No. 1927R
Rev. May 1989

DECONVOLUTION BASED SCORE TESTS IN MEASUREMENT ERROR MODELS

Leonard A. Stefanski
Department of Statistics
North Carolina State University
Raleigh, NC 27695

Raymond J. Carroll
Department of Statistics
Texas A&M University
College Station, TX 77843

ABSTRACT

Consider a generalized linear model with response Y and scalar predictor X. Instead of observing X, a surrogate $W = X + Z$ is observed where Z represents measurement error and is independent of X and Y. The efficient score test for the absence of association depends on $m(w) = E(X|W = w)$ which is generally unknown (Tosteson and Tsiatis, 1988). Assuming that the distribution of Z is known, asymptotically efficient tests are constructed using nonparametric estimators of $m(w)$. Rates of convergence for the estimator of $m(w)$ are established in the course of proving efficiency of the proposed test.

Key Words and Phrases: Deconvolution, Density estimation, Errors-in-variables, Generalized linear models, Maximum likelihood, Measurement Error Models, Score tests.

AMS 1980 *subject classifications.* Primary 62J05; secondary 62H25, 62G05.

Abbreviated title: Deconvolution Based Score Tests.

1. INTRODUCTION

Let X be a random variable with unknown density f_X and characteristic function ϕ_X. Given $X = x$, the response Y follows a generalized linear model with likelihood

$$\exp\left[\left\{y\zeta - b(\zeta)\right\}\Big/\gamma + c(y,\gamma)\right], \tag{1.1}$$

where $\zeta = g(\alpha + \beta x)$ and α, β and γ are unknown parameters. We study testing $H_0 : \beta = 0$ when a surrogate variable W is observed in place of X. This is a generalized linear measurement error model. Applications in epidemiology motivating our work are discussed by Carroll (1989).

Frequently H_0 is tested using the usual score test statistic

$$T_U = n^{-1/2}\sum_{i=1}^{n} W_i(Y_i - \bar{Y})/(S_W S_Y), \tag{1.2}$$

where S_W^2 and S_Y^2 are the sample variances of $\{W_i\}$ and $\{Y_i\}$ respectively. Although this test has the correct level asymptotically, it may be inefficient. For example, when Y and W are conditionally independent given X, the efficient score test statistic is

$$T_E = n^{-1/2}\sum_{i=1}^{n} m(W_i)(Y_i - \bar{Y})/(S_{m(W)} S_Y) \tag{1.3}$$

where $m(w) = E(X|W = w)$ and $S_{m(W)}^2$ is the sample variance of $\{m(W_i)\}$. Comparing (1.2) and (1.3) shows that the usual score test is inefficient when $m(w)$ is nonlinear in w. See Tosteson and Tsiatis (1988) for further details.

Since $m(w)$ is usually unknown, it must be estimated in order to construct an asymptotically efficient score test. In this paper we present a method of estimating $m(w)$ based on $\{W_i\}$ only, and then use this estimator to construct an asymptotically efficient test. We consider the additive measurement error model,

$$W = X + Z, \tag{1.4}$$

where Z is independent of (Y, X).

We assume that the error density f_Z is known, symmetric, and has finite second moment, and that its characteristic function $\phi_Z(t)$ is nonzero for all real t. The deconvolution kernel density estimator of Stefanski and Carroll (1988) is used to estimate f_X, which for known f_Z yields an estimator of $m(w)$. From this, we construct a fully efficient score test.

In the course of proving the efficiency of our test we investigate the performance of the estimator of $m(w)$. Although estimation of f_X is difficult when f_Z is smooth, estimation of $m(w)$ is feasible more generally. For example, Carroll and Hall (1988) have shown that unless it is assumed that f_X has more than two bounded derivatives, the best achievable mean squared error rate of convergence of *any* estimator of f_X is of order $\{\log(n)\}^{-2}$ when f_Z is normal and of order $n^{-4/9}$ when f_Z is Laplacian. The estimator proposed by Stefanski and Carroll (1988) achieves these rates. In contrast, we show that the pointwise expected mean squared error of our estimator of $m(w)$ decreases at the rates of $n^{-4/7}$ and $n^{-4/5}$ for normal and Laplacian errors respectively. We suspect that these rates are optimal although we have not pursued this problem. In general, the rate of convergence depends in a simple way on $h_Z'(t) = \phi_Z'(t)/\phi_Z(t)$.

2. CONDITIONAL EXPECTATIONS

Writing $f_X(x) = (2\pi)^{-1} \int e^{-itx}\phi_X(t)dt$ and $f_W(w) = (2\pi)^{-1} \int e^{-itw}\phi_X(t)\phi_Z(t)dt$ we have

$$m(w) = E(X|W = w) = \int x f_Z(w-x)f_X(x)dx \Big/ \int f_Z(w-x)f_X(x)dx, \tag{2.1}$$

$$= w + (i/2\pi)\int \phi_Z'(t)\phi_X(t)e^{-itw}dt \Big/ f_W(w),$$

$$= w + (i/2\pi)\int h_Z'(t)\phi_W(t)e^{-itw}dt \Big/ f_W(w). \tag{2.2}$$

We propose an estimator based on (2.1), later giving an interpretation with respect to (2.2).

We assume that f_X has two bounded continuous derivatives. Let $G(t)$ be a four-times continuously differentiable characteristic function with support $[-1,1]$. Define the real functions

$$K(t) = (2\pi)^{-1}\int_{-1}^{1} e^{itx}G(x)dx; \qquad K_*(t,\lambda) = (2\pi)^{-1}\int_{-1}^{1} e^{itx}G(x)\Big/\phi_Z(x/\lambda)dx.$$

By Fourier inversion, K is an even bounded density function, while $\lambda^{-1}K_*(t/\lambda)$ is an even bounded function that integrates to one although it is not nonnegative. The smoothness conditions on G insure that $\int v^2K(v)dv < \infty$. An ordinary kernel estimate of f_X based on the unobserved $\{X_i\}$ is $\hat{f}_{X,K}(x) = (n\lambda)^{-1}\sum_1^n K\{(X_j - x)/\lambda\}$. By standard calculations, $E\hat{f}_{X,K}(x) = f_X(x) + (1/2)\lambda^2 f_X''(x) + o(\lambda^2)$. The deconvolution estimator is $\hat{f}_X(x) = (n\lambda)^{-1}\sum_1^n K_*\{(W_j - x)/\lambda, \lambda\}$. See Stefanski and Carroll (1988), Carroll and Hall (1988), Stefanski (1989), Liu and Taylor (1988a,b) and Fan (1988) for motivation of this estimator and more specialized properties. Since $E\left[K_*\{(W - x)/\lambda, \lambda\} \mid X\right] = K\{(X-x)/\lambda\}$, it follows that $\hat{f}_X$ and $\hat{f}_{X,K}$ have the same expectation. However, the variance of $\hat{f}_X$ can be much larger than that of $\hat{f}_{X,K}$.

Equation (2.1) can be written $m(w) = E(X \mid W = w) = M_1(w, f_X)/M_0(w, f_X)$ where

$$M_p(w, f_X) = \int x^p f_Z(w-x)f_X(x)dx, \tag{2.3}$$

suggesting the estimator

$$\hat{m}(w) = M_1(w, \hat{f}_X)/M_0(w, \hat{f}_X). \tag{2.4}$$

Note that $M_0(w, \hat{f}_X) = \hat{f}_W(w)$ is a kernel density estimator of $f_W = f_Z * f_X$. Hence by standard results the denominator of (2.4) estimates f_W at the pointwise expected squared error rate $n^{-4/5}$. The pointwise convergence of $M_1(w, \hat{f}_X)$ to $M_1(w, f_X)$ is generally slower and thus determines the convergence rate of $\hat{m}(w)$ to $m(w)$. Squared bias in $\hat{m}(w)$ is of order λ^4.

The estimator (2.4) is based on (2.1). Alternatives might be based on (2.2), since f_W can be estimated directly by kernel techniques and ϕ_W can be estimated by the empirical characteristic function $\hat{\phi}_W$. This approach fails whenever $\int h_Z'(t)\hat{\phi}_W(t)e^{-itw}dt$ fails to exist, as in the case of normal measurement error. However, the lack of integrability can be circumvented by truncating the range of integration. Our estimator (2.4) has the representation

$$\hat{m}(w) = w + (i/2\pi)\int_{-1/\lambda}^{1/\lambda} h_Z'(t)\hat{\phi}_W(t)e^{-itw}G(\lambda t)dt \Big/ \hat{f}_W(w),$$

corresponding to (2.2), and thus performs the necessary truncation automatically.

If Z is normally distributed with mean zero and variance σ_Z^2, it follows from (2.2) that

$$m(w) = w + \sigma_Z^2 f_W'(w)/f_W(w). \tag{2.5}$$

In this case, $\hat{m}(w)$ has the form (2.5) with f_W and f_W' replaced by $\hat{f}_W$ and $\hat{f}_W'$ respectively.

Theorem 1 is the main result on rates of convergence and is proved in the appendix. The probability measure governing (Y, X) under $\theta = (\alpha, \beta)^T$ is denoted P_θ, and E_θ and Var_θ denote expectation and variance under P_θ.

THEOREM 1. Assume: i) f_X,f_X' and f_X'' are continuous and bounded; ii) $\int \phi_Z(t)dt < \infty$; iii) as $|t| \to \infty$, $|h_Z'(t)| = o(|t|^\gamma)$ for some $\gamma \geq 0$; and iv) $n \to \infty$ and $\lambda \to 0$. Then $\{\hat{m}(w) - m(w)\}^2 = O_{P_\theta}\left(\lambda^4 + \left(n\lambda^{1+2\gamma}\right)^{-1}\right)$.

For normally distributed errors, $\gamma = 1$ and the pointwise squared error rate of convergence is of order $n^{-4/7}$. For Laplacian errors, $\gamma = 0$ and the rate is $n^{-4/5}$.

Theorem 1 enables us to construct an asymptotically efficient test of $H_0 : \beta = 0$. Let η_n be a sequence of positive constants converging to zero, and let $\hat{f}_{X,i}$ be the deconvolution density estimator constructed without using W_i. Define

$$\hat{m}_{(i)}(W_i) = M_1(W_i, \hat{f}_{X,i}) \Big/ \{M_0(W_i, \hat{f}_{X,i}) + \eta_n\}. \tag{2.6}$$

The constants η_n are a technical convenience bounding the denominator of (2.6) away from zero for each n. If S_Y^2 is the sample variance of $\{Y_i\}$ and $\hat{S}^2_{m(W)}$ is the sample variance of the $\{\hat{m}_{(i)}(W_i)\}$, then the test statistic we propose is

$$T = \hat{C}_2/(\hat{S}_{m(W)}S_Y), \qquad \text{where} \qquad \hat{C}_2 = n^{-1/2}\sum_{i=1}^{n}\hat{m}_{(i)}(W_i)(Y_i - \bar{Y}).$$

Write C for the numerator of (1.3). Let P_n be the probability measure governing (Y, W) under $\theta_n = (\alpha, n^{-1/2}\beta)^T$. If $\hat{C}_2 - C$ and $\hat{S}_{m(W)} - S_{m(W)}$ are asymptotically negligible under P_n, then the score test based on T is asymptotically efficient. Theorem 2, proved in the appendix, gives sufficient conditions for this to occur.

THEOREM 2. Assume the conditions of Theorem 1 and also: i) $E\left[\{1 + m^2(W_1) + W_1^2\}\{1 + Var_\theta(Y_1|W_1)\}\right]$ is finite for all θ; ii) $E_\theta(Y|X)$ is mean square differentiable with respect to θ at $\theta = (\alpha, 0)^T$, i.e.,

$$\lim_{\epsilon\to 0}\epsilon^{-2}\int\left\{b'(g(\alpha+\epsilon x)) - b'(g(\alpha))\right\}^2 f_X(x)dx = \left\{b''(g(\alpha))g'(\alpha)\right\}^2 E(X^2) < \infty;$$

and iii) $\eta_n^{-2}\{\lambda^4 + (n\lambda^{1+2\gamma})^{-1}\} \to 0$. Then T is asymptotically efficient.

3. APPLICATIONS

Practical issues related to the use of T are discussed in detail in another paper, (Stefanski and Carroll, 1989), which also presents Monte Carlo evidence of the greater efficiency of T relative to T_U. We discuss these issues briefly here and give an example application.

The use of T requires specification of the error density, a kernel and $\{\eta_n\}$, as well as estimation/specification of λ. Theorem 2 indicates that asymptotically T is invariant to the kernel, $\{\eta_n\}$ and λ over a wide range of choices. Furthermore, it can be shown that misspecification of the error density does not affect the asymptotic vaidity of T, although it does affect efficiency. Our experience suggests that T is reasonably insensitive to these auxillary parameters in finite samples, although this is not guaranteed. Thus we employ T primarily as a means of examining the impact of measurement error on the usual test statistic.

As an illustration we consider logistic regression of breast cancer incidence on long-term log daily saturated fat intake in a cohort of 2888 women under the age of 50 at time examination. The data are a subset of those analyzed by Jones, *et al.* (1987). We calculated T using the kernel $K(t) = 3\{\sin(t)/t\}^4/(2\pi)$ assuming both normal and double-exponential errors. In both cases we took $\sigma_Z = .55$, see Stefanski and Carroll (1989) for details. The test statistic was calculated for a range of bandwidths with η_n fixed at 0. Although the latter assignment violates the assumptions of Theorem 2, we found good small-sample properties of a similar estimator in the Monte Carlo study cited above, suggesting that the third assumption of Theorem 2 might be weakened. For bandwidths $\lambda = 1.2, 1.1, ..., 0.7$, $-T = 1.73, 1.71, 1.68, 1.63, 1.55, 1.43$ under normality, and $-T = 1.79, 1.79, 1.79, 1.78, 1.77, 1.74$ for double-exponential errors respectively. The need to estimate a derivative explains the greater instability of the test statistics under normality. For these data $T_U = -1.76$.

4. CONCLUSION

Deconvolution to estimate a density function can be very difficult, with slow rates of convergence. For estimating $m(w) = E(X|W = w)$, faster rates are obtainable. This is noteworthy in the case of normal measurement error, where the squared error rate of convergence for estimating a density is of order $\{\log(n)\}^{-2}$, while that for estimating $m(w)$ is of order $n^{-4/7}$. Sufficiently good estimates of the regression function have been obtained to construct a fully efficient score test for the effect of a predictor measured with error.

ACKNOWLEDGEMENT

Our research was supported by the National Science Foundation, the Air Force Office of Scientific Research and the National Institutes of Health.

REFERENCES

Carroll, R. J. (1989). Covariance analysis in generalized linear measurement error models. *Statistics in Medicine*, to appear.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, **83** 1184-1186.

Fan, J. (1988). On the optimal rates of convergence for nonparametric deconvolution problem. Technical Report No. 157, Department of Statistics, University if California, Berkeley.

Jones, D. Y., Schatzkin, A., Green, S. B., Block, G., Brinton, L. A., Ziegler, R. G., Hoover, R. and Taylor, P. R. (1987). Dietary fat and breast cancer in the National Health and Nutrition

Survey I: epidemiologic follow-up study. J. N. C. I., **79** 465-471.

Liu, M. C. and Taylor, R. L. (1989a). A consistent nonparametric density estimator for the deconvolution problem. Technical Report STA 73, University of Georgia.

Liu, M. C. and Taylor, R. L. (1989b). Simulations and computations of nonparametric density estimates for the deconvolution problem. Technical Report STA 74, University of Georgia.

Stefanski, L. A. and Carroll, R. J. (1988). Deconvoluting kernel density estimators. Preprint.

Stefanski, L. A. and Carroll, R. J. (1989). Score tests in generalized linear measurement error models. Preprint.

Stefanski, L. A. (1989). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics & Probability Letters*, to appear.

Tosteson, T. and Tsiatis, A. (1988). The asymptotic relative efficiency of score tests in a generalized linear model with surrogate covariates. *Biometrika*, **75** 507-514.

APPENDIX

For $p = 0$ or 1 make the following definitions:

$$D_p(x,w) = x^p f_Z(w-x); \qquad L_p(t,w) = \int D_p(x,w)e^{-itx}dx$$

$$B(u,v,\lambda) = EK_*\big((W_1-u)/\lambda,\lambda\big)K_*\big((W_1-v)/\lambda,\lambda\big);$$

$$A_p(w,\lambda) = \lambda^{-2}\int\int D_p(u,w)D_p(v,w)B(u,v,\lambda)dudv.$$

Three lemmas are employed in the proofs of Theorems 1 and 2. In the following $c_1 < c_2 < c_3 < c_*$ are positive numbers used to bound certain constants encountered in the proofs.

Lemma A.1. Assume the conditions of Theorem 1. Then, for all w,

$$\Big\{EM_p(w,\hat{f}_X) - M_p(w,f_X)\Big\}^2 \le c_*\lambda^4(1+pw^2); \tag{A.2}$$

$$Var\Big\{M_p(w,\hat{f}_X)\Big\} \le c_*(\lambda n)^{-1}(1-p+p\lambda^{-2\gamma}+pw^2). \tag{A.3}$$

Proof of Lemma A.1. A direct calculation yields that for some $0 \le a(v,x) \le 1$,

$$EM_p(w,\hat{f}_X) = EM_p(w,\hat{f}_{X,K}) = \int\int x^p f_Z(w-x)f_X(x+\lambda v)K(v)dvdx$$

$$= M_p(w, f_X) + (1/2)\lambda^2 \int\int x^p f_Z(w-x) f_X''(x + a(v,x)\lambda v) v^2 K(v) dv dx.$$

Since f_X'' is bounded, $\{EM_p(w, \hat{f}_X) - M_p(w, f_X)\}^2 \le c_1 \lambda^4 \{\int |x^p| f_Z(w-x)dx\}^2$. However, since $|x^p| \le 1 - p + p(|w| + |w-x|)$ and $E(Z^2) < \infty$, $\int |x^p| f_Z(w-x)dx \le c_2(1 + p|w|)$, from which (A.2) is immediate. To prove (A.3) we first show that

$$Var\{M(w, \hat{f}_X)\} \le n^{-1} A_p(w, \lambda). \tag{A.4}$$

Note that

$$M_p(w, \hat{f}_X) = (n\lambda)^{-1} \sum_{j=1}^{n} \int x^p f_Z(w-x) K_*((W_j - x)/\lambda; \lambda) dx, \tag{A.5}$$

so that $Var\{M_p(w, \hat{f}_X)\} \le (n\lambda^2)^{-1} E\{\int x^p f_Z(w-x) K_*((W_1 - x)/\lambda; \lambda)dx\}^2 = n^{-1} A_p(w, \lambda)$, thus proving (A.4). By definition of K_*, it follows that

$$A_p(w, \lambda) = \int_{-1/\lambda}^{1/\lambda} \int_{-1/\lambda}^{1/\lambda} \frac{L_p(r,w) L_p(s,w) G(r\lambda) G(s\lambda) \phi_X(r+s) \phi_Z(r+s)}{4\pi^2 \phi_Z(r) \phi_Z(s)} dr ds. \tag{A.6}$$

Note that $L_0(t, w) = e^{-itw} \phi_Z(t)$. We now show that

$$L_1(t, w) = e^{-itw} \phi_Z(t) \big(w + i h_z'(t)\big). \tag{A.7}$$

Employing a change of variable, $L_1(t, w) = \int x f_Z(w-x) e^{-itx} dx = \int (w-u) f_Z(u) e^{it(u-w)} du$ $= we^{-itw}\phi_Z(t) - e^{-itw} \int u f_Z(u) e^{itu} du$, from which (A.7) follows. We now complete the proof of (A.3). Note that by (A.6),

$$Var\Big\{M_0(w, \hat{f}_X)\Big\} \le n^{-1} \int_{-1/\lambda}^{1/\lambda} \int_{-1/\lambda}^{1/\lambda} \phi_Z(r+s) G(r\lambda) G(s\lambda) dr ds, \tag{A.8}$$

while by (A.6) and (A.7),

$$Var\Big\{M_1(w, \hat{f}_X)\Big\} \le n^{-1} \int_{-1/\lambda}^{1/\lambda} \int_{-1/\lambda}^{1/\lambda} (|w| + |r|^{\gamma})(|w| + |s|^{\gamma}) \phi_Z(r+s) G(r\lambda) G(s\lambda) dr ds. \tag{A.9}$$

Furthermore, for any (a, b), since G is bounded and ϕ_Z is integrable,

$$\int_{-1/\lambda}^{1/\lambda} \int_{-1/\lambda}^{1/\lambda} |r|^a |s|^b \phi_Z(r+s) G(r\lambda) G(s\lambda) dr ds$$

$$\le c_3 \lambda^{-a-b} \int_{-1/\lambda}^{1/\lambda} \int_{-\infty}^{\infty} \phi_Z(r+s) dr ds \le c_* \lambda^{-1-a-b}. \tag{A.10}$$

Using (A.10) in (A.8) and (A.9) completes the proof. ••

Proof of Theorem 1. Immediate from Lemma A.1. ••

Let E_n and Var_n denote expectation and variance under P_n. Define $J_n(w) = E_n(Y|W = w)$ and $\mu(t) = b'(g(t))$.

Lemma A.2. Under the assumptions of Theorem 2, $nE_n\{J_n(W) - \mu(\alpha)\}^2 = \mathcal{O}(1)$.

Proof of Lemma A.2. By definition of $J_n(w)$,

$$nE_n\{J_n(W) - \mu(\alpha)\}^2 = n\int\left[\int\{\mu(\alpha + n^{-1/2}\beta x) - \mu(\alpha)\}f_Z(w-x)f_X(x)dx\right]^2\Big/ f_W(w)dw$$

$$\leq n\int\{\mu(\alpha + n^{-1/2}\beta x) - \mu(\alpha)\}^2 f_X(x)dx \to \{\mu'(\alpha)\}^2$$

by assumption. ••

Lemma A.3. Under the assumptions of Theorem 2,

$$nE_n\left[\{\hat{m}_{(1)}(W_1) - m(W_1)\}^2\{1 + Var_n(Y_1|W_1)\}\right] \to 0.$$

Proof of Lemma A.3. Let $d_j(W_1) = E\{M_j(W_1, \hat{f}_{X,1})|W_1\}$. Some tedious algebra shows that $(1/20)\{\hat{m}_{(1)}(W_1) - m(w_1)\}^2 \leq R_1 + R_2 + R_3 + R_4 + R_5$ where

$$R_1 = \eta_n^{-2}\Big\{M_1(W_1, \hat{f}_{X,1}) - d_1(W_1)\Big\}^2; \quad R_2 = \eta_n^{-2}\Big\{d_1(W_1) - M_1(W_1, f_X)\Big\}^2;$$

$$R_3 = m^2(W_1)\eta_n^{-2}\Big\{M_0(W_1, \hat{f}_{X,1}) - d_0(W_1)\Big\}^2; \quad R_4 = m^2(W_1)\eta_n^{-2}\Big\{d_0(W_1) - M_0(W_1, f_X)\Big\}^2;$$

$$R_5 = m^2(W_1)\eta_n^2\Big\{\eta_n + M_0(W_1, f_X)\Big\}^{-2}.$$

Assumption (i) of Theorem 2 and dominated convergence imply that

$$E_n\left[R_5\{1 + Var_n(Y_1|W_1)\}\right] = o(1).$$

It follows from $(A.2)$ and $(A.3)$ that

$$E_n(R_1 + R_2|W_1) \leq c_*\eta_n^{-2}\{(\lambda n)^{-1}(\lambda^{-2\gamma} + W_1^2) + \lambda^4(1 + W_1^2)\};$$

$$E_n(R_3 + R_4|W_1) \leq c_*\eta_n^{-2}m^2(W_1)\{(\lambda n)^{-1} + \lambda^4\};$$

and thus Assumptions (i) and (ii) of Theorem 2 and dominated convergence imply that

$$E_n\left[(R_1 + R_2 + R_3 + R_4 + R_5)\{1 + Var_n(Y_1|W_1)\}\right] = o(1),$$

completing the proof. ••

Proof of Theorem 2. Define $A_1 = n^{-1/2}\sum\{\hat{m}_{(i)}(W_i)-m(W_i)\}\{Y_i-J_n(W_i)\}$ and write $\hat{C}-C = A_1 - A_2$. Under P_n the summands in A_1 are identically distributed, uncorrelated and have mean zero. Thus Lemma A.3 implies that $Var_n(A_1) = E_n[\{\hat{m}_{(1)}(W_1) - m(W_1)\}^2 Var_n(Y_1|W_1)] = o(1)$, which in turn shows that $A_1 = o_{P_n}(1)$.

Write $A_2 = A_{2,1}A_{2,2} - A_{2,3}$ where $A_{2,3} = n^{-1/2}\sum\{\hat{m}_{(i)}(W_i) - m(W_i)\}\{J_n(W_i) - \mu(\alpha)\}$, $A_{2,2} = n^{-1}\sum\{\hat{m}_{(i)}(W_i) - m(W_i)\}$ and $A_{2,1} = n^{1/2}\{\bar{Y} - \mu(\alpha)\}$. Assumption (ii) of Theorem 2 implies that $Var_n(A_{2,1}) = \mathcal{O}(1)$. Using Lemma A.2, $\{E_n(A_{2,1})\}^2 \le nE_n\{J_n(W_1)-\mu(\alpha)\}^2 = \mathcal{O}(1)$. Since $E_n(A_{2,2})^2 \le E_n\{\hat{m}_{(1)}(W_1) - m(W_1)\}^2$, Lemma A.3 implies that $A_{2,2} = o_{P_n}(1)$. It follows that $A_{2,1}A_{2,2} = o_{P_n}(1)$.

By the Cauchy-Schwarz inequality

$$A_{2,3}^2 \le \left[n^{-1}\sum_{i=1}^{n}\{\hat{m}_{(i)}(W_i) - m(W_i)\}^2\right]\left[\sum_{i=1}^{n}\{J_n(W_i) - \mu(\alpha)\}^2\right]. \qquad (A.11)$$

The first bracketed term in $(A.11)$ is $o_{P_n}(1)$ since its expectation is $E_n\{\hat{m}_{(1)}(W_1)-m(W_1)\}^2 = o(1)$ by Lemma A.3. The second bracketed term has expectation $nE_n\{J_n(W_1) - \mu(\alpha)\}^2 = \mathcal{O}(1)$ by Lemma A.2 and thus is $\mathcal{O}_{P_n}(1)$. It follows that $A_{2,3} = o_{P_n}(1)$ thus showing that $\hat{C} - C = o_{P_n}(1)$. Finally, Lemma A.3 implies that $\hat{S}_{m(W)} - S_{m(W)} = o_{P_n}(1)$ completing the proof. ••