteroscedastic Error Terms," Econometrica, 34, 888.

[70] Paris, S.J. (1953): \A Note on Heteroscedastic Errors in Regression Analysis," Review of the International Statistical Institute, 21, 28-29.

[71] Paris, S.J., and H.S. Houthakker (1955): The Analysis of Family Budgets, New York: Cambridge University Press.

[72] Powell, J.L. (1986): \Symmetrically trimmed least squares estimation for Tobit models," Econometrica, 54, 1435-1460.

[73] Powell, J.L. (1989): \Semiparametric Estimation of Bivariate Latent Variable Models," University of Wisconsin-Madison.

[74] Powell, J.L. (1994): \Estimation of Semiparametric Models. Handbook of Econometrics," in Handbook of Econometrics Vol 4, ed. by R. Engle and D. McFadden. North-Holland: Elsevier Science, 2444-2521.

[75] Powell, J.L., J.H. Stock and T.M. Stoker (1989): \Semiparametric Estimation of Weighted Average Derivatives," Econometrica, 57, 1403-1430.

[76] Robinson, P.(1988): \Root-N-Consistent Semiparametric Regression," Econometrica, 56, 931-954.

[77] Rutemiller, H.C., and D.A. Bowers (1968): \Estimation in a Hete
l," Journal of the American Statistical Association, 63, 552-557.

[78] Sherman, R.P. (1993): \The Limiting Distribution of the Maximum Rank Correlation Estimator," Econometrica, 61, 123-137.

[79] Sherman, R.P. (1994): \U-processes in the Analysis of a Generalized Semiparametric Regression Estimator," Econometric Theory, 11, 372-395.

[80] Theil, H. (1951): \Estimates and Their S
bution," Review of the International Statistical Institute, 19, 141-147.

[81] Whang, Y.J. and D.W.K. Andrews (19 ¯cation for Parametric and Semiparametric Models,"Journal of Econometrics, 57, 277-318.

[82] Zheng, J.X. (1995): \Semiparametric E± ciency Bounds f mple Selection Models under Conditional Symmetry," Economics Letters, 47, 249-253.

[83 998): \Consistent Speci¯cation Testing for Conditional Symmetry," Econometric Theory, 14, 139-149.