The Annals of Statistics
1991, Vol. 19, No. 1, 249–259

DECONVOLUTION-BASED SCORE TESTS IN MEASUREMENT ERROR MODELS[1]

BY LEONARD A. STEFANSKI AND RAYMOND J. CARROLL

North Carolina State University and Texas A & M University

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. 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.

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

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

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 $\{m(W_i)\}$.

Received June 1988; revised January 1990.

[1]Research supported by NSF, the National Institutes of Health and a cooperative agreement among Harvard School of Public Health, Societal Institute of Mathematical Sciences and the Environmental Protection Agency.

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

Key words and phrases. Deconvolution, density estimation, empirical Bayes, errors-in-variables, generalized linear models, maximum likelihood, measurement error models, score tests.

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 additional 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 (1987, 1990a) 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 (1987, 1990a) 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. In general, the rate of convergence depends in a simple way on $h'_Z(t) = \phi'_Z(t)/\phi_Z(t)$. We suspect that these rates are optimal although we have not pursued this problem. The rate of convergence of higher-order moments of X given W can also be investigated using our techniques, although we do not do so in this paper.

Estimation of the posterior mean $m(w)$ is a problem central to empirical Bayes inference. Strong results are known for the case that an *exponential family* density is mixed with an unknown distribution; see Singh (1976, 1979, 1985). Viewed from the empirical Bayes perspective we are working with a *location family* density mixed with an unknown distribution. Thus our work overlaps with Singh's only when f_Z is normal or when Z, suitably scaled, has the density of $\log(Y)$, where Y has a particular gamma distribution; see Ferguson (1962). Even at the normal model, the estimators used are considerably different, as are the regularity conditions employed; we make assumptions about smoothness of f_X, whereas Singh makes assumptions about higher-order moments of X. The application of our results to the location-family empirical Bayes problem and a comparison of our rates with Singh's under similar regularity conditions would be interesting, but is beyond the scope of the present paper.

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). Throughout the paper we assume that f_X has two bounded continuous derivatives.

Let $G(t)$ be a four-times continuously differentiable, real characteristic function with bounded support, which we take to be $[-1, 1]$ without loss of generality. Define the real functions

$$K(t) = (2\pi)^{-1}\int_{-\infty}^{\infty} e^{itx}\, G(x)\,dx,$$

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

Note that since G has support $[-1, 1]$, both of these integrals exist for fixed t, under the assumption on ϕ_Z. By Fourier inversion, K is an even bounded density function, while $\lambda^{-1}K_*(t/\lambda)$ is an even bounded function that integrates to 1 although it is not nonnegative. The smoothness conditions on G ensure that $\int v^2K(v)\,dv < \infty$. An ordinary kernel estimate of f_X based on the unobserved $\{X_i\}$ is $\hat{f}_{X,K}(x) = (n\lambda)^{-1}\Sigma_1^n K\{(X_j - x)/\lambda\}$. 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 $\hat{f}_X(x) = (n\lambda)^{-1}\Sigma_1^n K_*\{(W_j - x)/\lambda, \lambda\}$. See Stefanski and Carroll (1987, 1990a), Carroll and Hall (1988), Stefanski (1988, 1990), Liu and Taylor (1988a, b) and Fan (1988) for motivation of this estimator and more specialized properties. Since $E[K_*\{(W - x)/\lambda, \lambda\}|X] = K\{(X - x)/\lambda\}$, 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 as

$$m(w) = E(X|W = w) = M_1(w, f_X)/M_0(w, f_X),$$

where

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

suggesting the estimator

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

Note that $M_0(w, \hat{f}_X) = \hat{f}_W(w)$ is a kernel density estimator of $f_W = f_Z * f_X$ based on the kernel K. 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

$$\hat{m}(w) = w + (i/2\pi)\int_{-\infty}^{\infty} h'_Z(t)\hat{\phi}_W(t)e^{-itw}G(\lambda t)\,dt\big/\hat{f}_W(w), \tag{2.5}$$

corresponding to (2.2). Since G vanishes outside $[-1, 1]$, the effective range of integration in (2.5) is $[-1/\lambda, 1/\lambda]$ and the integral exists for any h'_Z and all λ.

While the restriction on G ensures integrability in (2.5), it is apparent from (2.5) that for the purpose of estimating $m(w)$, it is sufficient to require only integrability of $h'_Z(t)G(\lambda t)$ and positivity of $\hat{f}_W$. However, we work with the representation of $\hat{m}(w)$ as a functional of $\hat{f}_X$, see (2.4), and thus we require the existence of the function $K_*(t, \lambda)$ defined previously. This imposes a much greater restriction on the tail behavior of G; see Stefanski and Carroll (1987, 1990a) for further details.

Kernels satisfying all of the required conditions can be derived from the family of densities, $K_m(x) = c_m\{\sin(x)/x\}^{2m}$, $m = 1, 2, \ldots$, where c_m is a normalizing constant, $m = 1, 2, \ldots$. The characteristic function, G_m corresponding to K_m is proportional to $U^{(2m)}$, the $2m$-fold convolution of the uniform density on $[-1, 1]$, with itself, and thus has bounded support. It is easy to find characteristic functions, G, that have bounded support although in most cases the corresponding density, K, does not have a closed form. The pairs (K_m, G_m), $m = 1, 2, \ldots$, are convenient in this respect. It follows from its relationship to $U^{(2m)}$ that K_m is approximately normal for large m; more precisely, with $a_m = \sqrt{3/m}$, $a_m K_m(a_m x)$ converges to the standard normal density. With the obvious exception of tail behavior, the approximation is good for m as small as 2.

If Z is normally distributed with mean 0 and variance σ_Z^2, it follows from (2.2) that

$$m(w) = w + \sigma_Z^2 f'_W(w)\big/f_W(w). \tag{2.6}$$

In this case, $\hat{m}(w)$ is derived from (2.6) by replacing f_W and f'_W with $\hat{f}_W$ and $\hat{f}'_W$, respectively, where $\hat{f}_W$ is an ordinary kernel density estimator of f_W based on the kernel K, and by adding a constant to the denominator of (2.6) to guard against small values of $\hat{f}_W$; see (2.7) below. Note that the restrictions on G ensure that K is analytic and thus so too is $\hat{f}_W$. The latter is a natural property in that for normal measurement error, f_W is analytic. Thus our choice of kernel imparts on $\hat{f}_W$, the same analyticity properties possessed by f_W.

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* $\gamma \geq 0$;
(iv) $n \to \infty$ *and* $\lambda \to 0$.

Then $\{\hat{m}(w) - m(w)\}^2 = O_{P_\theta}(\lambda^4 + (n\lambda^{1+2\gamma})^{-1})$.

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 0, and let $\hat{f}_{X,i}$ be the deconvolution density estimator constructed without using W_i. Define

$$\hat{m}_{(i)}(W_i) = M_1(W_i, \hat{f}_{X,i}) \big/ \{M_0(W_i, \hat{f}_{X,i}) + \eta_n\}. \tag{2.7}$$

The constants η_n are a technical convenience bounding the denominator of (2.7) away from 0 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

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

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) $E[\{1 + m^2(W_1) + W_1^2\}\{1 + \mathrm{Var}_\theta(Y_1|W_1)\}]$ *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.*,

$$\lim_{\varepsilon \to 0} \varepsilon^{-2} \int \{b'(g(\alpha + \varepsilon x)) - b'(g(\alpha))\}^2 f_X(x)\,dx$$

$$= \{b''(g(\alpha))g'(\alpha)\}^2 E(X^2) < \infty;$$

(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 (1990b)], 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 or 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 validity of T, although it does affect efficiency. Our experience suggests that T is reasonably insensitive to these auxiliary parameters in finite samples, although this is not guaranteed.

In our work we employ T primarily as a means of examining the impact of measurement error on the usual test statistic, calculating it under different assumed error distributions. Choice of $\{\eta_n\}$ and λ is more difficult. In the absence of a well-developed theory for estimating bandwidth, such as exists for usual kernel density estimation, we suggest calculating T for a range of bandwidths, a strategy consistent with the exploratory role suggested for T. Although taking $\eta_n \equiv 0$ 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. In the example below we also take $\eta_n \equiv 0$. Clearly, further theoretical support for this assignment would be desirable.

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 of examination. The data are a subset of those analyzed by Jones, Schatzkin, Green, Block, Brinton, Ziegler, Hoover and Taylor (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 $\sigma_Z = 0.55$; see Stefanski and Carroll (1990b) for details. The test statistic was calculated for a range of bandwidths with η_n fixed at 0. The range of bandwidths was chosen so that at the minimum and maximum, the corresponding estimators of $m(w)$

appeared to be under-smoothed and over-smoothed, respectively. For $\lambda = 1.2, 1.1, \ldots, 0.7$, $-T = 1.73, 1.71, 1.68, 1.63, 1.55, 1.43$ 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.

APPENDIX

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

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

$$B(u, v, \lambda) = EK_*((W_1 - u)/\lambda, \lambda) K_*((W_1 - v)/\lambda, \lambda), \tag{A.1}$$

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

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,

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

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

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

$$\begin{aligned} EM_p(w, \hat{f}_X) = EM_p(w, \hat{f}_{X,K}) &= \int\int x^p f_Z(w - x) f_X(x + \lambda v) K(v)\, dv\, dx \\ &= M_p(w, f_X) \\ &\quad + (1/2)\lambda^2 \int\int x^p f_Z(w - x) f_X''(x + a(v, x)\lambda v) v^2 K(v)\, dv\, dx. \end{aligned}$$

Since f_X'' is bounded,

$$\{EM_p(w, \hat{f}_X) - M_p(w, f_X)\}^2 \le c_1\lambda^4\left\{\int |x^p| f_Z(w - x)\, dx\right\}^2.$$

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

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

Note that

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

so that

$$\operatorname{Var}\{M_p(w, \hat{f}_X)\} \le (n\lambda^2)^{-1} E\left\{\int x^p f_Z(w - x) K_*((W_1 - x)/\lambda; \lambda)\, dx\right\}^2$$
$$= n^{-1}A_p(w, \lambda),$$

thus proving (A.4). By definition of K_*, it follows that

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

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

$$\text{(A.7)} \qquad L_1(t, w) = e^{-itw}\, \phi_Z(t)\big(w + ih_Z'(t)\big).$$

Employing a change of variable,

$$L_1(t, w) = \int x f_Z(w - x) e^{-itx}\, dx$$
$$= \int (w - u) f_Z(u) e^{it(u - w)}\, du$$
$$= w e^{-itw}\phi_Z(t) - e^{-itw}\int u f_Z(u) e^{itu}\, du,$$

from which (A.7) follows. We now complete the proof of (A.3). Note that by (A.6),

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

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

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

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

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

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,* $nE_n\{J_n(W) - \mu(\alpha)\}^2 = O(1)$.

PROOF. By definition of $J_n(w)$,

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

by assumption. □

LEMMA A.3. *Under the assumptions of Theorem 2,*

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

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

where

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

$$R_2 = \eta_n^{-2}\{d_1(W_1) - M_1(W_1, f_X)\}^2,$$

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

$$R_4 = m^2(W_1)\eta_n^{-2}\{d_0(W_1) - M_0(W_1, f_X)\}^2,$$

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

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

$$E_n[R_5\{1 + \mathrm{Var}_n(Y_1|W_1)\}] = o(1).$$

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

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

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

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

$$E_n[(R_1 + R_2 + R_3 + R_4 + R_5)\{1 + \mathrm{Var}_n(Y_1|W_1)\}] = o(1),$$

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 0. Thus Lemma A.3 implies that

$$\mathrm{Var}_n(A_1) = E_n\left[\{\hat{m}_{(1)}(W_1) - m(W_1)\}^2\mathrm{Var}_n(Y_1|W_1)\right] = o(1),$$

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 $\mathrm{Var}_n(A_{2,1}) = O(1)$. Using Lemma A.2,

$$\{E_n(A_{2,1})\}^2 \le nE_n\{J_n(W_1) - \mu(\alpha)\}^2 = O(1).$$

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

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

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 $nE_n\{J_n(W_1) - \mu(\alpha)\}^2 = O(1)$ by Lemma A.2 and thus is $O_{P_n}(1)$. 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. □

REFERENCES

CARROLL, R. J. (1989). Covariance analysis in generalized linear measurement error models. *Statist. in Medicine* **8** 1075–1095.

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

FAN, J. (1988). On the optimal rates of convergence for nonparametric deconvolution problem. Technical Report 157, Dept. Statist., Univ. California, Berkeley.

FERGUSON, T. S. (1962). Location and scale parameters in exponential families of distributions. *Ann. Math. Statist.* **33** 986–1001. [Correction note (1963) **34** 1603.]

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. Nat. Cancer Inst.* **79** 465–471.

LIU, M. C. and TAYLOR, R. L. (1988a). A consistent nonparametric density estimator for the deconvolution problem. Technical Report STA 73, Dept. Statist., Univ. Georgia.

LIU, M. C. and TAYLOR, R. L. (1988b). Simulations and computations of nonparametric density estimates for the deconvolution problem. Technical Report STA 74, Dept. Statist., Univ. Georgia.

SINGH, R. S. (1976). Empirical Bayes estimation with convergence rates in noncontinuous Lebesgue exponential families. *Ann. Statist.* **4** 431–439.

SINGH, R. S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate. *Ann. Statist.* **7** 890–902.

SINGH, R. S. (1985). Empirical Bayes estimation in a multiple linear regression model. *Ann. Inst. Statist. Math.* **37** 71–86.

STEFANSKI, L. A. (1988). Rates of convergence of some estimators in a class of deconvolution problems. Inst. Statist. Mimeograph Series 1935, North Carolina State Univ.

STEFANSKI, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229–235.

STEFANSKI, L. A. and CARROLL, R. J. (1987). Deconvoluting kernel density estimators. Inst. Statist. Mimeograph Series 1909, North Carolina State Univ.

STEFANSKI, L. A. and CARROLL, R. J. (1990a). Deconvoluting kernel density estimators. *Statistics* **21** 169–184.

STEFANSKI, L. A. and CARROLL, R. J. (1990b). Score tests in generalized linear measurement error models. *J. Roy. Statist. Soc. Ser. B* **52** 345–359.

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.

DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695-8203

DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843