COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1119

SEMIPARAMETRIC ESTIMATION OF A SAMPLE SELECTION MODEL

Donald W. K. Andrews and Marcia A. Schafgans

April 1996

Semiparametric Estimation of a Sample Selection Model

Donald W.K. Andrews
and
Marcia M.A. Schafgans*

April 1996

Abstract

This paper provides a consistent and asymptotically normal estimator for the intercept of a semiparametrically estimated sample selection model. The estimator uses a decreasingly small fraction of all observations as the sample size goes to infinity, as in Heckman (1990). In the semiparametrics literature, estimation of the intercept typically has been subsumed in the nonparametric sample selection bias correction term. The estimation of the intercept, however, is important from an economic perspective. For instance, it permits one to determine the "wage gap" between unionized and nonunionized workers, decompose the wage differential between different socio-economic groups (e.g., male–female and black–white), and evaluate the net benefits of a social program.

*Correspondence: Marcia Schafgans, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, U.K. Email: m.schafgans@lse.ac.uk. This paper was partially supported by National Science Foundation Grant SES–9121914 and by a C.A. Anderson Fellowship of the Cowles Foundation. The authors wish to thank James Heckman for his comments and Glena Ames for her able assistance with the typing of this document.

1 Introduction

Semiparametric estimation of sample selection models has received considerable attention in the last decade. The reason is that parameter estimators of the sample selection model are inconsistent when incorrect distributional assumptions are made about the errors (e.g., see Goldberger (1983) and Arabmazar and Schmidt (1981, 1982)).

This paper considers semiparametric estimation of the intercept parameter, μ_0, in a sample selection model. In the semiparametrics literature to date, the intercept has been absorbed in the nonparametric sample selection bias correction term. The only exceptions are the estimators of Gallant and Nychka (1987) and Heckman (1990). Gallant and Nychka's (1987) estimator of μ_0 has been shown to be consistent, but its asymptotic distribution is unknown. The asymptotic distribution of Heckman's (1990) estimator also is unknown. The estimator we consider here is a slight variant of Heckman's estimator. We show that it is consistent and asymptotically normal.

The economic interpretation of an estimated sample selection model makes estimation of the intercept important. It is required for the evaluation of the "wage gap," an issue that has received considerable attention in the literature (e.g., see Oaxaca (1973), Lewis (1986), Smith and Welch (1986), Wellington (1993), and Baker et al. (1993)), and for the evaluation of social programs. Estimation of the intercept permits evaluation of the net benefit of a social program by permitting comparisons of the actual outcome of participants with the expected outcome had they chosen not to participate.

The estimator, $\widehat{\mu}_n$, that we consider uses a decreasingly small fraction of all observations as the sample size, n, goes to infinity. This approach is advocated by Heckman (1990). He suggests estimating μ_0 using only those observations for which the probability of selection in the truncated or censored sample is close to one and in the limit as $n \to \infty$ is one. This approach works because the conditional mean of the outcome equation errors is close to zero for observations whose probability of selection is close to one. This is an example of Chamberlain's (1986) "identification at infinity."

Heckman suggests using all observations for which the estimated index of the participation equation, $X_i'\widehat{\beta}$, exceeds a certain threshold γ_n. We introduce a weighting scheme for these observations, where observations exceeding this threshold are weighted by a smooth monotone [0,1]-valued function, such as a distribution function. The introduction of this function, $s(\cdot)$, allows us to establish the distribution theory for the estimator. The smoothness we impose on this function, viz., differentiability of order three, is used to show that the asymptotic distribution of the estimator $\widehat{\mu}_n$ is not affected by preliminary estimators such as $\widehat{\beta}$.

Our distribution theory for the estimator $\widehat{\mu}_n$ assumes the existence of root-n consistent semiparametric estimators of (θ_0, β_0), where θ_0 is the vector of parameters of the outcome equation with the *exclusion* of the intercept and β_0 is the vector of parameters of the selection equation. Several such estimators are available in the literature.

The estimator $\widehat{\mu}_n$ depends on the bandwidth parameter γ_n. For consistency, γ_n is required to go to infinity as $n \to \infty$. The choice of γ_n is constrained by the requirements that the variability and bias of $\widehat{\mu}_n$ go to zero as $n \to \infty$. The first imposes an upper bound on how fast γ_n can increase; the second a lower bound. The thinner the upper tail of the errors in the selection equation compared to the upper tail of the index $X_i'\beta_0$, the greater is the latitude in the choice of γ_n. A formal method for determining an optimal bandwidth parameter γ_n has yet to be developed.

The remainder of this paper is organized as follows: In Section 2, the sample selection model is discussed. The discussion of several applications in this section motivates our interest in estimating γ_0. In Section 3, our proposed estimator is defined. Consistency and asymptotic normality of this estimator are established in Section 4 under the assumption that the regressors and errors are independent. In Section 5, analogous results are established with the latter assumption relaxed. Section 6 concludes. Appendices A and B contain proofs of the results stated in Sections 4 and 5 respectively.

2 The Sample Selection Model

An early discussion of the sample selection problem in the economics literature is given in Roy (1951). In his prototypical model of self-selection, agents choose among two professions, hunting and fishing, based on their comparative advantage.

The discussion of the econometric implications of sample selection started in the early seventies with the papers by Gronau (1974), Heckman (1974), and Lewis (1974). In their studies, the problem of sample selection bias is discussed in the context of the decision by women to participate in the labor force or not. The distribution of the wage offers sampled is truncated by the "self-selection" of women in the labor force, where women choose to be "in the sample" of workers if the offered wage exceeds their reservation wage.

The sample selection model has been used in a wide variety of other applications. For example, it has been used by Lee and Trost (1978) in the context of the demand for housing, by Willis and Rosen (1979) in the context of education, and by Lee (1986) in the context of labor unions and wages. The self-selection model also is used extensively in the evaluation of the benefits of social programs. For further references, see Maddala (1983) and Amemiya (1984).

The sample selection model can be written as

$$\begin{aligned} Y_i^* &= \mu_0 + Z_i'\theta_0 + U_i, \\ D_i &= 1(X_i'\beta_0 > \varepsilon_i), \quad \text{and} \\ Y_i &= Y_i^* D_i \quad \text{for} \quad i = 1, ..., n, \end{aligned} \tag{2.1}$$

where (Y_i, D_i, Z_i, X_i) are observed random variables. The first equation is the outcome equation and the second equation is the participation equation. For convenience, we let

$$W_i = X_i'\beta_0. \tag{2.2}$$

To express the model given in (2.1) in terms of the Gronau–Heckman–Lewis model, we note that in their model Y_i^* is the latent offered wage and D_i is a dummy variable indicating whether an individual is employed, i.e., whether $Y_i^* - Y_i^r$ exceeds zero, where Y_i^r denotes the individual's latent reservation wage. The observed wage is given by Y_i. The variables influencing the decision to participate in the labor market are given by X_i and the determinants of the wage offer are given by Z_i.

The standard approach to estimation of this model assumes that (U_i, ε_i) are bivariate normal with zero mean and unknown covariance matrix and are independent of (Z_i, X_i). With this assumption, the parameters can be estimated by maximum likelihood or the two–step estimator of Heckman (1976). This approach is known to yield inconsistent estimators if the normality assumption fails, e.g., see Arabmazar and Schmidt (1981, 1982) and Goldberger (1983).

Semiparametric estimation methods provide a way to overcome this deficiency. These methods consider the estimation of the parameters of interest without restricting the distribution function of the error terms or restricting the functional form of heteroskedasticity to lie in a finite-dimensional parametric family. Important progress on semiparametric estimation of selection models has been made by Gallant and Nychka (1987), Newey (1988), Robinson (1988), Powell (1989), Cosslett (1990), and Ichimura and Lee (1990). Also see Andrews (1991).

The point of departure in these papers is the conditional mean index function representation of the sample selection problem:

$$\begin{aligned} E(Y_i|D_i = 1, X_i, Z_i) &= Z_i'\theta_0 + E(U_i|D_i = 1, X_i) \quad \text{and} \\ E(U_i|D_i = 1, X_i) &= \kappa(X_i'\beta_0), \end{aligned} \tag{2.3}$$

where $X_i'\beta_0$ is an index function and $\kappa(\cdot)$ is an unknown (smooth) function. The function $\kappa(\cdot)$ is sometimes called the sample selection correction function. It equals the inverse Mill's ratio when (U_i, ε_i) are bivariate normal. In the above model, conditional heteroskedasticity of U_i is allowed, although only through the single index $X_i'\beta_0$. The objective of the papers referred to above is to eliminate the contaminating effect of $E(U_i|D_i = 1, X_i)$ in forming regression estimates of θ_0.

Cosslett (1990) approximates the conditional mean of U_i using step functions based on a nonparametric estimator of $\kappa(\cdot)$. Robinson (1988), Powell (1989), and Ichimura and Lee (1990) difference out the conditional mean. Newey (1988) and Andrews (1991) approximate the conditional mean of U_i by series expansion methods.

These papers present various criteria for identification of (θ_0, β_0). In order to obtain a consistent estimator of β_0, one needs information on X_i when $D_i = 0$. Even if β_0 is known, identification of θ_0 fails if $\kappa(X_i'\beta_0)$ lies in the space spanned by Z_i.

As pointed out by Heckman (1990), all of the above papers absorb the intercept, μ_0, into the definition of the conditional mean $E(U_i|D_i = 1, X_i)$. None of these papers produces a consistent estimator of μ_0. Gallant and Nychka, also using a series expansion method, do obtain a consistent estimator of μ_0. The distribution theory for their estimator, however, has not been developed. Gallant and Nychka assume that the errors and regressors are independent. They also impose a continuity condition on the distribution of the errors and regressors that is somewhat complicated and potentially difficult to verify.

The estimation of the intercept, μ_0, has economic importance. We show this below using several examples. This is what prompted our interest in obtaining a consistent and asymptotically normal estimator of the intercept.

Consider the Roy model. In this model, estimation of the intercept, μ_0, allows one to compute the gain from moving a worker with certain attributes from one profession into another. Let Y_{Hi}^* and Y_{Fi}^* denote the latent offered wages for hunting and fishing respectively. Let D_{Hi} be a dummy variable indicating whether an individual is a hunter, i.e., whether $Y_{Hi}^* - Y_{Fi}^*$ exceeds zero. The observed wages are given by Y_{Hi} and Y_{Fi}. The model is

$$\begin{aligned} Y_{Hi}^* &= \mu_H + Z_{Hi}'\theta_H + U_{Hi}, \\ Y_{Fi}^* &= \mu_F + Z_{Fi}'\theta_F + U_{Fi}, \\ D_{Hi} &= 1(Y_{Hi}^* - Y_{Fi}^* > 0) = 1(X_i'\beta_0 > \varepsilon_i), \\ Y_{Hi} &= Y_{Hi}^* D_{Hi}, \text{ and} \\ Y_{Fi} &= Y_{Fi}^*(1 - D_{Hi}) \text{ for } i = 1, ..., n, \end{aligned} \tag{2.4}$$

where $Z_{Fi} = [Z_i \vdots Z_{fi}]$ denotes variables that affect wage offers in the fishing profession, $Z_{Hi} = [Z_i \vdots Z_{hi}]$ denotes variables that affect wage offers in the hunting profession, and Z_i denotes variables that affect wage offers in both professions. Define $X_i = [Z_i \vdots Z_{fi} \vdots Z_{hi}]$. Model (2.1) for hunters consists of the equations for Y_{Hi}^*, D_{Hi}, and Y_{Hi}. Model (2.1) for fishermen consists of the equations for Y_{Fi}^*, $1 - D_{Hi}$, and Y_{Fi}.

The sample selection model permits the estimation of various interesting quantities related to wage differentials between the hunting and fishing professions. The focus of most empirical literature is the difference in expected wage between hunting and fishing for a *randomly selected* worker with attributes $(X_i = x, Z_{Hi} = z_H, Z_{Fi} = z_F)$:

$$\mu_H - \mu_F + z_H'\theta_H - z_F'\theta_F. \tag{2.5}$$

Another interesting quantity is the difference in expected wage between hunting and fishing for an individual with attributes $(X_i = x, Z_{Hi} = z_H, Z_{Fi} = z_F)$, who *self-selects* into one specific sector, say hunting:

$$\begin{aligned} & E(Y_{Hi}^*|D_{Hi} = 1, x, z_H, z_F) - E(Y_{Fi}^*|D_{Hi} = 1, x, z_H, z_F) \\ = \; & \mu_H - \mu_F + z_H'\theta_H - z_F'\theta_F + E(U_{Hi} - U_{Fi}|D_{Hi} = 1, x). \end{aligned} \tag{2.6}$$

When equations (2.5) and (2.6) are averaged over all individuals in the hunting profession, one obtains

$$\begin{aligned} & \mu_H - \mu_F + \bar{Z}_H'\theta_H - \bar{Z}_F'\theta_F \quad \text{and} \\ & \mu_H - \mu_F + \bar{Z}_H'\theta_H - \bar{Z}_F'\theta_F + E(U_{Hi} - U_{Fi}|D_{Hi} = 1), \end{aligned} \tag{2.7}$$

respectively, where $\bar{Z}_H = E(Z_{Hi}|D_{Hi} = 1)$ and $\bar{Z}_F = E(Z_{Fi}|D_{Hi} = 1)$ are the endowments of the skills in each sector available to the "average" worker in the hunting profession. The estimation of the intercepts (μ_F, μ_H) is necessary to estimate the quantities in (2.5)–(2.7).

In a similar fashion, Heckman (1990) discusses the parameters of economic interest in sample selection models in the context of the effect of unionism on wages. The difference in expected wage between the unionized and non-unionized sectors for individuals who self-select into the unionized sector is called the "wage gap" by Lewis (1986). Estimation of these quantities throws light on the effects of self-selection. If self-selection is based on comparative advantage, as in Roy's example, then the benefits of moving a randomly selected worker are less than the benefits of moving a worker under self-selection.

The sample selection model also is useful in the evaluation of social programs. Estimation of the intercept allows one to evaluate the net benefit of a social program, by allowing one to compare the actual outcome of participants with the expected outcome had they chosen not to participate. If individuals who have a comparative advantage with the program self-select into the program, then these individuals benefit more from it than would a randomly selected individual with the same characteristics. In consequence, the program produces greater benefit under self-selection than under random assignment.

Consistent estimation of μ_0 also is of great importance in the extensive literature analyzing wage differences between different socio-economic groups (e.g., male–female and black–white). Following Oaxaca (1973), many studies have attempted to decompose the "wage gap" between these groups (e.g., Smith and Welch (1986), Wellington (1993), and Baker et al. (1993)).[1] The decomposition attempts to illuminate what part of the gap can be explained by differences in wage-related characteristics and what part can be explained by differences in the wage structure. Consider the male–female wage gap. The model is given by

$$\begin{aligned} Y_{ki}^* &= \mu_k + Z_i'\theta_k + U_{ki}, \\ D_{ki} &= 1(X_i'\beta_k > \varepsilon_{ki}), \text{ and} \\ Y_{ki} &= = D_{ki}Y_{ki}^* \quad \text{for} \quad i = 1, ..., n, \end{aligned} \tag{2.8}$$

where $k =$ male, female and Y_{ki}^* and Y_{ki} denote the log of latent and observed wages respectively. Following Oaxaca (1973), the wage gap, $\bar{Y}_m - \bar{Y}_f$ $(= E(Y_{mi}|D_{mi} = 1) - E(Y_{fi}|D_{fi} = 1))$, can be decomposed as

$$\begin{aligned} \bar{Y}_m - \bar{Y}_f &= (\mu_m - \mu_f) + \bar{Z}_f(\theta_m - \theta_f) + (\bar{Z}_m - \bar{Z}_f)\theta_m + E(U_{mi}|D_{mi}=1) - E(U_{fi}|D_{fi}=1) \\ &= (\mu_m - \mu_f) + \overline{Z}_m(\theta_m - \theta_f) + (\bar{Z}_m - \bar{Z}_f)\theta_f + E(U_{mi}|D_{mi}=1) - E(U_{fi}|D_{fi}=1), \end{aligned} \tag{2.9}$$

where $\bar{Z}_m$ and $\bar{Z}_f$ are the endowments of the skills available to the "average" male and female worker respectively. The first two parts on the right-hand side of equation (2.9)

[1] In these studies "wage gap" is defined as the difference in average log earnings between these socio-economic groups.

measure the wage gap explained by the differences in male-female wage structures for the same observed job-related characteristics. The third part on the right-hand side measures the wage gap due to male-female differences in wage-related characteristics and the remainder is due to the self-selection correction. Equation (2.9) shows two ways in which one can measure the wage gap: using the male wage structure or using the female wage structure.[2] In either case, the estimation of (μ_m, μ_f) is required to estimate the extent to which the wage gap is explained by differences in male-female wage structures for the same observed job-related characteristics.

3 The Estimator

The estimator we consider is

$$\widehat{\mu}_n = \frac{\sum_{i=1}^n (Y_i - Z_i'\widehat{\theta}) D_i s(X_i'\widehat{\beta} - \gamma_n)}{\sum_{i=1}^n D_i s(X_i'\widehat{\beta} - \gamma_n)}, \tag{3.1}$$

where $s(\cdot)$ is a non-decreasing [0,1]-valued function that has three derivatives bounded over R and for which $s(x) = 0$ for $x \leq 0$ and $s(x) = 1$ for $x \geq b$ for some $0 < b < \infty$. The preliminary estimators $(\widehat{\theta}, \widehat{\beta})$ are root-n consistent estimators of (θ_0, β_0). The parameter γ_n is called the bandwidth or smoothing parameter. This bandwidth parameter is chosen such that $\gamma_n \to \infty$ as $n \to \infty$.

The estimator suggested by Heckman (1990) is

$$\widetilde{\mu}_n = \frac{\sum_{i=1}^n (Y_i - Z_i'\widehat{\theta}) D_i 1(X_i'\widehat{\beta} > \gamma_n)}{\sum_{i=1}^n D_i 1(X_i'\widehat{\beta} > \gamma_n)}. \tag{3.2}$$

Comparing the two formulae (3.1) and (3.2), it is clear that the estimator $\widehat{\mu}_n$ differs from Heckman's (1990) $\widetilde{\mu}_n$ only in that it replaces the indicator function $1(\cdot)$ with a smooth function $s(\cdot)$. The introduction of this function allows us to provide the estimator with a distribution theory. The smoothness we impose on this function, viz., differentiability of order three, is used to show that the preliminary estimators $(\widehat{\theta}, \widehat{\beta})$ do not affect the asymptotic results.

Heckman's estimator $\widetilde{\mu}_n$ is essentially a sample average of the random variables $U_i + \mu_0$ over a fraction of all observations, since $Y_i - Z_i'\widehat{\theta} \to_p U_i + \mu_0$ as $n \to \infty$ for all $i \geq 1$. The effective sample size is equal to the number of observations used for the estimation of μ_0. Since we introduce a weighting scheme for these observations, viz., the smooth function $s(\cdot)$, our estimator $\widehat{\mu}_n$ is a weighted sample average of the random variables $U_i + \mu_0$, where observations with $X_i'\widehat{\beta}$ greater than γ_n and with $X_i'\widehat{\beta}$ close to the threshold γ_n are weighted less than those further away.

[2] Oaxaca (1973) suggested the adoption of either the male wage structure or the female wage structure as the nondiscriminatory wage structure (with the actual nondiscriminatory structure being bracketed by them). Alternative weighting schemes are suggested by Cotton (1988), Neumark (1988), and Oaxaca and Ransom (1994).

4 Consistency and Asymptotic Normality

The consistency and asymptotic normality of $\widehat{\mu}_n$ are established in Sections 4.1 and 4.2 respectively below. Section 4.3 addresses the estimation of the asymptotic covariance matrix of $(\widehat{\mu}_n, \widehat{\theta}, \widehat{\beta})$.

4.1 Consistency

We now state the Assumptions 1–7 that are used to establish consistency of $\widehat{\mu}_n$. Each assumption is discussed below.

Assumption 1: $(Z_i,\, X_i,\, U_i,\, \varepsilon_i)$ *are* iid rv's *with* $E\|Z_i\|^p < \infty$*,* $E\|X_i\|^p < \infty$*, and* $E|U_i|^\lambda < \infty$ *for some* $p > 3$ *and* $\lambda > 2$*.*

Assumption 2: (a) $EU_i = 0$.
(b) $(U_i,\, \varepsilon_i)$ *is independent of* $(Z_i,\, X_i)$*.*

Assumption 3: $s(\cdot) : R \to [0,1]$ *is a nondecreasing three times differentiable function with* $s(x) = 0\ \forall x \leq 0$*,* $s(x) = 1\ \forall x \geq b$ *for some* $0 < b < \infty$*, and* $\sup_{x\in R} |s'''(x)| < \infty$*.*

Assumption 4: $P(W_i > w)^{1+\xi}/P(W_i > w+b) = O(1)$ *as* $w \to \infty$ *for some* $\xi \in [0,\, 1/3]$ *for* b *as in Assumption* 3*.*

Assumption 5: $\sqrt{n}(\widehat{\theta} - \theta_0) = O_p(1)$ *and* $\sqrt{n}(\widehat{\beta} - \beta_0) = O_p(1)$*.*

Assumption 6: $\gamma_n \to \infty$*.*

Assumption 7: $\dfrac{\sqrt{n} E D_i s(W_i - \gamma_n)}{(E D_i s^2(W_i - \gamma_n))^{1/2}} \to \infty$*.*

The first assumption imposes quite mild moment conditions on $(Z_i, X_i, U_i, \varepsilon_i)$. It rules out unconditional heteroskedasticity and time series applications.

Assumption 2(a) is not restrictive since a non-zero mean can be absorbed in the definition of μ_0. The assumption that (Z_i, X_i) and (U_i, ε_i) are independent, Assumption 2(b), can be restrictive. Some semiparametric estimation techniques for the sample selection model have considered the less restrictive case in which the errors are allowed to depend on X_i through the index $W_i = X_i'\beta_0$. In Section 5, we relax Assumption 2 to incorporate this less restrictive case.

Assumption 3 is an assumption of smoothness of the function $s(\cdot)$. An example of a function satisfying this condition is given by:

$$s(x) = \begin{cases} 1 - \exp(-\frac{x}{b-x}) & \text{for} \quad x \in (0, b) \\ 0 & \text{for} \quad x \leq 0 \\ 1 & \text{for} \quad x \geq b. \end{cases} \tag{4.1}$$

The boundedness of the first three derivatives of $s(\cdot)$ is not essential in all parts of the proof. To indicate *where* it is needed, the following weaker alternative, Assumption 3′, is used wherever possible. In particular, Assumption 3′ is used when (θ_0, β_0) is assumed to be known.

ASSUMPTION 3′: *$s(\cdot) : R \to [0,1]$ is a distribution function of a* rv *with support contained in $[0,b]$ for some $0 \leq b < \infty$.*

Note that Assumption 3′ allows for $s(x) = 1(x > 0)$, which generates Heckman's estimator $\widetilde{\mu}_n$.

Assumption 4 requires that W_i has unbounded support from above. It rules out distributions of W_i that are too thin upper tailed. For example, suppose the upper tail of W_i decays as $1 - F(x) \sim \exp(-\exp(\lambda x))$.[3] Then, for $\xi > 0$, Assumption 4 is satisfied as long as $\lambda \leq \ln(1+\xi)/b$, where b and ξ are defined in Assumptions 3 and 4 respectively. The upper tail of W_i is too thin when λ exceeds this cutoff point. In this case, Assumption 4 is never satisfied for $\xi = 0$. If W_i has a Weibull upper tail $1 - F(x) \sim \exp(-\lambda x^c)$ for some $c > 0$ and $\lambda > 0$, then Assumption 4 is satisfied with $\xi = 0$ if $c \leq 1$. If $c > 1$, the assumption holds for all $\xi > 0$. (A special case is $W_i \sim N(0, \sigma^2)$, in which case $c = 2$.) If the upper tail of the distribution of W_i is declining geometrically, i.e., W_i has a Pareto upper tail $1 - F(x) \sim x^{-\lambda}$ for $\lambda > 0$, then Assumption 4 is satisfied for any $\xi \geq 0$.

Assumption 5 requires that the preliminary estimators are root-n consistent. The estimators suggested by Newey (1988), Robinson (1988), Powell (1989), Ichimura and Lee (1990), and Andrews (1991) satisfy this assumption. The estimator suggested by Cosslett (1990) has not been shown to satisfy this for his estimator. To date, only a consistency result is available for his estimator. For the first-step estimation of β_0 in the participation equation, these papers typically rely on Ichimura's (1985) or Klein and Spady's (1993) estimator for the binary choice model.

The bandwidth parameter, γ_n, is required to go to infinity as the number of observations, n, goes to infinity, by Assumption 6. This guarantees that the estimation of μ_0 is based on only those values of X_i for which $P(D_i = 1|X_i)$ is close to one and in the limit is equal to one.

A bound on the speed with which γ_n is allowed to go to infinity is given by Assumption 7. This assumption imposes the restriction that the variability of $\widehat{\mu}_n$ is asymptotically zero. The reciprocal of the ℓhs of this assumption is equal to the standard error of

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - Z_i'\theta_0)D_i s(X_i'\beta_0 - \gamma_n)}{\sigma E D_i s(X_i'\beta_0 - \gamma_n)} = \widehat{\mu}_{n0}(1 + o_p(1))/\sigma, \tag{4.2}$$

where $\widehat{\mu}_{n0}$ is identical to $\widehat{\mu}_n$ with $(\widehat{\theta}, \widehat{\beta})$ replaced by the true values (θ_0, β_0). The equality holds by Lemma A–2 in the Appendix. Stronger, but simpler, alternative assumptions to Assumption 7 are given by the following Assumptions 7* and 7**.

[3] Here and below, "$a(x) \sim b(x)$" is defined to mean that the ratios $a(x)/b(x)$ and $b(x)/a(x)$ are $O(1)$ as $x \to \infty$.

Assumption 7^*: $nP(W_i > \gamma_n + b) \to \infty$.

Assumption 7^{**}: $n^{1/(1+\xi)}P(W_i > \gamma_n) \to \infty$ *for* ξ *as in Assumption* 4.

Assumptions 7^* and 7^{**} place restrictions on the speed at which γ_n can go to infinity relative to the upper tail probabilities of W_i. They ensure that the sample size on which the estimation of μ_0 is based is sufficiently large. For example, if W_i has a Weibull upper tail $1-F(x) \sim \exp(-\lambda x^c)$ for some $c > 0$ and $\lambda > 0$, then Assumptions 7^* and 7^{**} are satisfied when $\gamma_n \leq (\frac{1}{\lambda}\log n^{1-\tau_1})^{1/c}$ and $\gamma_n \leq (\frac{1}{\lambda}\log n^{(1-\tau_2)/(1+\xi)})^{1/c}$, respectively, for arbitrary $0 < \tau_1 < 1$ and $0 < \tau_2 < 1$.

The following Lemma shows that Assumptions 7^* and 7^{**} imply Assumption 7.

Lemma 1: (a) *Under Assumptions* 1, 2(b), $3'$, *and* 6, *Assumption* 7^* *implies Assumption* 7.
(b) *Under Assumptions* 4 *and* 6, *Assumption* 7^{**} *implies Assumption* 7^*.Our consistency result is given by Theorem 1.

Theorem 1: *Under Assumptions* 1–7, $\widehat{\mu}_n \xrightarrow{p} \mu_0$.

The proofs of Lemma 1, Theorem 1, and other results stated in this section are given in Appendix A.

4.2 Asymptotic Normality

Under Assumptions 1–7, the following two assumptions are necessary and sufficient for $\widehat{\mu}_n$ to be asymptotically normal.

Assumption 8: $\forall \varepsilon > 0,\ \lim_{n\to\infty} E\left(\frac{(B_n - EB_n)^2}{\mathrm{Var}(B_n)} 1\left[\frac{(B_n - EB_n)^2}{\mathrm{Var}(B_n)} > n\varepsilon\right]\right) = 0,$ *where* $B_n = U_i D_i s(W_i - \gamma_n)$.

Assumption 9: $\sqrt{n} E U_i D_i s(W_i - \gamma_n)/(ED_i s^2(W_i - \gamma_n))^{1/2} \to 0$.

Assumption 8 is the usual Lindeberg condition. Assumption 9 ensures that the bias of $\widehat{\mu}_n$ goes to zero asymptotically. Sufficient conditions for Assumptions 8 and 9 are Assumptions 7^* (given above) and 9^* respectively.

Assumption 9^*: $\sqrt{n} E|U_i| 1(\varepsilon_i > \gamma_n) \cdot P(W_i > \gamma_n)^{(1-\xi)/2} \to 0$ *for* $\xi \in [0, 1/3]$ *as in Assumption* 4.

Both Assumptions 7^* and 9^* are related to the choice of γ_n. The former implies an upper bound as described above; the latter implies a lower bound. A sufficient condition for the latter assumption is that ε_i is a bounded random variable, since then $1(\varepsilon_i > \gamma_n) \to 0$ *a.s.* by Assumption 6. Alternatively, if $\xi = 0$ and U_i is bounded, a sufficient condition is that $n^{1/2}P(\varepsilon_i > \gamma_n)P(W_i > \gamma_n)^{1/2} \to 0$. In Assumption

9^*, the relative upper tail thicknesses of ε_i and W_i are crucial. If ε_i and W_i both have similar Weibull upper tails $1 - F(x) \sim \exp(-\lambda x^c)$ for some $c > 0$ and $\lambda > 0$ and U_i is bounded, then Assumption 9^* is satisfied if $n^{1/(3-\xi)}P(W_i > \gamma_n) \to 0$ or $\gamma_n \geq (\frac{1}{\lambda} \log n^{(1+\tau_3)/(3-\xi)})^{1/c}$ for some $\tau_3 > 0$.

LEMMA 2: (a) *Under Assumptions* 1, 2, $3'$, 4, *and* 6, *Assumption* 7^* *implies Assumption* 8.
(b) *Under Assumptions* 1, 2 , $3'$, 4, *and* 6, *Assumption* 9^* *implies Assumption* 9.

Asymptotic normality of $\widehat{\mu}_n$ is established in the following theorem. Let $\sigma^2 = \mathrm{Var}(U_i)$.

THEOREM 2: *Under Assumptions* 1–7,
(a) $\dfrac{\sqrt{n}ED_i s(W_i - \gamma_n)}{\sigma(ED_i s^2(W_i - \gamma_n))^{1/2}} \left(\widehat{\mu}_n - \mu_0 - \dfrac{EU_i D_i s(W_i - \gamma_n)}{ED_i s(W_i - \gamma_n)} \right) \xrightarrow{d} N(0,1)$ *iff Assumption* 8 *holds and*

(b) $\dfrac{\sqrt{n}ED_i s(W_i - \gamma_n)}{\sigma(ED_i s^2(W_i - \gamma_n))^{1/2}} (\widehat{\mu}_n - \mu_0) \xrightarrow{d} N(0,1)$ *iff Assumptions* 8 *and* 9 *hold.*

The mutual compatibility of Assumptions 7–9, or their sufficient counterparts $7^*$and 9^*, is crucial for asymptotic normality. The thinner the upper tail of ε_i relative to that of W_i, the greater the latitude in the choice of γ_n, as can be inferred from the discussion of Assumptions 7^* and 9^* given above.

To test hypotheses and construct confidence intervals for functions of $(\mu_0, \theta_0, \beta_0)$, we need a joint asymptotic normality result for $(\widehat{\mu}_n, \widehat{\theta}, \widehat{\beta})$. For example, this result is needed to calculate a confidence region for $\widehat{Y}_i$. For this joint normality result, we need to impose the following additional assumption.

ASSUMPTION 5^*: $\sqrt{n}\begin{pmatrix} \widehat{\theta}-\theta_0 \\ \widehat{\beta}-\beta_0 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i + o_p(1)$ *for some* iid *mean zero rv's* $\{Q_i : i \geq 1\}$ *with* $\Omega = EQ_i Q_i'$ *positive definite,* $E\|Q_i\|^\lambda < \infty$ *for some* $\lambda > 2$, *and* $E\|U_i Q_i\|^p < \infty$ *for some* $p > 3$.

Assumption 5^* is stronger than Assumption 5. Assumption 5^* implies that the estimators $(\widehat{\theta}, \widehat{\beta})$ satisfy an asymptotic linear expansion. This holds typically by the asymptotic normality proofs for $\widehat{\theta}$ and $\widehat{\beta}$. For example, in Klein and Spady's (1993) estimator of β_0,

$$Q_i = E\left\{ \left[\frac{\partial P}{\partial \beta}\right] \left[\frac{\partial P}{\partial \beta}\right]' \left[\frac{1}{P(1-P)}\right] \right\}^{-1}_{\beta=\beta_0} \cdot \tau_i r_{in} w_i, \tag{4.3}$$

where P is the probability of selection, τ_i is a likelihood trimming function, w_i is a weight function, and r_{in} is equal to $(D_i - P_i(\beta_0))/[(E(D_i|X_i'\beta_0 = x_i'\beta_0) + \delta_{in})P_i(\beta_0)(1 - P_i(\beta_0))]$, where δ_{in} is a probability trimming function.[4]

[4]For details, see Klein and Spady (1993).

The assumption that $E\|U_iQ_i\|^p < \infty$ for some $p > 3$ is used in establishing the block diagonality of the asymptotic covariance matrix between $\widehat{\mu}_n$ and $(\widehat{\theta}, \widehat{\beta})$. Intuitively, this block diagonality makes sense, since $\widehat{\mu}_n$ is estimated using a decreasingly small fraction of all observations and estimators of (θ_0, β_0) that leave out these observations are asymptotically equivalent to $(\widehat{\theta}, \widehat{\beta})$.

The following asymptotic joint normality result holds.

THEOREM 3: *Under Assumptions* 1–4, 5*, 6, *and* 7,

$$\begin{pmatrix} \frac{\sqrt{n} ED_i s(W_i - \gamma_n)}{\sigma(ED_i s^2(W_i - \gamma_n))^{1/2}} (\widehat{\mu}_n - \mu_0) \\ \sqrt{n}\Omega^{-1/2} \binom{\widehat{\theta}-\theta_0}{\widehat{\beta}-\beta_0} \end{pmatrix} \xrightarrow{d} N(0, I) \text{ iff Assumptions 8 and 9 hold.}$$

4.3 Asymptotic Normality with Unknown Covariance Matrix

To compute standard errors and test statistics one needs an estimator of the asymptotic covariance matrix of $(\widehat{\mu}_n, \widehat{\theta}, \widehat{\beta})$. More specifically, one needs consistent estimators of σ^2, the asymptotic covariance matrix of $(\widehat{\theta}, \widehat{\beta})$, and the normalizing factors $ED_i s(W_i - \gamma_n)$ and $ED_i s^2(W_i - \gamma_n)$.

Define

$$\widehat{\sigma}_n^2 = \frac{\sum_{i=1}^n (Y_i - \widehat{\mu}_n - Z_i'\widehat{\theta})^2 D_i s(X_i'\widehat{\beta} - \gamma_n)}{\sum_{i=1}^n D_i s(X_i'\widehat{\beta} - \gamma_n)}. \tag{4.4}$$

To establish consistency of $\widehat{\sigma}_n^2$ for σ^2, we impose a stronger moment condition on the errors U_i than that given in Assumption 1.

ASSUMPTION 10: $EU_i^4 < \infty$.

Consistency of $\widehat{\sigma}_n^2$ is established in the following theorem.

THEOREM 4: *Under Assumptions* 1–7 *and* 10, $\widehat{\sigma}_n^2 \xrightarrow{p} \sigma^2$.

For a consistent estimator of Ω, we use the results given in the literature for the particular choice of preliminary estimators $(\widehat{\theta}, \widehat{\beta})$. Assumption 11 simply formalizes the existence of such an estimator.

ASSUMPTION 11: $\widehat{\Omega} \xrightarrow{p} \Omega$.

The normalizing factors $ED_i s(W_i - \gamma_n)$ and $ED_i s^2(W_i - \gamma_n)$ can be estimated by their sample analogues $\frac{1}{n}\sum_{i=1}^n D_i s(W_i - \gamma_n)$ and $\frac{1}{n}\sum_{i=1}^n D_i s^2(W_i - \gamma_n)$ respectively. To show that the ratio of $ED_i s^2(W_i - \gamma_n)$ to its sample analogue converges in probability to 1 as $n \to \infty$, we need to impose Assumption 7* rather than Assumption 7.

The joint asymptotic normality of $(\widehat{\mu}_n, \widehat{\theta}, \widehat{\beta})$ with estimated covariance matrix is given in the following theorem.

Theorem 5: *Under Assumptions* 1–4, 5*, 6, 7*, 10, *and* 11,

$$\begin{pmatrix} \dfrac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_i s(X_i'\widehat{\beta}-\gamma_n)}{\widehat{\sigma}_n\left(\frac{1}{n}\sum_{i=1}^{n} D_i s^2(X_i'\widehat{\beta}-\gamma_n)\right)^{1/2}}(\widehat{\mu}_n-\mu_0) \\ \sqrt{n}\widehat{\Omega}^{-1/2}\binom{\widehat{\theta}-\theta_0}{\widehat{\beta}-\beta_0} \end{pmatrix} \xrightarrow{d} N(0,I) \text{ iff Assumptions 8 and 9}$$

hold.

5 Dependence Between Errors and Regressors

In the previous section, we established consistency and asymptotic normality of $\widehat{\mu}_n$ under the assumption of independence of (U_i, ε_i) and (Z_i, X_i). This assumption can be restrictive. Some semiparametric estimators of (θ_0, β_0) allow for conditional heteroskedasticity of the errors (U_i, ε_i), although only in a restricted form. (They require that the distribution of the errors depends on X_i only through the single index $X_i'\beta_{0.}$) Here we extend the results of the previous section to incorporate conditional heteroskedasticity.

5.1 Consistency

First we state revisions of Assumptions 1 and 2 that are used to establish consistency. Let $\mathcal{W}$ be the support of W_i $(= X_i'\beta_0)$.

Assumption 1*: $(Z_i, X_i, U_i, \varepsilon_i)$ *are* iid rv's *with* $E\|Z_i\|^p < \infty$, $E\|X_i\|^p < \infty$, $E\|\max(|U_i|, 1)\cdot X_i\|^p < \infty$, *and* $\sup_{w\in\mathcal{W}} E(|U_i|^\lambda |W_i = w) < \infty$ *for some* $p > 3$ *and* $\lambda > 2$.

Assumption 2*: (a) $E(U_i|W_i) = 0$ a.s.
(b) $\sup_{w\in\mathcal{W}} P(\varepsilon_i \geq \gamma|W_i = w) \to 0$ *as* $\gamma \to \infty$.

Compared to Assumption 1, Assumption 1* imposes stronger moment conditions on U_i and X_i and places a restriction on the joint distribution of (U_i, W_i). Assumption 2* restricts the dependence of U_i and W_i by requiring the conditional mean of U_i to be zero almost surely. In addition, a restriction is made on the joint distribution of (ε_i, W_i).

An analogue of Lemma 1 is given by

Lemma 1*: (a) *Under Assumptions* 1*, 2*(b) , 3′, *and* 6, *Assumption* 7* *implies Assumption* 7.
(b) *Under Assumptions* 4 *and* 6, *Assumption* 7** *implies Assumption* 7*.

The consistency result is given by

Theorem 1*: *Under Assumptions* 1*, 2*, *and* 3–7, $\hat{\mu}_n \xrightarrow{p} \mu_0$.

The proofs of Lemma 1*, Theorem 1*, and other results of this section are given in Appendix B.

5.2 Asymptotic Normality

To obtain asymptotic normality of $\widehat{\mu}_n$ with dependence between the errors and regressors we need to add a new assumption and revise the sufficient condition Assumption 9* for Assumption 9.

ASSUMPTION 12: $\inf_{w\in\mathcal{W}} E(U_i^2|W_i = w) > 0$.

ASSUMPTION 9**: $\sqrt{n}\sup_{w\in\mathcal{W}} E(|U_i|\,1(\varepsilon_i > \gamma_n)|W_i = w)P(W_i > \gamma_n)^{(1-\xi)/2} \to 0$ *for* $\xi \in [0, 1/3]$ *as in Assumption* 4.

Assumption 12 ensures that $\mathrm{Var}(U_i D_i s(W_i - \gamma_n))$ is positive.

An analogue of Lemma 2 is now given by

LEMMA 2*: (a) *Under Assumptions* 1*, 2*, 3′, 4, 6, *and* 12*, Assumption* 7* *implies Assumption* 8*.*

(b) *Under Assumptions* 1*, 2*, 3′, 4, 6, *and* 12*, Assumption* 9** *implies Assumption* 9*.*

The asymptotic normality result is given by

THEOREM 2*: *Under Assumptions* 1*, 2*, 3–7, *and* 12,

(a) $\dfrac{\sqrt{n}ED_is(W_i-\gamma_n)}{\mathrm{Var}(U_iD_is(W_i-\gamma_n))^{1/2}}\left(\hat{\mu}_n-\mu_0-\dfrac{EU_iD_is(W_i-\gamma_n)}{ED_is(W_i-\gamma_n)}\right)\xrightarrow{d} N(0,1)$ *iff Assumption 8 holds and*

(b) $\dfrac{\sqrt{n}ED_is(W_i-\gamma_n)}{\mathrm{Var}(U_iD_is(W_i-\gamma_n))^{1/2}}(\hat{\mu}_n-\mu_0)\xrightarrow{d} N(0,1)$ *iff Assumptions* 8 *and* 9 *hold.*

The asymptotic joint normality result is given by

THEOREM 3*: *Under Assumptions* 1*, 2*, 3–4, 5*, 6–7, *and* 12,

$$\begin{pmatrix} \dfrac{\sqrt{n}ED_is(W_i-\gamma_n)}{\mathrm{Var}(U_iD_is(W_i-\gamma_n))^{1/2}}(\hat{\mu}_n-\mu_0) \\ \sqrt{n}\Omega^{-1/2}\binom{\hat{\theta}-\theta_0}{\hat{\beta}-\beta_0} \end{pmatrix} \xrightarrow{d} N(0,I) \text{ iff Assumptions 8 and 9 hold.}$$

Note that the normalization terms differ from those of Theorems 2 and 3. If we impose the rather strong assumption that the variance of U_i conditional on W_i equals σ^2 a.s., we get normalization terms that are identical to those obtained in the absence of dependence between (U_i, ε_i) and (X_i, Z_i).

5.3 Asymptotic Normality with Unknown Covariance Matrix

To compute standard errors and test statistics under the assumed dependence of the errors and regressors, one needs a consistent estimator of $\mathrm{Var}(U_i D_i s(W_i - \gamma_n))$.

Define

$$\widehat{V}_n = \frac{\sum_{i=1}^n (Y_i - \hat{\mu}_n - Z_i'\widehat{\theta})^2 D_i s^2(X_i'\widehat{\beta} - \gamma_n)}{\sum_{i=1}^n D_i s^2(X_i'\widehat{\beta} - \gamma_n)}. \tag{5.1}$$

$\widehat{V}_n$ is a consistent estimator of $\mathrm{Var}(U_i D_i s(W_i - \gamma_n))/E(D_i s^2(W_i - \gamma_n))$. To establish consistency of this estimator, we again impose stronger moment conditions on the errors U_i.

Assumption 10*: $\sup_{w \in \mathcal{W}} E(|U_i|^\nu \,|W_i = w) < \infty$ *for some* $\nu > 4$ *and* $E\left\|U_i^2 X_i\right\|^p < \infty$ *for* p *as in Assumption* 1*.

Consistency of $\widehat{V}_n$ is established in the following theorem.

Theorem 4*: *Under Assumptions* 1*, 2*, 3–6, 7*, 10*, *and* 12, $\widehat{V}_n E(D_i s^2(W_i - \gamma_n))/\mathrm{Var}(U_i D_i s(W_i - \gamma_n)) \xrightarrow{p} 1$.

The joint asymptotic normality of $(\widehat{\mu}_n, \widehat{\theta}, \widehat{\beta})$ with estimated covariance matrix is given by

Theorem 5*: *Under Assumptions* 1*, 2*, 3–4, 5*, 6, 7*, 10*, *and* 11–12,
$$\begin{pmatrix} \dfrac{\frac{1}{\sqrt{n}} \sum_{i=1}^n D_i s(X_i'\widehat{\beta} - \gamma_n)}{\widehat{V}_n^{1/2} \left(\frac{1}{n} \sum_{i=1}^n D_i s^2(X_i'\widehat{\beta} - \gamma_n)\right)^{1/2}} (\hat{\mu}_n - \mu_0) \\ \sqrt{n}\widehat{\Omega}^{-1/2} \binom{\hat{\theta} - \theta_0}{\hat{\beta} - \beta_0} \end{pmatrix} \xrightarrow{d} N(0, I)$$ *iff Assumptions* 8 *and* 9 *hold.*

6 Conclusions

In this paper we provide a consistent asymptotically normal estimator for the intercept of a semiparametrically estimated sample selection model. This parameter is of importance in determining a variety of economically interesting quantities as discussed in the second section. The estimation of this intercept has up to now been absorbed in the nonparametric sample selectivity bias correction term, with the exception of the estimator given by Gallant and Nychka. Their estimator is consistent, but its asymptotic distribution is unknown. Therefore, this paper provides a useful contribution to the literature on semiparametric sample selection models. We also present a consistent estimator of the asymptotic covariance matrix for $(\widehat{\mu}_n, \widehat{\theta}, \widehat{\beta})$.

In a simulation study, Schafgans (1995) compares our estimator to the estimator of the intercept when standard parametric estimation techniques are applied. In particular, the two-step estimator of Heckman (1976) is considered. Using a root mean squared error criterion, our estimator performs better for a range of bandwidth parameter choices for a variety of distributions of the errors and regressors. For error distributions that are close to normal, however, the two-step parametric estimator performs better.

APPENDIX OF PROOFS (A)

For notational simplicity, we let s_{ni} and $\widehat{s}_{ni}$ abbreviate $s(W_i - \gamma_n)$ and $s(X_i'\widehat{\beta} - \gamma_n)$, respectively, in the proofs below.

The proofs below use the following lemmas:

Lemma A–1: *Under Assumptions* 1 *and* 7, $\frac{\frac{1}{n}\sum_1^n D_i s(W_i-\gamma_n)}{ED_i s(W_i-\gamma_n)} \xrightarrow{p} 1$.

Lemma A–2: *Under Assumptions* 2(b), 3′, *and* 6, $\frac{ED_i s(W_i-\gamma_n)}{Es(W_i-\gamma_n)} \to 1$, $\frac{ED_i s^2(W_i-\gamma_n)}{Es^2(W_i-\gamma_n)} \to 1$, *and* $\frac{ED_i s^4(W_i-\gamma_n)}{Es^4(W_i-\gamma_n)} \to 1$.

Lemma A–3: *Under Assumptions* 1, 2(b), 3′, *and* 6, $\frac{EU_i^2 D_i s^2(W_i-\gamma_n)}{EU_i^2 s^2(W_i-\gamma_n)} \to 1$. *If Assumption* 4 *also holds,* $\frac{\mathrm{Var}(U_i D_i s(W_i-\gamma_n))}{\sigma^2 ED_i s^2(W_i-\gamma_n)} \to 1$.

Lemma A–4: *Under Assumptions* 1, 2(b), 3′, 6, *and* 7*, $\frac{\frac{1}{n}\Sigma_1^n D_i s^2(W_i-\gamma_n)}{ED_i s^2(W_i-\gamma_n)} \xrightarrow{p} 1$.

Proof of Lemma A–1: The rv $\frac{1}{n}\Sigma_1^n(D_i s_{ni} - ED_i s_{ni})/ED_i s_{ni}$ has mean zero and variance

$$\frac{\mathrm{Var}(D_i s_{ni})}{n(ED_i s_{ni})^2} = \frac{ED_i s_{ni}^2}{n(ED_i s_{ni})^2} - \frac{1}{n} \to 0 \qquad \text{(A.1)}$$

using Assumption 7. $\square$

Proof of Lemma A–2: We have

$$(1 - D_i)s_{ni} = 1(\gamma_n \le W_i \le \varepsilon_i)s_{ni} \le 1(\gamma_n \le \varepsilon_i)s_{ni} \qquad \text{(A.2)}$$

using the fact that $s_{ni} \ge 0$ only if $W_i - \gamma_n \ge 0$ by Assumption 3′. Thus,

$$\left|\frac{ED_i s_{ni}}{Es_{ni}} - 1\right| = \frac{E(1-D_i)s_{ni}}{Es_{ni}} \le \frac{E1(\gamma_n \le \varepsilon_i)s_{ni}}{Es_{ni}} = P(\gamma_n \le \varepsilon_i) \to 0 \qquad \text{(A.3)}$$

using Assumptions 2(b) and 6. The proof is identical with s_{ni} replaced by s_{ni}^2 or s_{ni}^4. $\square$

Proof of Lemma A–3: When s_{ni} in (A.3) is replaced by $U_i^2 s_{ni}^2$ the rhs becomes $EU_i^2 1(\gamma_n \le \varepsilon_i) \to 0$ using Assumption 1.

Next, we have

$$\begin{aligned}
\frac{\mathrm{Var}(U_i D_i s_{ni})}{\sigma^2 ED_i s_{ni}^2} &= \frac{EU_i^2 D_i s_{ni}^2}{\sigma^2 ED_i s_{ni}^2} - \frac{(EU_i D_i s_{ni})^2}{\sigma^2 ED_i s_{ni}^2}, \\
\frac{EU_i^2 D_i s_{ni}^2}{\sigma^2 ED_i s_{ni}^2} &= \frac{EU_i^2 s_{ni}^2}{\sigma^2 Es_{ni}^2}(1 + o(1)) = 1 + o(1), \quad \text{and} \qquad \text{(A.4)} \\
\frac{(EU_i D_i s_{ni})^2}{\sigma^2 ED_i s_{ni}^2} &\le \frac{(E|U_i|s_{ni})^2}{\sigma^2 Es_{ni}^2}(1+o(1)) \le \frac{(E|U_i|)^2 P(W_i \ge \gamma_n)^2}{\sigma^2 P(W_i > \gamma_n + b)}(1+o(1)) = o(1),
\end{aligned}$$

where the second and third equations use the results of the first part of this lemma, Lemma A–2, and Assumption 2(b) and the third equation also uses Assumptions 1, 4, and 6. □

Proof of Lemma A–4: The rv $\frac{1}{n}\Sigma_1^n(D_i s_{ni}^2 - ED_i s_{ni}^2)/ED_i s_{ni}^2$ has zero mean and variance

$$\frac{\mathrm{Var}(D_i s_{ni}^2)}{n(ED_i s_{ni}^2)^2} = \frac{ED_i s_{ni}^4}{n(ED_i s_{ni}^2)^2} - \frac{1}{n} \to 0 \tag{A.5}$$

using Assumptions 3$'$, 6, and 7* and Lemma A–2 since,

$$\begin{aligned} \frac{ED_i s_{ni}^4}{n(ED_i s_{ni}^2)^2} &\leq \frac{1}{nED_i s_{ni}^2} = \frac{1}{nEs_{ni}^2(1+o(1))} \\ &\leq \frac{1}{nP(W_i > \gamma_n + b)(1+o(1))} = o(1). \quad \square \end{aligned} \tag{A.6}$$

Proof of Lemma 1: Assumption 7* implies 7, because

$$\frac{\sqrt{n}ED_i s_{ni}}{(ED_i s_{ni}^2)^{1/2}} \geq (nED_i s_{ni})^{1/2} = (1+o(1))(nEs_{ni})^{1/2} \geq (1+o(1))(nP(W_i > \gamma_n + b))^{1/2}, \tag{A.7}$$

where both inequalities use Assumption 3$'$ and the equality uses Lemma A–2.

Assumption 7** implies 7* under Assumptions 4 and 6, because

$$\begin{aligned} &(nP(W_i > \gamma_n + b))^{-1/(1+\xi)} \\ = &(n^{1/(1+\xi)}P(W_i > \gamma_n))^{-1}(P(W_i > \gamma_n)^{1+\xi}/P(W_i > \gamma_n + b))^{1/(1+\xi)} \\ = &(n^{1/(1+\xi)}P(W_i > \gamma_n))^{-1}O_p(1). \quad \square \end{aligned} \tag{A.8}$$

Proof of Lemma 2: To establish part (a), suppose Assumption 7* holds. We have: $\forall \delta > 0$,

$$\begin{aligned} E|B_n - EB_n|^{2+2\delta} &\leq 4^{1+\delta}E|B_n|^{2+2\delta} = 4^{1+\delta}E|U_i D_i s_{ni}|^{2+2\delta} \\ &\leq 4^{1+\delta}E|U_i|^{2+2\delta}Es_{ni}^{2+2\delta} \leq 4^{1+\delta}E|U_i|^{2+2\delta}Es_{ni}^2, \end{aligned} \tag{A.9}$$

where the first inequality holds by Minkowski's inequality, the second inequality uses Assumption 2(b), and the last inequality uses Assumption 3$'$. Now, using Lemmas A–2 and A–3, for $0 < \delta \leq 1$, the ℓhs of Assumption 8 equals

$$\begin{aligned} &\lim_{n\to\infty} E\frac{(B_n - EB_n)^2}{\sigma^2 Es_{ni}^2(1+o(1))}1((B_n - EB_n)^2 > n\sigma^2 Es_{ni}^2\varepsilon(1+o(1))) \\ \leq &\lim_{n\to\infty}\frac{E|B_n - EB_n|^{2+2\delta}}{\sigma^{2+2\delta}\varepsilon^\delta n^\delta(Es_{ni}^2)^{1+\delta}} \\ \leq &\lim_{n\to\infty}\frac{4^{1+\delta}E|U_i|^{2+2\delta}}{\sigma^{2+2\delta}\varepsilon^\delta(nP(W_i > \gamma_n + b))^\delta} \to 0, \end{aligned} \tag{A.10}$$

where the second inequality uses (A.9) and the convergence to zero holds by Assumptions 1 and 7*.

Next, we establish part (b). By Assumption 2, $EU_i s_{ni} = 0$. Hence, the absolute value of the ℓhs of Assumption 9 equals

$$\begin{aligned}
&|\sqrt{n}EU_i(1-D_i)s_{ni}/(ED_i s_{ni}^2)^{1/2}| \\
\leq\ &(1+o(1))\sqrt{n}E|U_i|1(W_i<\varepsilon_i)1(W_i>\gamma_n)/(Es_{ni}^2)^{1/2} \\
\leq\ &(1+o(1))\sqrt{n}E|U_i|1(\varepsilon_i>\gamma_n)\cdot P(W_i>\gamma_n)/P(W_i>\gamma_n+b)^{1/2} \qquad \text{(A.11)} \\
=\ &(1+o(1))\sqrt{n}E|U_i|1(\varepsilon_i>\gamma_n)\cdot P(W_i>\gamma_n)^{(1-\xi)/2}(P(W_i>\gamma_n)^{1+\xi}/P(W_i>\gamma_n+b))^{1/2} \\
\to\ &0
\end{aligned}$$

where the first inequality uses Lemma A–2, the second inequality uses Assumptions 2(b) and 3′, and the convergence to zero uses Assumptions 4, 6, and 9*. □

Theorems 1, 2, and 3 follow from Theorems A–1 and A–2, A–1 and A–3, and A–1 and A–4, respectively, below.

THEOREM A–1: *Under Assumptions* 1, 2(b), *and* 3–7, $\dfrac{\sqrt{n}ED_i s(W_i-\gamma_n)}{(ED_i s^2(W_i-\gamma_n))^{1/2}}(\widehat{\mu}_n - \widehat{\mu}_{n0}) \xrightarrow{p} 0$.

THEOREM A–2: *Under Assumptions* 1, 2, 3′, 6, *and* 7, $\widehat{\mu}_{n0} \xrightarrow{p} \mu_0$.

THEOREM A–3: *Under Assumptions* 1, 2, 3′, 4, 6, *and* 7, *the results of parts* (a) *and* (b) *of Theorem* 2 *hold with* $\widehat{\mu}_n$ *replaced by* $\widehat{\mu}_{n0}$.

THEOREM A–4: *Under Assumptions* 1, 2, 3′, 4, 5*, 6, *and* 7, *the result of Theorem* 3 *holds with* $\widehat{\mu}_n$ *replaced by* $\widehat{\mu}_{n0}$.

PROOF OF THEOREM A–1: The ℓhs in the Theorem can be written as

$$\begin{aligned}
&C\left(\frac{\widehat{A}}{\widehat{B}}-\frac{A}{B}\right) = C\frac{\widehat{A}-A}{B}\,\frac{B}{\widehat{B}} - C\frac{\widehat{B}-B}{B}\,\frac{A}{B}\,\frac{B}{\widehat{B}}\ , \quad \text{where} \\
&C = \frac{\sqrt{n}ED_i s_{ni}}{(ED_i s_{ni}^2)^{1/2}}\ ,\quad A = \Sigma_1^n(Y_i - Z_i'\theta_0)D_i s_{ni}\ ,\quad \widehat{A} = \Sigma_1^n(Y_i - Z_i'\widehat{\theta})D_i\widehat{s}_{ni}, \qquad \text{(A.12)} \\
&B = \Sigma_1^n D_i s_{ni}\ ,\quad \text{and}\quad \widehat{B} = \Sigma_1^n D_i \widehat{s}_{ni}.
\end{aligned}$$

To show that the ℓhs $\xrightarrow{p} 0$, it suffices to show (i) $\widehat{B}/B \xrightarrow{p} 1$, (ii) $C(\widehat{A}-A)/B \xrightarrow{p} 0$, (iii) $A/B = O_p(1)$, and (iv) $C(\widehat{B}-B)/B \xrightarrow{p} 0$. Since $C \to \infty$ by Assumption 7, condition (i) follows from (iv). Using Lemma A–1, we find that the following two conditions (established below) are sufficient for (ii):

$$\frac{1}{\sqrt{n}}\Sigma_1^n(U_i+\mu_0)D_i(\widehat{s}_{ni}-s_{ni})/(ED_i s_{ni}^2)^{1/2} \xrightarrow{p} 0 \ \text{ and} \qquad \text{(A.13)}$$

$$(\widehat{\theta}-\theta_0)'\frac{1}{\sqrt{n}}\Sigma_1^n Z_i D_i \widehat{s}_{ni}/(ED_i s_{ni}^2)^{1/2} \xrightarrow{p} 0. \qquad \text{(A.14)}$$

By Lemmas A–1 and A–2, the following are sufficient for (iii) and (iv) respectively:

$$\frac{1}{n}\Sigma_1^n U_i D_i s_{ni}/E s_{ni} = O_p(1) \text{ and} \tag{A.15}$$

$$\frac{1}{\sqrt{n}}\Sigma_1^n D_i(\widehat{s}_{ni} - s_{ni})/(E s_{ni}^2)^{1/2} \xrightarrow{p} 0. \tag{A.16}$$

We establish (A.15) first. We have

$$\left|\frac{\frac{1}{n}\Sigma_1^n U_i D_i s_{ni}}{E s_{ni}}\right| \leq \left|\frac{\frac{1}{n}\Sigma_1^n (U_i D_i s_{ni} - E U_i D_i s_{ni})}{E D_i s_{ni}}\right| + \frac{E|U_i| s_{ni}}{E s_{ni}}. \tag{A.17}$$

The first term on the rhs is $O_p(1)$, because it has mean zero and variance $\sigma^2 E D_i s_{ni}^2/(n(E D_i s_{ni})^2) = o(1)$ using Lemmas A–2 and A–3 and Assumption 7. The second term on the rhs equals $E|U_i|$ using Assumption 2(b). Using Assumption 1, therefore, (A.15) holds.

Next, we establish (A.16). For notational simplicity, suppose $\widehat{\beta}$ and β_0 are scalars. The argument carries over to the vector case. Let G_n denote the ℓhs of (A.16). A three-term Taylor expansion of $s(X_i\widehat{\beta} - \gamma_n)$ about β_0 gives

$$\begin{aligned}|G_n| \leq & \left|\frac{1}{n}\Sigma_1^n X_i D_i s'(X_i\beta_0 - \gamma_n)\sqrt{n}(\widehat{\beta} - \beta_0)/(E s_{ni}^2)^{1/2}\right| \\ & + \left|\frac{1}{2n}\Sigma_1^n X_i^2 D_i s''(X_i\beta_0 - \gamma_n)[\sqrt{n}(\widehat{\beta} - \beta_0)]^2/(n E s_{ni}^2)^{1/2}\right| \\ & + \left|\frac{1}{6n}\Sigma_1^n X_i^3 D_i s'''(X_i\beta_* - \gamma_n)[\sqrt{n}(\widehat{\beta} - \beta_0)]^3/(n^2 E s_{ni}^2)^{1/2}\right| \\ \leq & \; O_p(1)\frac{1}{n}\Sigma_1^n |X_i| 1(X_i\beta_0 > \gamma_n)/(E s_{ni}^2)^{1/2} + \\ & + O_p(1)\frac{1}{n}\Sigma_1^n X_i^2 1(X_i\beta_0 > \gamma_n)/(n E s_{ni}^2)^{1/2} + O_p(1)\frac{1}{n}\Sigma_1^n |X_i|^3/(n^2 E s_{ni}^2)^{1/2},\end{aligned} \tag{A.18}$$

where $s'(\cdot)$ denotes the first derivative of $s(\cdot)$, etc., β_* is on the line segment joining $\widehat{\beta}$ and β_0, and the second inequality uses Assumptions 3 and 5 and the inequalities $|s'(x)| \leq 1(x > 0)\sup_{x_* \in R}|s'(x_*)|$, $|s''(x)| \leq 1(x > 0)\sup_{x_* \in R}|s''(x_*)|$, and $\sup_{x \in R}|s'''(x)| < \infty$.

For any identically distributed rv's $\{H_i : i \geq 1\}$, with $0 < E|H_i| < \infty$, $\left|\frac{1}{n}\Sigma_1^n H_i\right|/E|H_i| = O_p(1)$ by Markov's inequality. Thus, the first term on the rhs of (A.18) is

$$O_p(1)\frac{E|X_i| 1(W_i > \gamma_n)}{(E s_{ni}^2)^{1/2}} \leq O_p(1)\frac{(E|X_i|^p)^{1/p} P(W_i > \gamma_n)^{1/q}}{P(W_i > \gamma_n + b)^{1/2}} = o_p(1), \tag{A.19}$$

where $1/p + 1/q = 1$, p is as in Assumption 1, and the equality holds by Assumptions 4 and 6 since $2/q > 1 + \xi$.

The second term on the rhs of (A.18) is

$$O_p(1)\frac{EX_i^2 1(W_i>\gamma_n)}{(nEs_{ni}^2)^{1/2}} \le O_p(1)\frac{(E|X_i|^3)^{2/3}P(W_i>\gamma_n)^{1/3}Es_{ni}}{(n(Es_{ni})^2/Es_{ni}^2)^{1/2}Es_{ni}^2} \tag{A.20}$$
$$\le O_p(1)\left(\frac{n(Es_{ni})^2}{Es_{ni}^2}\right)^{-1/2}\frac{P(W_i>\gamma_n)^{1+1/3}}{P(W_i>\gamma_n+b)} = o_p(1),$$

where the equality holds by Lemma A–2 and Assumptions 4, 6, and 7.

The third term on the rhs of (A.18) is

$$O_p(1)\frac{E|X_i|^3}{(n^2Es_{ni}^2)^{1/2}} = O_p(1)\left(\frac{n(Es_{ni})^2}{Es_{ni}^2}\right)^{-1}\frac{(Es_{ni})^2}{(Es_{ni}^2)^{3/2}} \tag{A.21}$$
$$\le O_p(1)\left(\frac{n(Es_{ni})^2}{Es_{ni}^2}\right)^{-1}\left(\frac{P(W_i>\gamma_n)^{4/3}}{P(W_i>\gamma_n+b)}\right)^{3/2} = o_p(1).$$

The first equality uses Assumption 1 and the last equality uses Lemma A–2 and Assumptions 4, 6, and 7 and requires that $\xi \le 1/3$ in Assumption 4. This completes the proof of (A.16).

The proof of (A.13) is the same as that of (A.16) except that the factor $|U_i+\mu_0|$ appears in various sums and expectations. In particular, using Assumption 2(b), each expression in (A.19)–(A.20) is multiplied by $E|U_i+\mu_0|$.

We now establish (A.14). By Assumption 5 and a two-term Taylor expansion about β_0, the absolute value of the ℓhs of (A.14) is bounded by

$$\begin{aligned}
&O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|\widehat{s}_{ni}}{(Es_{ni}^2)^{1/2}}\\
&\le O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|s_{ni}}{(Es_{ni}^2)^{1/2}} + O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|\cdot|X_i|\cdot|s'(X_i\beta_0-\gamma_n)|\cdot|\sqrt{n}(\widehat{\beta}-\beta_0)|}{2(nEs_{ni}^2)^{1/2}}\\
&\quad + O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|X_i^2|s''(X_i\beta_*-\gamma_n)|\cdot|\sqrt{n}(\widehat{\beta}-\beta_0)|^2}{6(n^2Es_{ni}^2)^{1/2}} &\text{(A.22)}\\
&\le O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|1(W_i>\gamma_n)}{(Es_{ni}^2)^{1/2}} + O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|\cdot|X_i|1(W_i>\gamma_n)}{(nEs_{ni}^2)^{1/2}}\\
&\quad + O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|X_i^2}{(n^2Es_{ni}^2)^{1/2}},
\end{aligned}$$

where β_* lies on the line segment joining $\widehat{\beta}$ and β_0 and the inequalities hold for the same reasons as in (A.18). The rhs of (A.22) is the same as the rhs of (A.18) except that $|X_i|$, X_i^2, and $|X_i|^3$ are replaced by $\|Z_i\|$, $\|Z_i\|\cdot|X_i|$, and $\|Z_i\|\cdot X_i^2$ respectively. In consequence, the rhs of (A.22) is $o_p(1)$ by the same proof as for (A.18), provided the following moment conditions hold: $E\|Z_i\|^p<\infty$ for $p>3$, $E\|Z_i\|^{3/2}\cdot|X_i|^{3/2}<\infty$, and $E\|Z_i\|X_i^2<\infty$. The latter hold by Assumption 1 and the proof of (A.14) is complete. $\square$

Proof of Theorem A–2: We have

$$\widehat{\mu}_{n0} - \mu_0 = \frac{\frac{1}{n}\Sigma_1^n U_i D_i s_{ni}}{E D_i s_{ni}} \times \frac{E D_i s_{ni}}{\frac{1}{n}\Sigma_1^n D_i s_{ni}}. \tag{A.23}$$

The second multiplicand on the rhs is $1+o_p(1)$ by Lemma A–1. The first multiplicand on the rhs has mean $EU_iD_is_{ni}/ED_is_{ni}$, whose absolute value by Assumption 2 is

$$\frac{|EU_i(1-D_i)s_{ni}|}{ED_is_{ni}} \leq \frac{E|U_i|1(\gamma_n < \varepsilon_i)s_{ni}}{Es_{ni}(1+o(1))} = E|U_i|1(\gamma_n < \varepsilon_i)(1+o(1)) = o(1), \tag{A.24}$$

where the inequality uses (A.2) and Lemma A–2, the first equality uses Assumption 2, and the second equality uses Assumptions 1 and 6. The first multiplicand on the rhs of (A.23) has variance equal to

$$\frac{\mathrm{Var}(U_iD_is_{ni})}{n(ED_is_{ni})^2} = \frac{\sigma^2 ED_is_{ni}^2(1+o(1))}{n(ED_is_{ni})^2} \to 0 \tag{A.25}$$

using Lemma A–2 and Assumption 7. □

Proof of Theorem A–3: For part (a), we write

$$\widehat{\mu}_{n0} - \mu_0 = \frac{\Sigma_1^n U_i D_i s_{ni}}{\Sigma_1^n D_i s_{ni}}, \quad \text{and so,} \tag{A.26}$$

$$\begin{aligned} &\frac{\sqrt{n}ED_is_{ni}}{\sigma(ED_is_{ni}^2)^{1/2}}\left(\widehat{\mu}_{n0} - \mu_0 - \frac{EU_iD_is_{ni}}{\frac{1}{n}\Sigma_1^n D_is_{ni}}\right) \\ &= \frac{ED_is_{ni}\mathrm{Var}^{1/2}(U_iD_is_{ni})}{\sigma(ED_is_{ni}^2)^{1/2}\frac{1}{n}\Sigma_1^n D_is_{ni}} \times \frac{1}{\sqrt{n}}\Sigma_1^n \frac{U_iD_is_{ni} - EU_iD_is_{ni}}{\mathrm{Var}^{1/2}(U_iD_is_{ni})}. \end{aligned} \tag{A.27}$$

The second multiplicand of the rhs is asymptotically $N(0,1)$ by the CLT for "infinitesimal" independent non-identically distributed rv's (and its converse) iff the Lindeberg condition (Assumption 8) holds, see Chow and Teicher (1978, Cor. 12.2.2, p. 434). (We note that the summands of the triangular array are infinitesimal, because $\sup_{1\leq i\leq n} P(\frac{1}{\sqrt{n}}(U_iD_is_{ni} - EU_iD_is_{ni})/\mathrm{Var}^{1/2}(U_iD_is_{ni}) > \varepsilon) \leq 1/(n\varepsilon^2) \to 0$ using Markov's inequality.) The first multiplicand on the rhs of (A.27) equals $1 + o_p(1)$ by Lemmas A–1 and A–3.

For part (a), it remains to show that

$$\frac{\sqrt{n}ED_is_{ni}}{\sigma(ED_is_{ni}^2)^{1/2}}\left(\frac{EU_iD_is_{ni}}{\frac{1}{n}\Sigma_1^n D_is_{ni}} - \frac{EU_iD_is_{ni}}{ED_is_{ni}}\right) = o_p(1). \tag{A.28}$$

The ℓhs of (A.28) equals

$$-\frac{EU_iD_is_{ni}}{\sigma(ED_is_{ni}^2)^{1/2}}\left(\frac{\frac{1}{\sqrt{n}}\Sigma_1^n(D_is_{ni} - ED_is_{ni})}{\frac{1}{n}\Sigma_1^n D_is_{ni}}\right)$$

$$= -\frac{EU_iD_is_{ni}}{\sigma ED_is_{ni}}\left(\frac{1}{\sqrt{n}}\Sigma_1^n\frac{D_is_{ni}-ED_is_{ni}}{(ED_is_{ni}^2)^{1/2}}\right)(1+o_p(1)) \tag{A.29}$$
$$= -\frac{EU_iD_is_{ni}}{\sigma ED_is_{ni}}O_p(1) = o_p(1),$$

where the first equality uses Lemma A–1, the second equality holds because the second multiplicand has mean zero and variance $\leq 1 \ \forall n \geq 1$, and the third equality holds by (A.24). This completes the proof of part (a).

Part (b) follows from part (a), because the ℓhs of part (b) differs from that of part (a) by a non-stochastic quantity that goes to zero as $n \to \infty$ if and only if Assumption 9 holds. □

Proof of Theorem A–4: The converse holds by the converse result of Theorem A–3(b).

Hence, we suppose Assumptions 8 and 9 hold. We use the Cramer–Wold device and the proof of Theorem A–3. Let $c = (c_1, c_2')'$ be an arbitrary unit vector. We need to show that

$$\frac{\sqrt{n}ED_is_{ni}}{\sigma(ED_is_{ni}^2)^{1/2}}c_1(\widehat{\mu}_{n0}-\mu_0) + \sqrt{n}c_2'\Omega^{-1/2}\begin{pmatrix}\widehat{\theta}-\theta_0\\ \widehat{\beta}-\beta_0\end{pmatrix} \xrightarrow{d} N(0,1). \tag{A.30}$$

By the proof of Theorem A–3 and Assumptions 5* and 9, the ℓhs above equals

$$(1+o_p(1))\sum\nolimits_1^n\left(c_1\frac{U_iD_is_{ni}-EU_iD_is_{ni}}{\mathrm{Var}^{1/2}(U_iD_is_{ni})} + c_2'\Omega^{-1/2}Q_i\right)/\sqrt{n} + o_p(1). \tag{A.31}$$

The summands are easily seen to be infinitesimal and we show below that the sum of their variances equals $1+o(1)$, so the same CLT as used in the proof of Theorem A–3 applies here provided the Lindeberg condition holds.

We now compute the variance of the summands of (A.31):

$$\begin{aligned}&\mathrm{Var}\left(c_1\left(\frac{U_iD_is_{ni}-EU_iD_is_{ni}}{\mathrm{Var}^{1/2}(U_iD_is_{ni})}\right) + c_2'\Omega^{-1/2}Q_i\right)\\ &= 1 + 2\mathrm{Cov}\left(\frac{c_1U_iD_is_{ni}}{\mathrm{Var}^{1/2}(U_iD_is_{ni})},\ c_2'\Omega^{-1/2}Q_i\right)\ ,\end{aligned} \tag{A.32}$$

and

$$\begin{aligned}&\left|\mathrm{Cov}\left(\frac{c_1U_iD_is_{ni}}{\mathrm{Var}^{1/2}(U_iD_is_{ni})},\ c_2'\Omega^{-1/2}Q_i\right)\right| = \frac{|c_1EU_iD_is_{ni}c_2'\Omega^{-1/2}Q_i|}{\mathrm{Var}^{1/2}(U_iD_is_{ni})}\\ &\leq \frac{(E|U_ic_2'\Omega^{-1/2}Q_i|^p)^{1/p}(E|s_{ni}|^q)^{1/q}}{\sigma(Es_{ni}^2)^{1/2}}(1+o(1))\\ &\leq \sigma^{-1}(E|U_ic_2'\Omega^{-1/2}Q_i|^p)^{1/p}\left(\frac{P(W_i>\gamma_n)^{2/q}}{P(W_i>\gamma_n+b)}\right)^{1/2}(1+o(1)) = o(1),\end{aligned} \tag{A.33}$$

where the first inequality uses Lemmas A–2 and A–3, p is as in Assumption 5*, $1/p+1/q=1$, and the last equality holds by Assumptions 4, 5*, and 6 since $2/q>1+\xi$.

It remains to verify the Lindeberg condition. Let

$$A_n = c_1(U_i D_i s_{ni} - EU_i D_i s_{ni})/\mathrm{Var}^{1/2}(U_i D_i s_{ni}) \ \text{ and } \ Q = c_2'\Omega^{-1/2}Q_i\ . \qquad \text{(A.34)}$$

Assumption 8 implies the Lindeberg condition holds for $\{A_n : n\geq 1\}$:

$$\lim_{n\to\infty} EA_n^2 1(A_n^2 \geq n\varepsilon) = 0 \ \ \forall \varepsilon > 0. \qquad \text{(A.35)}$$

We need to show that the Lindeberg condition holds for $\{A_n+Q : n\geq 1\}$:

$$\lim_{n\to\infty} E(A_n+Q)^2 1((A_n+Q)^2 > n\varepsilon) = 0 \ \ \forall\varepsilon>0. \qquad \text{(A.36)}$$

The left-hand side of (A.36) is less than or equal to

$$\lim_{n\to\infty} E(2A_n^2+2Q^2)1(2A_n^2+2Q^2>n\varepsilon). \qquad \text{(A.37)}$$

We consider the two summands of (A.37) separately:

$$\begin{aligned} &\lim_{n\to\infty} EQ^2 1(2A_n^2+2Q^2>n\varepsilon) \leq \lim_{n\to\infty}(E|Q|^\lambda)^{2/\lambda} P(2A_n^2+2Q^2>n\varepsilon)^{\lambda/(\lambda-2)} \\ &\leq (E|Q|^\lambda)^{2/\lambda}\lim_{n\to\infty}\left(\frac{2EA_n^2+2EQ^2}{n\varepsilon}\right)^{\lambda/(\lambda-2)} = 0, \end{aligned} \qquad \text{(A.38)}$$

where the equality holds because $EA_n^2=1$ and $E|Q|^\lambda<\infty$ by Assumption 5*. Next, we have

$$\begin{aligned} &\lim_{n\to\infty} EA_n^2 1(A_n^2+Q^2>n\varepsilon/2) \\ &\leq \lim_{n\to\infty} EA_n^2 1(Q^2>M_n) + \lim_{n\to\infty} EA_n^2 1(A_n^2+Q^2>n\varepsilon/2,\, Q^2\leq M_n) \\ &\leq \lim_{n\to\infty} EA_n^2 1(A_n^2\geq n, Q^2>M_n) + \lim_{n\to\infty} EA_n^2 1(A_n^2\leq n, Q^2>M_n) \\ &\quad + \lim_{n\to\infty} EA_n^2 1(A_n^2>n\varepsilon/2-M_n) \qquad \text{(A.39)} \\ &\leq \lim_{n\to\infty} EA_n^2 1(A_n^2\geq n) + \lim_{n\to\infty} nP(Q^2>M_n) + \lim_{n\to\infty} EA_n^2 1(A_n^2>n\varepsilon/2-M_n) \\ &\leq \lim_{n\to\infty} nE|Q|^\lambda/M_n^{\lambda/2} + \lim_{n\to\infty} EA_n^2 1(A_n^2>n\varepsilon/2-M_n) = 0, \end{aligned}$$

where the last inequality holds by the Lindeberg condition for $\{A_n : n\geq 1\}$ and Markov's inequality, $\lambda>2$ is as in Assumption 5*, and the equality holds by taking $M_n = n\varepsilon/4$ using Assumption 5* and the Lindeberg condition for $\{A_n : n\geq 1\}$. This completes the proof of the Lindeberg condition for $\{A_n+Q : n\geq 1\}$. □

Theorem 4 follows from Theorems A–5 and A–6 below. Define

$$\widehat{\sigma}_{n0}^2 = \frac{\Sigma_1^n U_i^2 D_i s_{ni}}{\Sigma_1^n D_i s_{ni}}. \qquad \text{(A.40)}$$

Theorem A–5: *Under Assumptions* 1, 2(b), 3′, 4, 6, 7, *and* 10, $\widehat{\sigma}^2_{n0} \xrightarrow{p} \sigma^2$.

Theorem A–6: *Under Assumptions* 1, 2(b), 3–7, *and* 10, $\widehat{\sigma}^2_n - \widehat{\sigma}^2_{n0} \xrightarrow{p} 0$.

Proof of Theorem A–5: By Lemma A–1,

$$\widehat{\sigma}^2_{n0} = \frac{\frac{1}{n}\Sigma_1^n U_i^2 D_i s_{ni}}{ED_i s_{ni}}(1+o_p(1)). \tag{A.41}$$

Let $V_{n0} = \frac{1}{n}\Sigma_1^n U_i^2 D_i s_{ni}/ED_i s_{ni}$. We have

$$EV_{n0} = \frac{EU_i^2 D_i s_{ni}}{ED_i s_{ni}} = \frac{EU_i^2 s_{ni}(1+o(1))}{ED_i s_{ni}} = \frac{\sigma^2 E s_{ni}(1+o(1))}{ED_i s_{ni}} = \sigma^2(1+o(1)) \tag{A.42}$$

using Lemmas A–2 and A–3 and Assumption 2(b). We also have

$$\mathrm{Var}(V_{n0}) = \frac{\mathrm{Var}(U_i^2 D_i s_{ni})}{n(ED_i s_{ni})^2} \leq \frac{EU_i^4 s_{ni}^2}{n(ED_i s_{ni})^2} = \frac{EU_i^4 ED_i s_{ni}^2(1+o(1))}{n(ED_i s_{ni})^2} = o(1), \tag{A.43}$$

where the second to last equality holds by Lemma A–2 and Assumption 2(b) and the last equality holds by Assumptions 7 and 10. □

Proof of Theorem A–6: The ℓhs in this theorem can be written as:

$$\begin{aligned} &\left(\frac{\widehat{A}}{\widehat{B}} - \frac{A}{B}\right) = \frac{\widehat{A}-A}{B}\cdot\frac{B}{\widehat{B}} - \frac{\widehat{B}-B}{B}\cdot\frac{A}{B}\cdot\frac{B}{\widehat{B}}, \quad \text{where} \\ &A = \Sigma_1^n U_i^2 D_i s_{ni}, \quad \widehat{A} = \Sigma_1^n (Y_i - \widehat{\mu}_n - Z_i'\widehat{\theta})^2 D_i \widehat{s}_{ni}, \\ &B = \Sigma_1^n D_i s_{ni}, \quad \text{and} \quad \widehat{B} = \Sigma_1^n D_i \widehat{s}_{ni}. \end{aligned} \tag{A.44}$$

To show that the ℓhs $\xrightarrow{p} 0$, it suffices to show that: (i) $\widehat{B}/B \xrightarrow{p} 1$, and (ii) $(\widehat{A}-A)/B \xrightarrow{p} 0$, and (iii) $A/B = O_p(1)$. By the proof of Theorem A–1, (i) holds. (It is implied by condition (iv) of that proof.) By Theorem A–5, $A/B \xrightarrow{p} \sigma^2$, so (iii) is satisfied.

Using Lemmas A–1, A–2, and A–3, Assumptions 3, 5, and 7, and Theorem 1, we find that the following six conditions are sufficient for (ii):

$$\frac{\frac{1}{\sqrt{n}}\Sigma_1^n U_i^2 D_i(\widehat{s}_{ni} - s_{ni})}{(ED_i s_{ni}^2)^{1/2}} \xrightarrow{p} 0, \tag{A.45}$$

$$\frac{\frac{1}{n}\Sigma_1^n D_i \widehat{s}_{ni}}{E s_{ni}} = O_p(1), \tag{A.46}$$

$$\frac{(\widehat{\theta}-\theta_0)'\frac{1}{n}\Sigma_1^n Z_i D_i \widehat{s}_{ni}}{E s_{ni}} = O_p(1), \tag{A.47}$$

$$\frac{\frac{1}{n}\Sigma_1^n U_i D_i \widehat{s}_{ni}}{E s_{ni}} = O_p(1), \tag{A.48}$$

$$\frac{(\widehat{\theta}-\theta_0)'\frac{1}{\sqrt{n}}\Sigma_1^n Z_i U_i D_i \widehat{s}_{ni}}{(ED_i s_{ni}^2)^{1/2}} \xrightarrow{p} 0\ , \quad \text{and} \tag{A.49}$$

$$\frac{(\widehat{\theta}-\theta_0)'\frac{1}{\sqrt{n}}\Sigma_1^n Z_i Z_i' D_i \widehat{s}_{ni}(\widehat{\theta}-\theta_0)}{(ED_i s_{ni}^2)^{1/2}} \xrightarrow{p} 0. \tag{A.50}$$

The proof of (A.45) is the same as that of (A.13) except that the factor U_i^2 appears in various sums and expectations instead of $|U_i+\mu_0|$. By Assumption 2(b), this requires $EU_i^2<\infty$, which holds by Assumption 1.

By the proof of Theorem A–1, (A.46), (A.47), and (A.48) hold. In particular, equation (A.46) is implied by condition (i) of that proof together with Lemmas A–1 and A–2. Equation (A.47) follows from (A.14) and Assumption 7. Equation (A.48) follows from (A.13), (A.15) and Assumption 7.

The proof of (A.49) is the same as that of (A.14) except that the factor $|U_i|$ appears in various sums and expectations. In analogy with the proof of (A.13), using Assumption 2(b), this requires $E|U_i|<\infty$, which holds by Assumption 1.

We now establish (A.50). By Assumption 5 and a one-term Taylor expansion about β_0, the absolute value of the ℓhs of (A.50) is bounded by:

$$\begin{aligned}
&O_p(1)\frac{\frac{1}{n^{3/2}}\Sigma_1^n\|Z_i\|^2\widehat{s}_{ni}}{(Es_{ni}^2)^{1/2}}\\
\le\ & \frac{O_p(1)\frac{1}{n}\Sigma_1^n\|Z_i\|^2 s_{ni}}{(nEs_{ni}^2)^{1/2}}+\frac{O_p(1)\frac{1}{n}\Sigma_1^n\|Z_i\|^2\|X_i s'(X_i\beta_*-\gamma_n)\sqrt{n}(\widehat{\beta}-\beta_0)\|}{(n^2Es_{ni}^2)^{1/2}}\\
&O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|^2 1(W_i>\gamma_n)}{(nEs_{ni}^2)^{1/2}}+O_p(1)\frac{\frac{1}{n}\Sigma_1^n\|Z_i\|^2|X_i|}{(n^2Es_{ni}^2)^{1/2}}.
\end{aligned} \tag{A.51}$$

By the same proof as for (A.22), the rhs of (A.51) is $o_p(1)$ provided $E\|Z_i\|^\lambda<\infty$ for $\lambda>3$ and $E\|Z_i\|^2|X_i|<\infty$. The latter hold by Assumption 1. □

Proof of Theorem 5: By the proof of Theorem A–1, $\frac{1}{n}\Sigma_1^n D_i\widehat{s}_{ni}/\frac{1}{n}\Sigma_1^n D_i s_{ni}\xrightarrow{p}1$. (It is condition (i) of that proof.) If we show, analogously, that

$$\frac{1}{n}\Sigma_1^n D_i\widehat{s}_{ni}^2/\frac{1}{n}\Sigma_1^n D_i s_{ni}^2\xrightarrow{p}1, \tag{A.52}$$

then the ℓhs of the theorem equals

$$\begin{pmatrix}\frac{\sqrt{n}ED_i s_{ni}}{\sigma(ED_i s_{ni}^2)^{1/2}}(1+o_p(1))(\widehat{\mu}_n-\mu_0)\\ \sqrt{n}\Omega^{-1/2}(1+o_p(1))\binom{\widehat{\theta}-\theta_0}{\widehat{\beta}-\beta_0}\end{pmatrix} \tag{A.53}$$

using Lemmas A–1 and A–4, Assumption 11, and Theorem 4. Theorem 5 follows from (A.53) and Theorem 3. It remains to establish (A.52). Using Lemma A–4, we find that the following condition is sufficient:

$$\frac{\frac{1}{\sqrt{n}}\Sigma_1^n D_i(\widehat{s}_{ni}^2-s_{ni}^2)}{(Es_{ni}^4)^{1/2}}\xrightarrow{p}0. \tag{A.54}$$

The proof is analogous to that of (A.16) with s_{ni} replaced by s_{ni}^2 and s_{ni}^2 replaced by s_{ni}^4. In addition to the assumptions required for (A.16), the proof uses (A.6) of the proof of Lemma A–4. $\square$

APPENDIX OF PROOFS (B)

This Appendix contains the adjustments required when we relax the assumption of independence of (U_i, ε_i) and (X_i, Z_i).

Lemmas A–1, A–2, and A–4 used in the previous Appendix also apply in this case, except that we replace Assumptions 1 and 2(b) with Assumptions 1^* and 2^*(b) respectively. We call the revised lemmas Lemmas A–1^*, A–2^* and A–4^*. Lemma A–3 is replaced by Lemma A–3^* given below. The proofs make use of two additional lemmas, both of which are required to show consistency of the estimator $\widehat{V}_n$.

Lemma A–3^*: *Under Assumptions* 1^*, 2^*(b), $3'$, 6, *and* 12, $\frac{EU_i^2 D_i s^2(W_i - \gamma_n)}{EU_i^2 s^2(Wi - \gamma_n)} \to 1$. *If Assumption 4 also holds,* $\frac{\mathrm{Var}(U_i D_i s(W_i - \gamma_n))}{EU_i^2 D_i s^2(W_i - \gamma_n)} \to 1$.

Lemma B–1: *Under Assumptions* 1^*, 2^*(b), $3'$, 6, 10^*, *and* 12, $\frac{EU_i^4 D_i s^4(W_i - \gamma_n)}{EU_i^4 s^4(Wi - \gamma_n)} \to 1$.

Lemma B–2: *Under Assumptions* 1^*, 2^*(b), $3'$, 6, 7^*, 10^*, *and* 12, $\frac{\frac{1}{n}\sum_1^n U_i^2 D_i s^2(W_i - \gamma_n)}{EU_i^2 D_i s^2(W_i - \gamma_n)}$ $\xrightarrow{p} 1$.

The need for Assumption 2^*(b) can be seen, for instance, by considering the proof of Lemma A-2^*.

Proof of Lemma A–2^*: We change equation (A.3) to

$$\left|\frac{ED_i s_{ni}}{Es_{ni}} - 1\right| = \frac{E(1-D_i)s_{ni}}{Es_{ni}} \le \frac{E1(\gamma_n \le \varepsilon_i)s_{ni}}{Es_{ni}} \le \sup_{w \in \mathcal{W}} P(\gamma_n \le \varepsilon_i | W_i = w), \quad \text{(B.1)}$$

which converges to zero using Assumptions 2^*(b) and 6. $\square$

The need for Assumption 12 becomes clear when we consider the proof of Lemma A–3^*.

Proof of Lemma A–3^*: We have

$$\begin{aligned} \left|\frac{EU_i^2 D_i s_{ni}^2}{EU_i^2 s_{ni}^2} - 1\right| &= \frac{EU_i^2(1-D_i)s_{ni}^2}{EU_i^2 s_{ni}^2} \le \frac{EU_i^2 1(\gamma_n \le \varepsilon_i)s_{ni}^2}{EU_i^2 s_{ni}^2} \\ &\le \frac{\sup_{w \in \mathcal{W}} E(U_i^2 1(\gamma_n \le \varepsilon_i) | W_i = w)}{\inf_{w \in \mathcal{W}} E(U_i^2 | W_i = w)} \\ &\le \frac{(\sup_{w \in \mathcal{W}} E(|U_i|^\lambda | W_i = w))^{2/\lambda} (\sup_{w \in \mathcal{W}} P(\gamma_n \le \varepsilon_i | W_i = w))^{1/q}}{\inf_{w \in \mathcal{W}} E(U_i^2 | W_i = w)} \to 0, \end{aligned} \quad \text{(B.2)}$$

where λ is defined in Assumption 1^* and q satisfies $2/\lambda + 1/q = 1$. The first inequality uses Assumption $3'$ and the convergence to zero holds by Assumptions 1^*, 2^*(b), 6, and 12.

Next, we have

$$\begin{aligned}
\frac{\operatorname{Var}(U_i D_i s_{ni})}{EU_i^2 D_i s_{ni}^2} &= 1 - \frac{(EU_i D_i s_{ni})^2}{EU_i^2 D_i s_{ni}^2}, \\
\frac{(EU_i D_i s_{ni})^2}{EU_i^2 D_i s_{ni}^2} &\leq \frac{(E\,|U_i|\,s_{ni})^2}{\inf_{w\in\mathcal{W}} E(U_i^2|W_i = w) E D_i s_{ni}^2}(1 + o(1)) \\
&\leq \frac{\sup_{w\in\mathcal{W}} E\left(|U_i|\,|W_i = w\right)^2 P(W_i \geq \gamma_n)^2}{\inf_{w\in\mathcal{W}} E(U_i^2|W_i = w)\cdot P(W_i > \gamma_n + b)}(1 + o(1)) = o(1),
\end{aligned} \tag{B.3}$$

where the second equation uses the first part of this Lemma, Lemma A–2^*, and Assumptions 1^*, 4, 6, and 12. □

Proof of Lemma B–1: The proof is analogous to the first part of Lemma A–3^*. The proof additionally requires Assumption 10^*. □

Proof of Lemma B–2: The proof is analogous to that of Lemma A–4^*. A similar argument to that of (B.3) can be used to show that $EU_i^4 D_i s_{ni}^4 / n(EU_i^2 D_i s_{ni}^2)^2 = o(1)$. The proof uses Lemma B–1. □

Proof of Lemma 1^*: The proof follows Lemma 1 exactly. □

Proof of Lemma 2^*: To establish part (a), note that the first equation of (A.9) in this case is bounded by

$$4^{1+\delta} \sup_{w\in\mathcal{W}} E\left(|U_i|^{2+2\delta}|W_i = w\right) E s_{ni}^{2+2\delta} \leq 4^{1+\delta} \sup_{w\in\mathcal{W}} E\left(|U_i|^{2+2\delta}|W_i = w\right) E s_{ni}^2. \tag{B.4}$$

Using Lemma A–3^*, for $0 < \delta \leq 1$, the ℓhs of Assumption 8 is then bounded by

$$\begin{aligned}
&\lim_{n\to\infty} E\left(\frac{(B_n - EB_n)^2}{EU_i^2 s_{ni}^2(1+o(1))} 1((B_n - EB_n)^2 > nEU_i^2 s_{ni}^2 \varepsilon(1+o(1)))\right) \\
&\leq \lim_{n\to\infty} \frac{E|B_n - EB_n|^{2+2\delta}}{(EU_i^2 s_{ni}^2)^{1+\delta}\varepsilon^\delta n^\delta} \\
&\leq \lim_{n\to\infty} \frac{4^{1+\delta} \sup_{w\in\mathcal{W}} E\left(|U_i|^{2+2\delta}|W_i = w\right)}{\inf_{w\in\mathcal{W}} E\left(U_i^2|W_i = w\right)^{1+\delta} \varepsilon^\delta (nP(W_i > \gamma_n + b))^\delta (1 + o(1))} \to 0,
\end{aligned} \tag{B.5}$$

where the second inequality uses (B.4) and the convergence to zero holds by Assumptions 1^*, 7^*, and 12.

To establish part (b) we need to use Assumption 2^*(a) to establish that $EU_i s_{ni} = 0$. In (A.11) we need to replace $E|U_i|1(\varepsilon_i > \gamma_n)$ by $\sup_{w\in\mathcal{W}} E\left(|U_i|1(\varepsilon_i > \gamma_n)|W_i = w\right)$.
□

Theorems 1*, 2*, and 3* follow from Theorems A–1 and A–2, A–1 and A–3, and A–1 and A–4 respectively, given in the previous Appendix, with Assumptions 1* and 2* replacing Assumptions 1 and 2, and Theorems 2* and 3* replacing Theorems 2 and 3 respectively. We call the revised theorems Theorems A–1*, A–2*, A–3*, and A–4*.

Proof of Theorem A–1*: The proof of Theorem A–1 needs to be revised at two points. In particular, they concern the establishment of the conditions given in (A.13) and (A.15).

First we consider (A.15). The first term on the rhs of (A.17) is $O_p(1)$, because it has mean zero and variance $o(1)$:

$$\frac{\mathrm{Var}(U_i D_i s_{ni})}{n(ED_i s_{ni})^2} \leq \frac{E(U_i^2 D_i s_{ni}^2)}{n(ED_i s_{ni})^2} \leq \frac{\sup_{w\in\mathcal{W}} E(U_i^2|W_i=w)\, ED_i s_{ni}^2(1+o(1))}{n(ED_i s_{ni})^2} \to 0 \quad \text{(B.6)}$$

using Lemmas A–2* and A–3* and Assumptions 1* and 7. The second term on the rhs of (A.17) is bounded by $\sup_{w\in\mathcal{W}} E(|U_i|\,|W_i=w)$, which is $O_p(1)$ by Assumption 1*.

Next, we establish (A.13). By (A.16) and Lemma A–2* it is sufficient to show

$$\frac{1}{\sqrt{n}}\Sigma_1^n U_i D_i(\widehat{s}_{ni} - s_{ni})/(ED_i s_{ni}^2)^{1/2} \xrightarrow{p} 0. \quad \text{(B.7)}$$

The proof is analogous to that of (A.16), where $|X_i|$, X_i^2, $|X_i|^3$ are replaced by $|U_i X_i|$, $|U_i|X_i^2$, $|U_i||X_i|^3$. In consequence, the proof requires the following moment conditions to hold $E|U_i X_i|^p < \infty$, where p is as in Assumption 1*, $E\,|U_i|^{3/2}\,|X_i|^3 < \infty$, and $E\,|U_i|\,|X_i|^3 < \infty$, which hold by Assumption 1*.

Proof of Theorem A–2*: The first multiplicand on the rhs of (A.23) has mean $EU_i D_i s_{ni}/ED_i s_{ni}$, whose absolute value by Assumption 2*(a) is

$$\frac{|EU_i(1-D_i)s_{ni}|}{ED_i s_{ni}} \leq \frac{E|U_i|1(\gamma_n < \varepsilon_i)s_{ni}}{Es_{ni}(1+o(1))} \leq \sup_{w\in\mathcal{W}} E(|U_i|1(\gamma_n < \varepsilon_i)|W_i=w)(1+o(1)) = o(1) \quad \text{(B.8)}$$

using Lemma A–2* and Assumptions 1*, 2*(b), and 6. The first multiplicand on the rhs of (A.23) has variance equal to $o(1)$ by (B.6). □

Proof of Theorem A–3*: Equation (A.27) in this case reduces to

$$\frac{\sqrt{n}ED_i s_{ni}}{\mathrm{Var}^{1/2}(U_i D_i s_{ni})}\left(\widehat{\mu}_{n0} - \mu_0 - \frac{EU_i D_i s_{ni}}{\frac{1}{n}\Sigma_1^n D_i s_{ni}}\right) = \frac{ED_i s_{ni}}{\frac{1}{n}\Sigma_1^n D_i s_{ni}}\frac{1}{\sqrt{n}}\Sigma_1^n \frac{U_i D_i s_{ni} - EU_i D_i s_{ni}}{\mathrm{Var}^{1/2}(U_i D_i s_{ni})}. \quad \text{(B.9)}$$

For part (a), it remains to show

$$\frac{\sqrt{n}ED_i s_{ni}}{\mathrm{Var}^{1/2}(U_i D_i s_{ni})}\left|\frac{EU_i D_i s_{ni}}{\frac{1}{n}\Sigma_1^n D_i s_{ni}} - \frac{EU_i D_i s_{ni}}{ED_i s_{ni}}\right| = o_p(1). \quad \text{(B.10)}$$

Using Lemma A–3*, the following inequality holds:

$$\mathrm{Var}(U_i D_i s_{ni}) \geq \inf_{w\in\mathcal{W}} E(U_i^2|W_i = w) E s_{ni}^2 (1 + o(1)). \tag{B.11}$$

By analogy with (A.29), the ℓhs of (B.10) can then be shown to equal $o_p(1)$ using (B.11), Lemmas A–1* and A–2*, and Assumption 12. □

PROOF OF THEOREM A–4*: Equation (A.30) is given by

$$\frac{\sqrt{n} E D_i s_{ni}}{\mathrm{Var}^{1/2}(U_i D_i s_{ni})} c_1(\widehat{\mu}_{n0} - \mu_0) + \sqrt{n} c_2' \Omega^{-1/2} \begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{\beta} - \beta_0 \end{pmatrix} \xrightarrow{d} N(0,1). \tag{B.12}$$

The only difference from the proof of Theorem A–4 is in showing that $\left(\mathrm{Cov} \frac{c_1 U_i D_i s_{ni}}{\mathrm{Var}^{1/2}(U_i D_i s_{ni})},\ c_2' \Omega^{-1/2} Q_i\right) = o(1)$. Using (B.11), we can show that this holds if we add Assumption 12. □

Theorem 4* follows from Theorems B–1 and B–2 below. Define

$$\widehat{V}_{n0} = \Sigma_1^n U_i^2 D_i s_{ni}^2 / \Sigma_1^n D_i s_{ni}^2. \tag{B.13}$$

THEOREM B–1: *Under Assumptions* 1*, 2*(b), 3′, 4, 6, 7, 10*, *and* 12, $\widehat{V}_{n0} \mathrm{E}(D_i s_{ni}^2) / \mathrm{Var}(U_i D_i s_{ni}) \xrightarrow{p} 1$.

THEOREM B–2: *Under Assumptions* 1*, 2*(b), 3–6, 7*, 10*, *and* 12, $\widehat{V}_n - \widehat{V}_{n0} \xrightarrow{p} 0$.

PROOF OF THEOREM B–1: By Lemma A–2*, it is sufficient to show

$$\tfrac{1}{n} \sum\nolimits_1^n U_i^2 D_i s_{ni}^2 / \mathrm{Var}(U_i D_i s_{ni}) \xrightarrow{p} 1. \tag{B.14}$$

Using Lemma A–3*, the ℓhs of equation (B.14) has mean $1 + o_p(1)$, and variance

$$\frac{\mathrm{Var}(U_i^2 D_i s_{ni}^2)}{n \mathrm{Var}\left(U_i D_i s_{ni}\right)^2} \leq \frac{E(U_i^4 D_i s_{ni}^4)}{n E\left(U_i^2 D_i s_{ni}^2\right)^2 (1 + o(1))} = o(1), \tag{B.15}$$

where the inequality holds by Lemma A–3* and the equality holds by the proof of Lemma B–2. □

PROOF OF THEOREM B–2: The proof is analogous to that of Theorem A–6 with s_{ni} and $\widehat{s}_{ni}$ replaced by s_{ni}^2 and $\widehat{s}_{ni}^2$ respectively. Condition (i) of the proof holds by the proof of Theorem A–1*. By Theorem B–1, condition (iii) is satisfied. It remains to show

$$\frac{\frac{1}{n}\sum_1^n \widehat{U}_i^2 D_i \widehat{s}_{ni}^2 - \frac{1}{n}\sum_1^n U_i^2 D_i s_{ni}^2}{E D_i s_{ni}^2} = o_p(1). \tag{B.16}$$

The six conditions, sufficient for (B.16), are identical to (A.45)–(A.50) with s_{ni}, $\widehat{s}_{ni}$, and s_{ni}^2 replaced by s_{ni}^2, $\widehat{s}_{ni}^2$, and s_{ni}^4, respectively (call them (A.45*)–(A.50*)). The proofs are similar to those given in the previous appendix.

Due to replacing s_{ni} by s_{ni}^2 and s_{ni}^2 by s_{ni}^4, we need to impose Assumption 7* rather than Assumption 7, which was used in Theorem A.6. The adaptations to the proofs required follow the example given. Let L_n denote the ℓhs of (A.46*). A three-term Taylor expansion of $s^2(X_i\widehat{\beta}-\gamma_n)$ about β_0 gives:

$$\begin{aligned}
|L_n| \leq & \left|\frac{1}{n}\Sigma_1^n X_i D_i 2 s_{ni} s_{ni}' \sqrt{n}(\widehat{\beta}-\beta_0)/(Es_{ni}^4)^{1/2}\right| \\
& + \left|\frac{1}{2n}\Sigma_1^n X_i^2 D_i (2(s_{ni}')^2 + 2 s_{ni} s_{ni}'')[\sqrt{n}(\widehat{\beta}-\beta_0)]^2/(nEs_{ni}^4)^{1/2}\right| \\
& + \left|\frac{1}{6n}\Sigma_1^n X_i^3 D_i (6s'(X_i\beta_*-\gamma_n)s''(X_i\beta^*-\gamma_n) + 2s(X_i\beta_*-\gamma_n)s'''(X_i\beta^*-\gamma_n))\right. \quad \text{(B.17)} \\
& \left. \times\ [\sqrt{n}(\widehat{\beta}-\beta_0)]^3/(n^2Es_{ni}^4)^{1/2}\right| \\
\leq & \ O_p(1)\frac{1}{n}\Sigma_1^n |X_i| 1(X_i\beta_0>\gamma_n)/(Es_{ni}^4)^{1/2} \\
& + O_p(1)\frac{1}{n}\Sigma_1^n X_i^2 1(X_i\beta_0>\gamma_n)/(nEs_{ni}^4)^{1/2} + O_p(1)\frac{1}{n}\Sigma_1^n |X_i|^3/(n^2Es_{ni}^4)^{1/2},
\end{aligned}$$

where β_* is on the line segment joining $\widehat{\beta}$ and β_0, and the second inequality uses Assumptions 3 and 5 and the inequalities $|s'(x)| \leq 1(x>0)\sup_{x_*\in R}|s'(x_*)|$, $|s''(x)| \leq 1(x>0)\sup_{x_*\in R}|s''(x_*)|$, and $\sup_{x\in R}|s'''(x)| < \infty$.

The first term on the rhs of (B.17) is

$$O_p(1)\frac{E|X_i|1(W_i>\gamma_n)}{(Es_{ni}^4)^{1/2}} \leq O_p(1)\frac{(E|X_i|^p)^{1/p}P(W_i>\gamma_n)^{1/q}}{P(W_i>\gamma_n+b)^{1/2}} = o_p(1), \qquad \text{(B.18)}$$

where $1/p+1/q=1$, p is as in Assumption 1*, and the equality holds by Assumptions 4 and 6 since $2/q > 1+\xi$.

The second term on the rhs of (B.17) is

$$\begin{aligned}
O_p(1)\frac{EX_i^2 1(W_i>\gamma_n)}{(nEs_{ni}^4)^{1/2}} & \leq O_p(1)\frac{(E|X_i|^3)^{2/3}P(W_i>\gamma_n)^{1/3}Es_{ni}^2}{(n(Es_{ni}^2)^2/Es_{ni}^4)^{1/2}Es_{ni}^4} \\
\leq O_p(1)\left(\frac{n(Es_{ni}^2)^2}{Es_{ni}^4}\right)^{-1/2} & \frac{P(W_i>\gamma_n)^{1+1/3}}{P(W_i>\gamma_n+b)} = o_p(1),
\end{aligned}$$

where the equality holds by Assumptions 4, 6 and 7* (see (A.6) in proof of Lemma A–4). The third term can be shown to be $o_p(1)$ by analogy. This completes the proof of (A.46*).

Additional changes in the proofs of (A.45)–(A.50) occur in the cases where U_i or U_i^2 appear. These changes are similar to those alluded to in the proof of Theorem A-1*. $\square$ $\int$

Proof of Theorem 5*: Using Lemma A–1* and Theorem 4*, the ℓhs of the theorem equals

$$\begin{pmatrix} \dfrac{\sqrt{n} E D_i s_{ni}}{\mathrm{Var}^{1/2}(U_i D_i s_{ni})} (1 + o_p(1))(\widehat{\mu}_n - \mu_0) \\ \sqrt{n}\Omega^{-1/2}(1 + o_p(1))\binom{\widehat{\theta}-\theta_0}{\widehat{\beta}-\beta_0} \end{pmatrix}. \tag{B.19}$$

Theorem 5* follows from (B.19) and Theorem 3*. □

REFERENCES

Amemiya, T. (1984): "Tobit Models: A Survey," *Journal of Econometrics*, 24, 3–61.

Andrews, D.W.K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, 307–345.

Arabmazar, A. and P. Schmidt (1981): "Further Evidence on the Robustness of the Tobit Estimator to Heteroskedasticity," *Journal of Econometrics*, 17, 253–258.

_____ (1982): "An Investigation of the Robustness of the Tobit Estimator to Non-Normality," *Econometrica*, 50, 1055–1063.

Baker, M., D. Benjamin, A. Desoulniers, and M. Grant (1993): "The Distribution of the Male/Female Earnings Differential," Working paper No. 9307, Department of Economics and Institute for Policy Analysis, University of Toronto.

Chamberlain, G. (1986): "Asymptotic Efficiency in Semiparametric Models With Censoring," *Journal of Econometrics,* 32, 189–218.

Chow, Y.S. and H. Teicher (1978): *Probability Theory, Independence, Interchangeability, Martingales.* New York: Springer–Verlag.

Cosslett, S. (1990): "Semiparametric Estimation of a Regression Model with Sample Selectivity," in *Nonparametric and Semiparametric Estimation Methods in Econometrics and Statistics*, ed. by W.A. Barnett, J. Powell, and G.E. Tauchen. Cambridge: Cambridge University Press.

Cotton, J. (1988): "On the Decomposition of Wage Differentials," *Review of Economics and Statistics*, 70, 236–243.

Gallant, R. and D. Nychka (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

Goldberger, A.S. (1983): "Abnormal Selection Bias," in *Studies in Econometrics, Time Series and Multivariate Statistics,* ed. by S. Karlin, T. Amemiya, and L.A. Goodman. New York: Wiley.

Gronau, R. (1974): "Wage Comparisons: A Selectivity Bias," *Journal of Political Economy*, 82, 1119–1143.

Heckman, J.J. (1974): "Shadow Prices, Market Wages and Labor Supply," *Econometrica*, 42, 679–694.

_____ (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.

_____ (1990): "Varieties of Selection Bias," *American Economic Review*, 80, 2, 313–318.

Ichimura, H. and L.F. Lee (1990): "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in *Nonparametric and Semiparametric Estimation Methods in Econometrics and Statistics*, ed. by W.A. Barnett, J. Powell, and G.E. Tauchen. Cambridge: Cambridge University Press.

Ichimura, H. (1985): "Estimation of Single Index Models," unpublished discussion paper, Department of Economics, M.I.T.

Klein, R.W. and R.H. Spady (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61, 387–421.

Lee, L.F. (1986): "Unionism and Relative Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19, 415–433.

Lee, L.F. and R.P. Trost (1978): "Estimation of Some Limited Dependent Variable Models with Applications to Housing Demand," *Journal of Econometrics,* 8, 357–382.

Lewis, H.G. (1974): "Comments on Selectivity Biases in Wage Comparisons," *Journal of Political Economy*, 82, 1145–1155.

_____ (1986): *Union Relative Wage Effects: A Survey.* Chicago: University of Chicago Press.

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

Neumark, D. (1988): "Employers' Discriminatory Behavior and the Estimation of Wage Discrimination,' *Journal of Human resources*, 23, 279–295.

Newey, W. (1988): "Two Step Series Estimation of Sample Selection Models," unpublished manuscript, Department of Economics, Princeton University.

Oaxaca, R.L. (1973): "Male-Female Wage Differentials in Urban Labour Markets," *International Economic Review*, 14, 693–709.

Oaxaca, R.L. and M.R. Ransom (1994): "On Discrimination and the Decomposition of Wage Differentials," *Journal of Econometrics*, 61, 5–21.

Powell, J. (1989): "Semiparametric Estimation of Censored Selection Models," unpublished manuscript, University of Wisconsin-Madison.

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

Roy, A.D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

Schafgans, M.M.A. (1995): "Semiparametric Estimation of a Sample Selection Model: A Simulation Study," mimeo, Yale University.

Smith, J.P. and F. Welch (1986): "Closing the Gap: Forty Years of Economic Progress for Blacks," Working paper No. R-3330-DOL, The Rand Corporation, Santa Monica.

Wellington, A.J. (1993): "Changes in the Male/Female Wage Gap, 1976–85," *Journal of Human Resources*, 28, 383–411.

Willis, R.J. and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, Part 2, 507–526.