Econometric Theory, **18**, 2002, 40–50. Printed in the United States of America.

ON INTERCEPT ESTIMATION IN THE SAMPLE SELECTION MODEL

MARCIA M.A. SCHAFGANS
London School of Economics

VICTORIA ZINDE-WALSH
McGill University

We provide a proof of the consistency and asymptotic normality of the estimator suggested by Heckman (1990, *American Economic Review* 80, 313–318) for the intercept of a semiparametrically estimated sample selection model. The estimator is based on "identification at infinity," which leads to nonstandard convergence rate.

1. INTRODUCTION

Semiparametric estimation of sample selection models has attracted considerable interest in the last decade. More recently the estimation of the intercept of the semiparametric estimated sample selection model, typically absorbed in the nonparametric sample selection bias correction term, has received due attention (see Heckman, 1990; Andrews and Schafgans, 1998; Chen, 1999). Estimation of the intercept, e.g., allows one to evaluate the net benefit of a social program. It also permits a decomposition of the wage gap between socioeconomic groups (e.g., male–female) to assess the extent of "discrimination" (Schafgans, 1998a) and allows for a discussion of its evolution over time (Buchinsky, 1998).

Heckman (1990) suggested estimating the intercept μ_0 using only those observations for which the probability of selection is close to one and in the limit as $n \to \infty$ is one by introducing a bandwidth for the selection index that goes to infinity. The justification of this approach is that the conditional mean of the errors in the outcome equation for the observations having probability of selection close to one is close to zero. Andrews and Schafgans (1998) proved consistency and asymptotic normality for a smoothed estimator based on this idea.

In this note, we derive the consistency and asymptotic distribution of the Heckman estimator. This provides the empirical researcher with the advantage of not having to specify the smoothing function introduced in Andrews and

The authors thank two anonymous referees and Joel Horowitz for their comments and suggestions. This work was supported by the Social Sciences and Humanities Research Council of Canada and by the Fonds FCAR (Québec). Address correspondence to: Marcia Schafgans, Department of Economics, London School of Economics, London WC2A 2AE, UK; e-mail: m.schafgans@lse.ac.uk.

 © 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

Schafgans (1998). Schafgans (1998b) uses simulations whereas Schafgans and Zinde-Walsh (2000) use the asymptotic rate of the mean square error (MSE) to compare the two estimators and discuss the selection of the bandwidth parameter over a wide class of distributional assumptions. Although the choice of bandwidths usually has the most important consequences for the properties of the estimators, there are cases where the choice of estimator itself affects the optimal asymptotic rate.

The remainder of this note is organized as follows. Section 2 introduces the sample selection model considered and the estimators of Heckman (1990) and Andrews and Schafgans (1998). The asymptotic normality result for the Heckman estimator is given in Section 3. All of the proofs appear in the Appendix.

2. INTERCEPT ESTIMATION

The sample selection model that we consider can be written as

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

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 selection equation. For convenience, we set

$$W_i = X_i'\beta_0. \tag{2}$$

The Heckman (1990) estimator is

$$\hat{\mu}_I = \frac{\sum_{i=1}^{n} (Y_i - Z_i'\hat{\theta}) D_i 1(X_i'\hat{\beta} > \gamma_n)}{\sum_{i=1}^{n} D_i 1(X_i'\hat{\beta} > \gamma_n)}. \tag{3}$$

The Andrews and Schafgans (1998) estimator replaces the indicator function by a smooth function $s(\cdot)$:

$$\hat{\mu}_s = \frac{\sum_{i=1}^{n} (Y_i - Z_i'\hat{\theta}) D_i s(X_i'\hat{\beta} - \gamma_n)}{\sum_{i=1}^{n} D_i s(X_i'\hat{\beta} - \gamma_n)}. \tag{4}$$

Here $(\hat{\theta}, \hat{\beta})$ are root-n consistent preliminary estimators of (θ_0, β_0) (for references, see Andrews and Schafgans, 1998). The parameter γ_n is the bandwidth or smoothing parameter; $\gamma_n \to \infty$ as $n \to \infty$.

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

3. ASYMPTOTIC NORMALITY OF THE HECKMAN ESTIMATOR

Here we prove the conjecture made in Andrews and Schafgans (1998) that the Heckman estimator also is asymptotically normal. In the unrealistic case where the true θ_0 and β_0 are known, Andrews and Schafgans already showed that the Heckman estimator, $\hat{\mu}_{I,0}$, which is identical to $\hat{\mu}_I$ with $(\hat{\theta}, \hat{\beta})$ replaced by their true values (θ_0, β_0), is asymptotically normal.

For our purposes, all we need to show now is that (denoting $\mathrm{Var}(U_i)$ by σ^2)

$$\frac{\sqrt{nED_i 1(W_i > \gamma_n)}}{\sigma} (\hat{\mu}_I - \hat{\mu}_{I,0}) \xrightarrow{p} 0. \tag{5}$$

This requires us to deal directly with the nondifferentiability of the indicator function. Because the expectation of the generalized derivative of the indicator function equals the value of the probability density function (p.d.f.) at zero, our Assumption A involves the p.d.f.

Denote in the partition $y = [y_{(1)} \quad y'_{(-1)}]'$ of an arbitrary $k \times 1$ vector y the first element by $y_{(1)}$, the rest by $y_{(-1)}$. Consider the linear transformation $LX_i = [W_i: \quad \Gamma_i']'$ with $L = \begin{pmatrix} \beta_{0,(1)} & \beta'_{0,(-1)} \\ 0 & I_{k-1} \end{pmatrix}$, where I_{k-1} is an identity matrix of dimension $k-1$; then $\Gamma_i = X_{i,(-1)}$. For β_0 we assume that $\beta_{0,(1)} \neq 0$.

We add the following assumption to the list of assumptions in Andrews and Schafgans (1998).[1]

Assumption A.

(a) For some $A > 0$ the conditional p.d.f. $\mathrm{pdf}_{W_i, \Gamma_i | \Gamma_i}(W|\cdot)$ *exists for all $W > A$ and declines monotonically. The marginal p.d.f.* $\mathrm{pdf}_{W_i}(W)$ *is such that for some $d \to 0$, $dn^{1/6} \to \infty$*

$$\frac{\mathrm{pdf}_{W_i}(\gamma_n - d)}{\Pr(W_i > \gamma_n)^{3/4}} = O(1).$$

(b) For any $W > A$ the conditional moment $E_{|W_i = W}(\|X_{i,(-1)}\|^3)$ exists; $\beta_{0,(1)} \neq 0$.

Similar to Assumptions 4 and 7 of Andrews and Schafgans, Assumption A(a) relates to the upper tail behavior of the selection index $W_i = X_i' \beta_0$. It is satisfied if W_i has a Weibull, Pareto, or "combined" upper tail. If Assumption 4 of Andrews and Schafgans is satisfied with $\xi = 0$, then this condition can be replaced by

$$\frac{\mathrm{pdf}_{W_i}(\gamma_n)}{\Pr(W_i > \gamma_n)^{3/4}} = O(1).$$

In both cases, the condition is less strong than requiring a bounded hazard function on W_i, because $\mathrm{pdf}_{W_i}(\gamma_n - d)^{1/4} = o(1)$.

The second part of Assumption A requires the existence of the conditional moment but does not place any restrictions on its behavior as a function of W_i.

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

This assumption is satisfied, e.g., if the unbounded components of X_i have a joint normal or spherical distribution with W_i.

The following theorem summarizes our result for the Heckman estimator.

THEOREM 1. *Under Assumptions A and 1, 2, and 4–7 of Andrews and Schafgans (with $b = 0$ in Andrews and Schafgans)*

(a) $$\frac{\sqrt{nED_i 1(W_i > \gamma_n)}}{\sigma}\left(\hat{\mu}_I - \mu_0 - \frac{EU_i D_i 1(W_i > \gamma_n)}{ED_i 1(W_i > \gamma_n)}\right) \xrightarrow{d} N(0,1)$$

(b) $$\frac{\sqrt{nED_i 1(W_i > \gamma_n)}}{\sigma}(\hat{\mu}_I - \mu_0) \xrightarrow{d} N(0,1)$$ *iff Assumption 8 of Andrews and Schafgans holds.*

Proof. See the Appendix.

To test hypotheses and construct confidence intervals for functions of $(\mu_0, \theta_0, \beta_0)$, we obtain a joint asymptotic normality result for $(\hat{\mu}_I, \hat{\theta}, \hat{\beta})$ (similarly to Andrews and Schafgans).[2]

THEOREM 2. *Under Assumptions A and 1, 2, and 4–8 of Andrews and Schafgans (with $b = 0$ in Andrews and Schafgans)*

$$\begin{pmatrix} \frac{\sqrt{nED_i 1(W_i > \gamma_n)}}{\sigma}(\hat{\mu}_I - \mu_0) \\ \sqrt{n}\Omega^{-1/2}\begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \end{pmatrix} \xrightarrow{d} N(0, I).$$

Proof. See the Appendix.

NOTES

1. Essentially, Assumptions 1 and 2 of Andrews and Schafgans require existence of moments and independence between (U_i, ε_i) and (Z_i, X_i); Assumptions 3 and 3′ deal with the shape of $s(\cdot)$; Assumption 4 characterizes the upper tail of W_i in terms of a parameter $0 \leq \xi < \frac{1}{3}$ with "fatter" tails if $\xi = 0$; Assumption 5 is root-n consistency and asymptotic normality of $(\hat{\theta}, \hat{\beta})$; Assumption 6 is $\gamma_n \to \infty$, where its speed is restricted by Assumption 7 in terms of the tail of $W_i > \gamma_n$. Assumption 8 ensures that the bias of $\hat{\mu}_n$ goes to zero asymptotically.

2. As in Andrews and Schafgans, a joint asymptotic normality result of $(\hat{\mu}_n, \hat{\theta}, \hat{\beta})$ with estimated covariance matrix can be obtained also.

REFERENCES

Andrews, D.W.K. & M.M.A. Schafgans (1998) Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–518.

Buchinsky, M. (1998) The dynamics of changes in the female wage distribution in the USA: A quantile regression approach. *Journal of Applied Econometrics* 13, 1–30.

Chen, S. (1999) Distribution-free estimation of the random coefficient dummy endogenous variable model. *Journal of Econometrics* 91, 171–199.

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

Heckman, J.J. (1990) Varieties of selection bias. *American Economic Review* 80, 313–318.
Schafgans, M.M.A. (1998a) Ethnic wage differences in Malaysia: Parametric and semiparametric estimation of the Chinese–Malay wage-gap. *Journal of Applied Econometrics* 13, 481–504.
Schafgans, M.M.A. (1998b) A Monte Carlo study of the semiparametric estimation of the intercept of a sample selection model. Manuscript, London School of Economics.
Schafgans, M.M.A. & V. Zinde-Walsh (2000) On intercept estimation in the sample selection model. Sticerd Discussion paper EM/00/380, London School of Economics and Political Science.

APPENDIX: ASYMPTOTIC NORMALITY RESULT FOR THE HECKMAN ESTIMATOR

To prove asymptotic normality, all we need to show is that (5) holds. As in the proof of Theorem A-1 of Andrews and Schafgans the left-hand side of (5) can be written as $C\big((\hat{A}/\hat{B}) - (A/B)\big) = C\big((\hat{A} - A)/B\big)(B/\hat{B}) - C\big((\hat{B} - B)/B\big)(A/B)(B/\hat{B})$, where $C = \sqrt{nED_i 1(W_i > \gamma_n)}/\sigma$, $A = \sum_{i=1}^n (Y_i - Z_i'\theta_0)D_i 1(W_i > \gamma_n)$, $\hat{A} = \sum_{i=1}^n (Y_i - Z_i'\hat{\theta})D_i 1(X_i'\hat{\beta} > \gamma_n)$, $B = \sum_{i=1}^n D_i 1(W_i > \gamma_n)$, and $\hat{B} = \sum_{i=1}^n D_i 1(X_i'\hat{\beta} > \gamma_n)$. To show (5), therefore, it suffices to show that

$$\begin{aligned}
&\text{(i)}\ \ \hat{B}/B \xrightarrow{p} 1\\
&\text{(ii)}\ \ C(\hat{A} - A)/B \xrightarrow{p} 0\\
&\text{(iii)}\ \ A/B = O_p(1)\\
&\text{(iv)}\ \ C(\hat{B} - B)/B \xrightarrow{p} 0.
\end{aligned} \tag{A.1}$$

From Assumption 7 and Lemma A-2 of Andrews and Schafgans (note that Assumption 3′ in Andrews and Schafgans is trivially satisfied here) it follows that $C \to \infty$, which means that (iv) implies (i) in (A.1). From Lemmas A-1 and A-2 of Andrews and Schafgans one gets that $\sigma C/B$ equals $n^{-1/2}\Pr(W_i > \gamma_n)^{-1/2}$ in probability and thus (ii) is implied by the following sufficient conditions:

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n (U_i + \mu_0)D_i\alpha_{in}}{\Pr(W_i > \gamma_n)^{1/2}} \xrightarrow{p} 0, \tag{A.2}$$

$$\frac{-(\hat{\theta} - \theta_0)'\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i D_i \alpha_{in}}{\Pr(W_i > \gamma_n)^{1/2}} \xrightarrow{p} 0, \quad \text{and} \tag{A.3}$$

$$\frac{-(\hat{\theta} - \theta_0)'\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i D_i 1(W_i > \gamma_n)}{\Pr(W_i > \gamma_n)^{1/2}} \xrightarrow{p} 0, \tag{A.4}$$

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

where the nondifferentiable function α_{in} is given by

$$\alpha_n(\hat{\beta}, W_i, X_i) \quad \text{with } \alpha_n(\beta, W, X) = 1(X'\beta > \gamma_n) - 1(W > \gamma_n). \tag{A.5}$$

Condition (iii) holds following the proof of Theorem A-1 in Andrews and Schafgans. It provides

$$\frac{\frac{1}{n}\sum_{i=1}^{n} U_i D_i 1(W_i > \gamma_n)}{\Pr(W_i > \gamma_n)} = O_p(1). \tag{A.6}$$

Finally, condition (iv) would follow if we show that

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_i \alpha_{in}}{\Pr(W_i > \gamma_n)^{1/2}} \xrightarrow{p} 0. \tag{A.7}$$

Note that under Assumptions 1, 5, and 6 of Andrews and Schafgans and using Hölder's inequality the expression in (A.4) is bounded by

$$\frac{O_p(1)E\|Z_i\|1(W_i > \gamma_n)}{\Pr(W_i > \gamma_n)^{1/2}} \leq O_p(1)(E\|Z_i\|^3)^{1/3}(\Pr(W_i > \gamma_n))^{2/3-1/2} \to 0. \tag{A.8}$$

The proof of Theorem 1, therefore, requires us to show that (A.2), (A.3), and (A.7), which involve the discontinuous function α, hold.

We do this in three steps. First, a technical lemma is given (Lemma 1), then a lemma is given that examines terms in the expressions of interest (Lemma 2), and finally we give the proofs of the theorems, which combine the intermediate results.

LEMMA 1. *Under Assumptions 1, 2, 6, and 7 of Andrews and Schafgans and Assumption A, there exists a sufficiently slowly increasing* $\{M_n\}$ *such that* $M_n = O(n^{1/6})$ *and*

$$\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})} \frac{E(\sqrt{n}\|\Omega_i\|\alpha_n(\beta, W_i, X_i))}{\Pr(W_i > \gamma_n)^{1/2}} \to 0 \quad \text{as } n \to \infty \quad \text{for } \Omega_i = Z_i/\sqrt{n},$$

$$\Omega_i = U_i, \qquad \text{or } \Omega_i = 1, \quad \forall i.$$

Proof of Lemma 1. It will be sufficient to show that for the sequence $\{n^{1/3}\}$

$$\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})} \frac{\sqrt{n}E(\|\Omega_i\|\cdot|\alpha_n(\beta, W_i, X_i)|\cdot 1(\|X_i\| > n^{1/3}))}{\Pr(W_i > \gamma_n)^{1/2}} \to 0 \tag{A.9}$$

and

$$\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})} \frac{\sqrt{n}E(\|\Omega_i\|\cdot|\alpha_n(\beta, W_i, X_i)|\cdot 1(\|X_i\| \leq n^{1/3}))}{\Pr(W_i > \gamma_n)^{1/2}} \to 0. \tag{A.10}$$

Noting that $|\alpha_n(\beta, W_i, X_i)| \leq 1$, we get for (A.9) in the case $\Omega_i = U_i$ (using the independence condition of Assumption 2(b) of Andrews and Schafgans) and similarly for $\Omega_i = 1$, under Assumptions 1 and 7 of Andrews and Schafgans,

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

$$\frac{\sqrt{n}E(\|\Omega_i\|\cdot|\alpha_n(\beta,W_i,X_i)|\cdot 1(\|X_i\|>n^{1/3}))}{\Pr(W_i>\gamma_n)^{1/2}} \le \frac{\sqrt{n}E(\|\Omega_i\|\cdot 1(\|X_i\|>n^{1/3}))}{\Pr(W_i>\gamma_n)^{1/2}}$$

$$\le \frac{\sqrt{n}E\|\Omega_i\|\cdot\Pr(\|X_i\|>n^{1/3})}{\Pr(W_i>\gamma_n)^{1/2}} \le \frac{\sqrt{n}E\|\Omega_i\|\cdot E\|X_i\|^3/n}{\Pr(W_i>\gamma_n)^{1/2}}$$

$$= \frac{E\|\Omega_i\|\cdot E\|X_i\|^3}{(n\Pr(W_i>\gamma_n))^{1/2}} \to 0 \qquad \textbf{(A.11)}$$

using Jensen's and Markov's inequalities. In the case $\Omega_i = Z_i/\sqrt{n}$, (A.9) converges to zero even faster, because similarly

$$\frac{E(\|Z_i\|\cdot|\alpha_n(\beta,W_i,X_i)|\cdot 1(\|X_i\|>n^{1/3}))}{\Pr(W_i>\gamma_n)^{1/2}} \le \frac{(E\|Z_i\|^3)^{1/3}(E\|X_i\|^3)^{2/3}}{n^{1/6}\cdot(n\Pr(W_i>\gamma_n))^{1/2}} \to 0. \qquad \textbf{(A.12)}$$

The left-hand side of (A.10) for $\Omega_i = Z_i/\sqrt{n}$ is bounded by

$$\frac{(E\|Z_i\|^2)^{1/2}}{(n\Pr(W_i>\gamma_n))^{1/4}}\left(\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})}\frac{\sqrt{n}E|\alpha_n(\beta,W_i,X_i)|1(\|X_i\|<n^{1/3})}{\Pr(W_i>\gamma_n)^{1/2}}\right)^{1/2} \qquad \textbf{(A.13)}$$

using Hölder's inequality. Given Assumptions 1 and 7 of Andrews and Schafgans, therefore, it remains only to show (A.10) for $\Omega_i = 1$ and U_i.

Using the independence of Ω_i and X_i for $\Omega_i = 1$ and U_i (by Assumption 2(b) of Andrews and Schafgans) we can rewrite the numerator of (A.10) as

$$E(\|\Omega_i\|)\cdot\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})}\sqrt{n}E|\alpha_n(\beta,W_i,X_i)|1(\|X_i\|<n^{1/3}). \qquad \textbf{(A.14)}$$

By examining the function $\alpha_n(\beta,W_i,X_i)$ from equation (A.5), we can see that $|\alpha_n(\beta,W_i,X_i)|$ equals one if either $\gamma_n < W_i < \gamma_n + |X_i'(\beta-\beta_0)|$ for negative $X_i'(\beta-\beta_0)$ or if $\gamma_n - |X_i'(\beta-\beta_0)| < W_i < \gamma_n$ for positive $X_i'(\beta-\beta_0)$ and zero otherwise. In view of the restrictions on $\|\beta-\beta_0\|$ and $\|X_i\|$ we have $|X_i'(\beta-\beta_0)| < (M_n/n^{1/6})$, which implies $W_i > \gamma_n - (M_n/n^{1/6})$. Hence,

$$\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})}\sqrt{n}E|\alpha_n(\beta,W_i,X_i)|1(\|X_i\|<n^{1/3})$$

$$= \sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})}\sqrt{n}E|\alpha_n(\beta,W_i,X_i)|1\left(W_i>\gamma_n-\frac{M_n}{n^{1/6}}\right)1(\|X_i\|<n^{1/3}). \qquad \textbf{(A.15)}$$

Recall the linear transformation $LX = (W,\Gamma')'$, L nonsingular. Using $X = L^{-1}(W,\Gamma')'$ we write $X'(\beta-\beta_0) = W(L'^{-1}(\beta-\beta_0))_{(1)} + \Gamma'(L'^{-1}(\beta-\beta_0))_{(-1)}$. We denote $(L'^{-1}(\beta-\beta_0))_{(1)} = h$ and $(L'^{-1}(\beta-\beta_0))_{(-1)} = H$, where $|h| \le \|L^{-1}\|\cdot\|\beta-\beta_0\| \le \|L^{-1}\|(M_n/\sqrt{n})$ and similarly $\|H\| \le \|L^{-1}\|(M_n/\sqrt{n})$. Assume that M_n is such that $\|L^{-1}\|(M_n/\sqrt{n}) < 1$. Reexamining, for negative $X_i'(\beta-\beta_0)$, the condition $\gamma_n < W_i < \gamma_n + |X_i'(\beta-\beta_0)| = \gamma_n - W_i(L'^{-1}(\beta-\beta_0))_{(1)} - \Gamma_i'(L'^{-1}(\beta-\beta_0))_{(-1)}$ we realize that

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

for negative $X_i'(\beta-\beta_0)\,|\alpha_n(\beta,W_i,X_i)|$ equals one for $\gamma_n < W_i < (\gamma_n - \Gamma_i' H)/(1+h)$. Similarly for positive $X_i'(\beta-\beta_0)\,|\alpha_n(\beta,W_i,X_i)|$ equals one for $(\gamma_n - \Gamma_i' H)/(1+h) < W_i < \gamma_n$. Thus (A.15) equals

$$\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})} \sqrt{n} E\Bigg[\int 1(X_i'(\beta-\beta_0)<0)1\left(\gamma_n < W < \frac{\gamma_n - \Gamma_i' H}{1+h}\right) \times 1(\|X_i\| < n^{1/3})\,\mathrm{pdf}_{W_i,\Gamma_i|\Gamma_i}(W,\Gamma_i)\,dW + \int 1(X_i'(\beta-\beta_0)>0)1\left(W > \gamma_n - \frac{M_n}{n^{1/6}}\right) \times 1\left(\frac{\gamma_n - \Gamma_i' H}{1+h} < W < \gamma_n\right) \times 1(\|X_i\| < n^{1/3})\,\mathrm{pdf}_{W_i,\Gamma_i|\Gamma_i}(W,\Gamma_i)\,dW\Bigg]. \tag{A.16}$$

The $\mathrm{pdf}_{W_i,\Gamma_i|\Gamma_i}$ exists for large enough n, because the expression under the integrals is nonzero only if $W > \gamma_n - (M_n/n^{1/6})$ and thus as $\gamma_n \to \infty$ for $M_n = O(n^{1/6})$ becomes greater than A; the conditional $\mathrm{pdf}_{W_i,\Gamma_i|\Gamma_i}(W|\cdot)$ declines monotonically in W. Denote $\gamma_n - (M_n/n^{1/6})$ by $\tilde{\gamma}_n$. Each of the integrals can be bounded by $\mathrm{pdf}_{W_i,\Gamma_i|\Gamma_i}(\tilde{\gamma}_n,\Gamma_i)\cdot\left[\left(|\Gamma_i' H|/(1+h)\right) + |h/(1+h)|\gamma_n\right]$. Thus the sum is bounded by

$$\sup_{\|\beta-\beta_0\|<(M_n/\sqrt{n})} 2\sqrt{n}\int\left(\mathrm{pdf}_{W_i,\Gamma_i|\Gamma_i}(\tilde{\gamma}_n,\Gamma)\left[\frac{|\Gamma' H|}{1+h} + \left|\frac{h}{1+h}\right|\gamma_n\right]\right) d\Pr_\Gamma$$

$$\leq \frac{2M_n\|L^{-1}\|}{1+h}\,\mathrm{pdf}_{W_i}(\tilde{\gamma}_n)\int[\|\Gamma\| + \gamma_n]\,d\Pr_{\Gamma|W_i=\tilde{\gamma}_n}$$

$$= \frac{2M_n\|L^{-1}\|}{1+h}\,\mathrm{pdf}_{W_i}(\tilde{\gamma}_n)\left[E_{|W_i=\tilde{\gamma}_n}\|\Gamma_i\| + \tilde{\gamma}_n + \frac{M_n}{n^{1/6}}\right], \tag{A.17}$$

where $\Gamma_i = X_{i,(-1)}$. The notation $d\Pr_\Gamma$ indicates that a Stieltjes integral with respect to the cumulative probability function Γ is taken; if a marginal density exists $d\Pr_\Gamma = \mathrm{pdf}_\Gamma(\Gamma)d\Gamma$. By Jensen's inequality, and in view of Assumption A(b), this is bounded almost surely by

$$\frac{2M_n\|L^{-1}\|}{1+h}\cdot\mathrm{pdf}_{W_i}(\tilde{\gamma}_n)\cdot\left[E_{|W_i=\tilde{\gamma}_n}(\|X_{i,(-1)}\|^3)^{1/3} + \tilde{\gamma}_n + \frac{M_n}{n^{1/6}}\right]. \tag{A.18}$$

First we consider $E_{|W_i=\tilde{\gamma}_n}(\|X_{i,(-1)}\|^3)$. By Assumption 1 of Andrews and Schafgans, we know that the unconditional expectation $E(\|X_{i,(-1)}\|^3)$ is bounded; this implies that for the fixed A of Assumption A $E(\|X_{i,(-1)}\|^3 1(W_i > A))$ is bounded also.

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

$$E(\|X_{i,(-1)}\|^3 1(W_i > A)) = \iint \|X_{(-1)}\|^3 1(W > A)\,\mathrm{pdf}_{W_i}(W)\,d\,\mathrm{Pr}_{\Gamma|W_i=W}\,dW$$

$$= \int J(W)\,\mathrm{pdf}_{W_i}(W)1(W > A)dW, \tag{A.19}$$

where $J(W) = \int \|X_{(-1)}\|^3 d\,\mathrm{Pr}_{\Gamma|W_i=W}$ is a function of W only. Because the integral $\int J(W)\mathrm{pdf}_{W_i}(W)dW \equiv E(\|X_{i,(-1)}\|^3)$ exists, it implies that $\mathrm{pdf}_{W_i}(W)J(W) = o(W^{-1})$.

Because EW_i^3 exists from Assumption 1 in Andrews and Schafgans we also have $\mathrm{pdf}_{W_i}(\gamma)\gamma^3 = o(\gamma^{-1})$ as $\gamma \to \infty$. Rewriting (A.18), we get

$$\frac{2M_n\|L^{-1}\|}{1+h}\cdot \mathrm{pdf}_{W_i}(\tilde{\gamma}_n)^{2/3}\cdot[(\mathrm{pdf}_{W_i}(\tilde{\gamma}_n)J(\tilde{\gamma}_n))^{1/3} + (\mathrm{pdf}_{W_i}(\tilde{\gamma}_n)\tilde{\gamma}_n^3)^{1/3}]$$

$$+ \frac{2M_n^2\|L^{-1}\|}{n^{1/6}(1+h)}\cdot \mathrm{pdf}_{W_i}(\tilde{\gamma}_n)$$

$$= \frac{2M_n\|L^{-1}\|}{1+h}\cdot \mathrm{pdf}_{W_i}\left(\gamma_n - \frac{M_n}{n^{1/6}}\right)^{2/3} o(\gamma_n^{-1/3}), \tag{A.20}$$

where the equality uses in addition the assumption that $M_n n^{-1/6} = O(1)$.

This implies that for $\Omega_i = 1$ and U_i (A.10) can be bounded as

$$\frac{E(\|\Omega_i\|)\,\dfrac{2M_n\|L^{-1}\|}{1+h}\cdot \mathrm{pdf}_{W_i}\left(\gamma_n - \dfrac{M_n}{n^{1/6}}\right)^{2/3} o(\gamma_n^{-1/3})}{\mathrm{Pr}(W_i > \gamma_n)^{1/2}}$$

$$\leq E(\|\Omega_i\|)\,\frac{2M_n\|L^{-1}\|}{1+h}\left(\frac{\mathrm{pdf}_{W_i}\left(\gamma_n - \dfrac{M_n}{n^{1/6}}\right)}{\mathrm{Pr}(W_i > \gamma_n)^{3/4}}\right)^{2/3} o(\gamma_n^{-1/3}). \tag{A.21}$$

Set $M_n = \min\{dn^{1/6}, \gamma_n^{1/3}\}$ with $d > 0$ as in Assumption A. Using Assumptions 1 and 6 of Andrews and Schafgans and Assumption A(a) the right-hand side of (A.21) converges to zero. ■

LEMMA 2. *Under Assumptions 1, 2, and 5–7 of Andrews and Schafgans and under Assumption A*

$$\frac{\dfrac{1}{\sqrt{n}}\sum \|\Omega_i\||\alpha_{in}|}{\mathrm{Pr}(W_i > \gamma_n)^{1/2}} \xrightarrow{p} 0 \quad \text{as } n \to \infty \quad \text{for } \Omega_i = Z_i/\sqrt{n}, \qquad \Omega_i = U_i, \quad \text{or } \Omega_i = 1, \quad \forall i.$$

Proof of Lemma 2. We show that for any $\zeta > 0$

$$\mathrm{Pr}\left(\frac{\dfrac{1}{\sqrt{n}}\sum \|\Omega_i\||\alpha_{in}|}{\mathrm{Pr}(W_i > \gamma_n)^{1/2}} > \zeta\right) \to 0 \quad \text{as } n \to \infty. \tag{A.22}$$

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

From Lemma 1, let M_n satisfy $M_n o(\gamma_n^{-1/3}) \to 0$ and $(M_n/n^{1/6}) \leq d$. The left-hand side of (A.22) is equal to

$$\Pr\left(\frac{\frac{1}{\sqrt{n}}\sum \|\Omega_i\||\alpha_{in}|}{\Pr(W_i > \gamma_n)^{1/2}} > \zeta, \|\hat{\beta} - \beta_0\| \leq \frac{M_n}{\sqrt{n}}\right)$$

$$+ \Pr\left(\frac{\frac{1}{\sqrt{n}}\sum \|\Omega_i\||\alpha_{in}|}{\Pr(W_i > \gamma_n)^{1/2}} > \zeta, \|\hat{\beta} - \beta_0\| > \frac{M_n}{\sqrt{n}}\right)$$

$$\leq \Pr\left(\frac{\frac{1}{\sqrt{n}}\sum \|\Omega_i\||\alpha_{in}|}{\Pr(W_i > \gamma_n)^{1/2}} > \zeta, \|\hat{\beta} - \beta_0\| \leq \frac{M_n}{\sqrt{n}}\right) + \Pr\left(\|\hat{\beta} - \beta_0\| > \frac{M_n}{\sqrt{n}}\right). \quad \textbf{(A.23)}$$

The second expression on the right-hand side of (A.23) converges to zero by Assumption 5 of Andrews and Schafgans, because $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$ implies $\sqrt{n}(\hat{\beta} - \beta_0)/M_n = o_p(1)$. The first expression on the right-hand side of (A.23) is bounded by

$$\sup_{\|\beta - \beta_0\| \leq M_n/\sqrt{n}} \Pr\left(\frac{\frac{1}{\sqrt{n}}\sum \|\Omega_i\||\alpha_n(\beta, W_i, X_i)|}{\Pr(W_i > \gamma_n)^{1/2}} > \zeta\right)$$

$$\leq \sup_{\|\beta - \beta_0\| \leq M_n/\sqrt{n}} \frac{E\left(\frac{\frac{1}{\sqrt{n}}\sum \|\Omega_i\||\alpha_n(\beta, W_i, X_i)|}{\Pr(W_i > \gamma_n)^{1/2}}\right)}{\zeta}$$

$$= \sup_{\|\beta - \beta_0\| \leq M_n/\sqrt{n}} \frac{\sqrt{n}E\|\Omega_i\||\alpha_n(\beta, W_i, X_i)|}{\zeta \Pr(W_i > \gamma_n)^{1/2}}, \quad \textbf{(A.24)}$$

where the inequality is based on Markov's inequality. This term converges to zero for all $\zeta > 0$ by Lemma 1. ■

Proof of Theorem 1. For our proof it is sufficient to show (A.2), (A.3), and (A.7). By Assumptions 1 and 5 for $\hat{\theta} - \theta_0$ of Andrews and Schafgans, the left-hand sides of (A.2) and (A.3) can be bounded by

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} |U_i| \cdot |\alpha_{in}|}{\Pr(W_i > \gamma_n)^{1/2}} + |\mu_0| \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} |\alpha_{in}|}{\Pr(W_i > \gamma_n)^{1/2}}, \quad \textbf{(A.25)}$$

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press

$$\frac{O_p(1)\,\frac{1}{n}\sum_{i=1}^{n}\|Z_i\|\cdot|\alpha_{in}|}{\Pr(W_i>\gamma_n)^{1/2}}, \tag{A.26}$$

respectively. From Lemma 2, (A.25) and (A.26) have zero probability limits.

Finally, to prove (A.7), we note that its left-hand side can be bounded by

$$\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|\alpha_{in}|}{\Pr(W_i>\gamma_n)^{1/2}}, \tag{A.27}$$

which by Lemma 2 again converges in probability to zero. This completes the proof of our theorem. ■

Proof of Theorem 2. By the Cramér–Wold device, this result follows directly from (5) and Theorem A-4 in Andrews and Schafgans. In the latter, the result in Theorem 2 is shown for the case where $\hat{\mu}_I$ is replaced by $\hat{\mu}_{I,0}$. ■

https://doi.org/10.1017/S0266466602181035 Published online by Cambridge University Press