1928R

SCORE TESTS IN GENERALIZED LINEAR 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

No. 1928R

SCORE TESTS IN GENERALIZED LINEAR 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

SUMMARY

Hypothesis tests in generalized linear models are studied under the condition that a surrogate w is observed in place of the true predictor x. The efficient score test for the hypothesis of no association depends on the conditional expectation $E(x \mid w)$ which is generally unknown. The ususal test substitutes w for $E(x \mid w)$ and is asymptotically valid but not efficient. We investigate two new test statistics appropriate when $w = x + z$ where z is an independent measurement error. The first is a Wald test based on estimators corrected for measurement error. Despite the correction for attenuation in the estimator, this test has the same local power as the usual test. The second test employs an estimator of $E(x \mid w)$ and is asymptotically efficient for normal errors and approximately efficient when the measurement error variance is small.

Key Words:

DECONVOLUTION; DENSITY ESTIMATE; ERRORS-IN-VARIABLES; GENERALIZED LINEAR MODELS; LOGISTIC REGRESSION; MAXIMUM LIKELIHOOD; MEASUREMENT ERROR MODELS; PROBIT REGRESSION; QUASILIKELIHOOD; SCORE TESTS

1. INTRODUCTION

1.1. Summary

Let x be a random variable with unknown density function f_x. Given x, the response y follows a generalized linear model with likelihood

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

where $\theta = g(\alpha_0 + \alpha_1 x)$. Based on an n-sample, we wish to test the hypothesis $H_0 : \alpha_1 = 0$. We are concerned with the case that x cannot be observed exactly, and instead we observe a proxy w. This is a generalized linear measurement error model. Examples from epidemiology motivating our work are discussed by Whittemore and Keller (1988), Carroll (1989), Tosteson, et al. (1989), Hasabelnaby, et al. (1989), Tosteson and Tsiatis (1988) and Armstrong, et al. (1989).

Tosteson and Tsiatis addressed the question of testing when measurement error is ignored. The usual score test statistic which ignores measurement error is asymptotically equivalent to

$$T_w = n^{-1/2} \sum_{i=1}^{n} w_i(y_i - \bar{y})/(s_w s_y), \tag{1.2}$$

where s_w^2 and s_y^2 are the usual sample variances of the w's and y's respectively. T_w has an asymptotic standard normal distribution under the null hypothesis, which is rejected when T_w exceeds standard normal critical points.

For a general function $q(w)$, let s_q^2 be the sample variance of the $\{q(w_i)\}$. Consider the class of test statistics

$$T_* = n^{-1/2} \sum_{i=1}^{n} q(w_i)(y_i - \bar{y})/(s_y s_q) = S_a/(s_y s_q). \tag{1.3}$$

The statistic T_* has an asymptotic standard normal distribution under the null hypothesis. In model (1.1), for local alternatives $\alpha_1 = \Delta/n^{1/2}$, T_* is asymptotically normally distributed with variance 1 and mean

$$\Delta g^{(1)}(\alpha_0) Cov\{m(w), q(w)\} \left[Var\{q(w)\}\gamma/b^{(2)}\{g(\alpha_0)\}\right]^{-1/2},$$

where $m(w) = E(x \mid w)$. Taking $q(w) = m(w)$ maximizes local power and yields a test which is asymptotically equivalent to the efficient score test. The usual test (1.2) can be inefficient when $m(w)$ is far from linear (Tosteson and Tsiatis, 1988).

Since $m(w)$ is usually unknown, it must be estimated to construct an efficient score test. The function $m(w)$ also plays an important role in the estimation of (α_0, α_1). Whittemore and Keller

(1988) show that knowledge of $m(w)$ and $Var(x \mid w)$ is sufficient to obtain useful and simple approximations to maximum likelihood estimators in many measurement-error models.

In this paper we discuss two alternatives to the usual test. The first is based on the small measurement error asymptotics of Stefanski and Carroll (1985), Stefanski (1985) and Amemiya and Fuller (1988). The second employs an estimate of $m(w) = E(x|w)$. The methods are illustrated on data from a study of saturated fat and breast cancer. We now outline the two alternative tests.

1.2. Wald Tests Based on Corrected Estimates

A standard model (Fuller, 1987) for measurement error is the homoscedastic additive model

$$w = x + z, \tag{1.4}$$

where the measurement error z is independent of x, has mean zero and variance σ_z^2. A method which has been used successfully for estimation is based on small measurement error approximations, see the previous references. Assuming that σ_z is small, an estimate $\hat{\alpha}_{1,c}(\sigma_z)$ is constructed, with estimated standard error $se\{\hat{\alpha}_{1,c}(\sigma_z)\}$ derived using the theory of M-estimation. The test statistic is $\hat{\alpha}_{1,c}(\sigma_z)/se\{\hat{\alpha}_{1,c}(\sigma_z)\}$, and the hypothesis is rejected when the test statistic exceeds standard normal critical values.

Data for checking (1.4) often are not available and thus tests that are robust to departures from the additive homoscedastic model are desired. We show in Section 2 that the Wald test statistic is robust to (1.4), i.e., it achieves its nominal level asymptotically and has the same local power as (1.2) regardless of the distribution of (x, w). Consequently when (1.4) holds, the estimator $\hat{\alpha}_{1,c}(\sigma_z)$, which corrects for much of the measurement error induced bias, can also be used for testing without sacrificing local power.

In the linear normal case, the fact that local power is not lost by correcting for attenuation is easily shown. Consider the classical simple linear regression errors-in-variables model (Fuller, 1987) $y = \alpha_0 + \alpha_1 x + \epsilon$, and $w = x + z$, under the additional assumptions that x, ϵ and z are uncorrelated random variables with variances $\sigma_x^2, \sigma_\epsilon^2$ and σ_z^2 respectively, and that σ_z^2 is known. Let $\hat{\alpha}_1$ denote the least squares estimator based on $(y_i, w_i)_1^n$ and let $\hat{\alpha}_{1,A} = \hat{D}\hat{\alpha}_1$ where $\hat{D} = s_w^2/(s_w^2 - \sigma_z^2)$. The estimate $\hat{\alpha}_{1,A}$ is a standard method-of-moments errors-in-variables estimator (Fuller, 1987, page 14). Under the local alternatives $\alpha_{1,n} = \Delta n^{-1/2}$, $n^{1/2}\hat{\alpha}_1 \Rightarrow \mathrm{N}(\mu_\Delta, \sigma_\Delta^2)$ say. Since $\hat{D} \rightarrow D = \sigma_w^2/\sigma_u^2$ in probability, it follows that $n^{1/2}\hat{\alpha}_{1,A} \Rightarrow \mathrm{N}(D\mu_\Delta, D^2\sigma_\Delta^2)$, and thus local power is not affected by correcting for attenuation. If all errors are normally distributed, however, the observed data and the least squares estimate are normally distributed and least squares would be preferred.

1.3. Approximating m(w)

Consider again the error model (1.4). Let f_w be the density of w and $f_w^{(1)}$ its derivative. Then in general as $\sigma_z \to 0$,

$$E(x|w) = w + \sigma_z^2 f_w^{(1)}(w)/f_w(w) + \mathcal{O}(\sigma_z^4). \tag{1.5}$$

Equation (1.5) is exact if (1.4) holds and z is normally distributed. It suggests use of the test (1.3) based on an estimate of $w + \sigma_z^2 f_w^{(1)}(w)/f_w(w)$. This estimate is itself interesting as an informal check on the hypothesis that $E(x \mid w)$ is linear in w. In Section 3, we show that the test is robust in the sense defined earlier and is approximately efficient when (1.4)–(1.5) hold, and fully efficient for the additive model (1.4) with normal errors.

Both methods assume knowledge or consistent estimability of σ_z. In cases where (1.4) fails, σ_z is defined as the limit in probability of $\hat{\sigma}_z$.

2. WALD TESTS BASED ON CORRECTED ESTIMATES

2.1. Introduction

Carroll (1989) reviews various methods for constructing estimates of the parameters (α_0, α_1). Among the more appealing estimates are those obtained by correcting the naive estimator for attenuation through a small measurement error analysis. Given a corrected estimate and its standard error, a test of the hypothesis $H_0 : \alpha_1 = 0$ is constructed in the usual manner. In this section we describe the level and local power of such tests. The starting point is Theorems 1 and 2 of Stefanski (1985), whose notation we follow. Our main result is Theorem 1. The usual and Wald tests are defined later in this section.

Theorem 1: Under local alternatives $\alpha_1 = \Delta/n^{1/2}$, the usual and corrected Wald tests attain their nominal level and have the same local power asymptotically. • •

Let $\theta = (\alpha_0, \ \alpha_1)^t$ be the true value of the parameter. Write $X_v = (1, \ \ w)^t$, $e^t = (0, \ \ 1)$; $\Omega = ee^t$. Let $\theta_0 = (\alpha_0, \ \ 0)^t$ be the value of θ under the hypothesis, and write

$$\ell_i\left(X_{v,i}^t \theta\right) = g^{(1)}(\alpha_0 + \alpha_1 w_i)\left[y_i - b^{(1)}\left\{g(\alpha_0 + \alpha_1 w_i)\right\}\right].$$

2.2. The Usual Wald Test

Let $\hat{\theta} = (\hat{\alpha}_0 \ \ \hat{\alpha}_1)^t$ be the maximum likelihood estimate of θ ignoring measurement error. Dropping the dependence on the individual observations, let $\theta(\sigma_z)$ satisfy

$$E\ell\left\{X_v^t \theta(\sigma_z)\right\} X_v = 0. \tag{2.1}$$

Appealing to Theorem 1 of Stefanski (1985), $n^{1/2}\{\hat{\theta} - \theta(\sigma_z)\} \rightarrow \mathcal{N}\left(0, A^{-1}BA^{-1}\right)$, where $A = E\ell^{(1)}\{X_v^t\theta(\sigma_z)\}X_vX_v^t$, $B = E\ell^2\{X_v^t\theta(\sigma_z)\}X_vX_v^t$. Here A and B are consistently estimated by $A_n(\hat{\theta})$ and $B_n(\hat{\theta})$, where $A_n(\theta) = n^{-1}\sum \ell_i^{(1)}(X_{v,i}^t\theta)X_{v,i}X_{v,i}^t$ and $B_n(\theta) = n^{-1}\sum \ell_i^2(X_{v,i}^t\theta)X_{v,i}X_{v,i}^t$. The usual Wald test statistic is $U = n^{1/2}e^t\hat{\theta}/\left\{e^tA_n^{-1}(\hat{\theta})B_n(\hat{\theta})A_n^{-1}(\hat{\theta})e\right\}^{1/2}$, since $e = (0,\ 1)^t$ and thus $\alpha_1 = e^t\theta$. The hypothesis is rejected if U is large relative to standard normal percentiles. In Appendix A, we prove that under the local alternative $\alpha_1 = \Delta/n^{1/2}$, $U \Rightarrow \mathcal{N}(a, 1)$, where for some function $\Lambda_2(\Delta)$,

$$a^2 = \Lambda_2(\Delta)Var(w)\left\{E\ell^{(1)}(\alpha_0)\right\}^2/E\ell^2(\alpha_0). \tag{2.2}$$

2.3. The Corrected Wald Test, σ_z known

We first assume that σ_z^2 is known, later showing that estimating it makes no difference to the results. Define $C_n(\theta) = n^{-1}\sum\{2\ell_i^{(1)}(X_{v,i}^t\theta) + \ell_i^{(2)}(X_{v,i}^t\theta)X_{v,i}\theta^t\}\Omega$. Let $C\{\theta(\sigma_z)\} = EC_n\{\theta(\sigma_z)\}$. The corrected estimate is

$$\hat{\theta}_c = \left\{I + .5\sigma_z^2A_n^{-1}(\hat{\theta})C_n(\hat{\theta})\right\}\hat{\theta}, \tag{2.3}$$

which converges to

$$\theta_c(\sigma_z) = \left[I + .5\sigma_z^2A^{-1}\{\theta(\sigma_z)\}C\{\theta(\sigma_z)\}\right]\theta(\sigma_z). \tag{2.4}$$

There exist matrices $V_n(\hat{\theta})$ and $V\{\theta(\sigma_z)\}$ such that $V_n(\hat{\theta})$ consistently estimates $V\{\theta(\sigma_z)\}$ and $n^{1/2}\{\hat{\theta}_c - \theta_c(\sigma_z)\} \Rightarrow \mathcal{N}[0, V\{\theta(\sigma_z)\}]$ (Stefanski, 1985). Under the null hypothesis, $C\{\theta(\sigma_z)\}\theta(\sigma_z) = (0 \quad 0)^t$, implying that $\theta(\sigma_z) = \theta_c(\sigma_z) = (\alpha_0\ 0)^t$. The corrected Wald test is

$$U_c = n^{1/2}e^t\hat{\theta}_c/\left\{e^tV_n(\hat{\theta})e\right\}^{1/2}. \tag{2.5}$$

In Appendix A, we show that under the local alternative $\alpha_1 = \Delta/n^{1/2}$,

$$U_c \Rightarrow \mathcal{N}(a, 1), \tag{2.6}$$

where a is defined by (2.2). This completes the proof of Theorem 1 when σ_z is known.

2.4. The Corrected Wald Test, σ_z unknown

Suppose that σ_z^2 is unknown, and is estimated by $\hat{\sigma}_z^2$ where

$$n^{1/2}(\hat{\sigma}_z^2 - \sigma_z^2) \Rightarrow \mathcal{N}(0, \tau^2). \tag{2.7}$$

The corrected estimate (2.3) becomes

$$\hat{\theta}_c^* = \left\{I + .5\hat{\sigma}_z^2A_n^{-1}(\hat{\theta})C_n(\hat{\theta})\right\}\hat{\theta}, \tag{2.8}$$

and its asymptotic covariance matrix is consistently estimated by

$$V_n^*(\hat{\theta}) = V_n(\hat{\theta}) + (1/4)\tau^2 A_n^{-1}(\hat{\theta})C_n(\hat{\theta})\hat{\theta}\hat{\theta}^t C_n^t(\hat{\theta})A_n^{-1}(\hat{\theta}).$$

A Wald statistic can be formed in the usual manner. However, since $C_n(\hat{\theta})\hat{\theta}$ is consistently estimating $C\{\theta(\sigma_z)\}\theta(\sigma_z) = 0$ under the null hypothesis, we choose to define the corrected Wald statistic as

$$U_c^* = n^{1/2} e^t \hat{\theta}_c^* / \left\{ e^t V_n(\hat{\theta}) e \right\}^{1/2}. \tag{2.9}$$

The equivalence of (2.5) and (2.9) with respect to local power when (2.7) holds follows from the fact that $C\{\theta(\sigma_z)\}\theta(\sigma_z) = 0$ under the null hypothesis. Condition (2.7) can be replaced by the weaker assumption that the right side of (2.7) is $O_p(1)$.

3. AN APPROXIMATELY EFFICIENT SCORE TEST

3.1. Introduction and Theory

Let $\hat{\sigma}_z^2$ be a consistent estimate of σ_z^2. Let $q(w) = w + \sigma_z^2 f_w^{(1)}(w)/f_w(w)$. Following (1.3), if s_q^2 is the sample variance of the $\{q(w_i)\}$, an approximately efficient score test is

$$T_a = n^{-1/2} \sum_{i=1}^{n} q(w_i)(y_i - \bar{y})/(s_y s_q) = S_a/(s_y s_q). \tag{3.1}$$

Since σ_z and f_w are unknown, T_a is not a statistic. In constructing an estimated version of (3.1), special care must be taken with the tail behavior of f_w. We will do this through censoring, truncation and special choice of the estimate of f_w.

Fix a sequence of constants $\eta_n \to 0$ and let C be a set with indicator function $I(w \in C)$. Let K be a bounded symmetric density function and define a leave-one-out estimate for f_w by

$$\hat{f}_{w,i}(v) = \{\lambda(n-1)\}^{-1} \sum_{j \neq i} K\{(w_j - v)/\lambda\},$$

and let $\hat{f}_{w,i}^{(1)}(v)$ be the derivative of $\hat{f}_{w,i}(v)$. Now define an approximation to $m(w)$ applicable on the set C:

$$\hat{q}_i(w) = I(w \in C)\left[w + \hat{\sigma}_z^2 \hat{f}_{w,i}^{(1)}(w) / \left\{ \hat{f}_{w,i}(w) + \eta_n \right\} \right]. \tag{3.2}$$

Censoring comes from η_n, which insures that the denominator in (3.2) does not approach zero. If we do not force censoring, truncation comes from the set C. In our main result, we consider the cases that censoring or truncation is used, along with the case that neither is used and $\eta_n = 0$ and C is the whole line.

If $\hat{s}_q^2$ is the sample variance of the $\{\hat{q}_i(w_i)\}$, an estimated version of (3.1) is

$$\hat{T}_a = n^{-1/2} \sum_{i=1}^{n} \hat{q}_i(w_i)(y_i - \bar{y})/(s_y \hat{s}_q) = \hat{S}_a/(s_y \hat{s}_q). \tag{3.3}$$

Note that (3.2) is an estimated version of $q_*(w) = I(w \in \mathcal{C})\Big\{w + \sigma_z^2 f_w^{(1)}(w)/f_w(w)\Big\}$. The larger the set $\mathcal{C}$, the closer $q_*(w)$ is to $q(w)$, see the remark below.

The test statistic (3.3) was suggested by heuristics based on the additive error model. However, in order to be broadly applicable, we need that $\hat{T}_a$ attains its nominal level asymptotically even when (1.4) fails to hold. In addition, in order to be approximately efficient, we want $\hat{T}_a$ and T_a to have the same behavior, at least asymptotically for local alternatives. Both these desirable features hold in our case.

In the following result, we show that the numerators of (3.3) and (3.1) with q replaced by q_* are asymptotically equivalent under the null hypothesis, i.e., $\hat{S}_a - S_a \to 0$ in probability. Strictly speaking, the equivalence of (3.1) and (3.3) also requires that $\hat{s}_q^2 - s_q^2 \to 0$ in probability. The latter result uses nearly identical techniques as does the former, and is not given here. The asymptotic equivalence of the tests under local alternatives follows from contiguity.

THEOREM 2: Let w be any random variable, i.e., not necessarily following the error model (1.4) for additive, homoscedastic z. Assume that f_w is bounded and three times boundedly and continuously differentiable. Assume further that

$$E\Big\{f_w^{(1)}(w)/f_w(w)\Big\}^2 < \infty. \tag{3.4}$$

Then, under the hypothesis, $\hat{S}_a - S_a \to 0$ in probability if any of the following hold:

Case 1 (Censoring): $\lambda^2/\eta_n^2 \to 0$ and $n\lambda^3\eta_n^4 \to \infty$. (3.5)

Case 2 (Truncation): $|K^{(1)}(a)/K(a)| \le c < \infty$ for all $-\infty < a < \infty$, $\eta_n = 0$, $\mathcal{C}$ is compact, f_w is positive on $\mathcal{C}$ and $n\lambda^3 \to \infty$. (3.6).

Case 3 (Neither censoring nor truncation): Make the assumptions of Case 2, except that $\mathcal{C}$ need not be compact. Define $h_n(v) = E\hat{f}_{w,i}(v)$ and $L_n(v) = \int K(x) f_w(v + \lambda x)dx$, $R_n(v) = \int K^2(x) f_w(v + \lambda x)dx/h_n^2(v)$, $H_n(x, v) = \sup_{|\alpha| \le 1} |f_w^{(2)}(v + \alpha x\lambda)/f_w(v)|$. Assume that

$$E\left\{\lambda^2 \int K(x) x^{2(1+p)} H_n^2(x, w)dx\right\} \to 0, \tag{3.7}$$

$$E\left\{\lambda^{-2} \min\left[c^{2(1+p)}, R_n(w)/(n\lambda)\right]\right\} \to 0, \tag{3.8}$$

for $p = 0, 1$. • •

Remark 1: The practical difficulty with Case 1 is that the constants η_n must be specified. In Case 2, the set C should be chosen so that there are enough points that f_w can be estimated accurately, but C is big enough so that any curvature in the regression function is noticable. Case 3 is the easiest to implement since it depends only on the choice of the kernel function and the bandwidth. Kernel functions satisfying the conditions of Cases 2 and 3 include the t-densities. For Case 3, conditions (3.7) and (3.8) look awkward but arise naturally in the proof. • •

Remark 2: That the test statistic $\hat{T}_a$ attains its nominal value asymptotically even when (1.4) fails means that the test can be applied even when a homoscedastic additive error model does not hold. • •

Remark 3: It would be surprising if the choice of kernel had a significant effect on the estimated test statistic and it seems likely that Theorem 2 holds in greater generality. The simulation results in the next subsection were obtained using a normal kernel with $\eta_n = 0$ and C the whole real line, a case not explicitly covered by the theorem, and the choice of kernel did not matter much in the application of Section 4. • •

3.2. Simulation Results

The relevance of Theorem 2 in finite samples is unclear due to the complexity of $\hat{T}_a$ and the technical nature of (3.7) - (3.8). A Monte Carlo experiment was performed to study the statistic's performance in typical applications, i.e., large samples and substantial measurement error, see Section 4. In our experiment $\hat{T}_a$ performed comparably to the usual test with respect to level and out performed the usual test at local alternatives.

We considered a logistic model with $Pr(y = 1 \mid x) = F(\alpha_0 + \alpha_1 x)$, where $F(t) = 1/\{1 + \exp(-t)\}$. The true predictor x was generated as a χ^2_4 random variable centered and scaled to mean zero and unit variance. Four parameter settings $(\alpha_{0,i}, \alpha_{1,i})$ (i=1,2,3,4) were studied. These were determined so that $\Pr(y = 1) = 0.1$ and $\mathrm{Corr}(y, x) = 0$, $1/40$, $1/20$ and $1/10$ respectively. Observed predictors were generated according to (1.4) with $z \sim \mathcal{N}(0,\ 0.75)$. Sample size was fixed at $n = 300$. Similar results were obtained for $n = 2,500$ but are not reported here.

The estimate of $m(w)$ was constructed using a normal kernel without truncation or censoring and assuming σ_z known. Obviously the performance of the test will depend on the quality of the information on σ_z. In our experiment we attempted only to establish the credibility of $\hat{T}_a$ in the case of complete knowledge of σ_z.

We required a fast and reliable bandwidth selection procedure for the simulation runs. Our solution to this problem was motivated by the normal discriminant derivation of the logistic model. Let

$$f_{ND}(w) = \sum_{i=0}^{1} \pi_i \phi\left(\frac{w - \mu_i}{\sigma_i}\right) / \sigma_i, \tag{3.9}$$

where ϕ is the standard normal density. Define

$$c_{ND} = c_{ND}(\pi_0, \pi_1, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2) = \int \left\{ f_{ND}^{(3)}(w) \right\}^2 dw, \tag{3.10}$$

where $f_{ND}^{(3)}$ is the third derivative of f_{ND}. Partition the data $(y_i, w_i)_1^n$ according to whether y_i $= 0$ or 1. Let $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ be the sample mean and variance of the w's for the groups defined by $y = j, j = 0, 1$. Define $\hat{\pi}_0 = 1 - \bar{y}$ and $\hat{\pi}_1 = \bar{y}$. Let $\hat{c}_{ND}$ be as in (3.10) with parameters replaced by their estimates. The bandwidth used in the simulation was

$$\hat{\lambda} = \left\{ 3/(4n\hat{c}_{ND}\pi^{1/2}) \right\}^{1/7}. \tag{3.11}$$

If w has the density (3.9) and if $\hat{f}$ is a kernel density estimator of f constructed using a normal kernel, then the choice of bandwidth λ defined by replacing $\hat{c}_{ND}$ by c_{ND} minimizes IMSE$(\hat{f}^{(1)}, \lambda)$ asymptotically, so the use of (3.11) is natural.

We performed 1,000 replications for each parameter set and experiments at each parameter setting are independent. Table 3.1 summarizes the results of using $\hat{T}_a$ and the usual test statistic for testing H$_0$: $\alpha_1 = 0$ vs. H$_a$: $\alpha_1 > 0$ at level .025. The two tests were compared using a paired t-test on the binary indicators of rejection.

Table 3.1

Power comparison of the usual test and the test statistic (3.3).

Corr(y, x)	Usual Test	EIV Test (3.3)	T–statistic
0.000	0.026	0.030	1.26
0.025	0.043	0.054	2.20
0.050	0.097	0.115	3.01
0.100	0.255	0.297	5.15

The additional power in $\hat{T}_a$ is modest although statistically significant. Differences in distributions were more profound. The statistic $\hat{T}_a$ had a smaller variance and larger mean than the usual test statistic although only the latter difference was highly significant. It was encouraging that the level of the test statistic (3.3) was approximately correct.

The amount of additional power in $\hat{T}_a$ depends on the extent of nonlinearity in $m(w)$. For certain distributions on x and z, $m(w)$ can be highly nonlinear; however, for distributions more commonly encountered in practice, such as those used in the monte carlo experiment, $m(w)$ has a strong linear component and the gains realized by $\hat{T}_a$ will be modest though not insignificant.

For the simulation we were forced to automate the smoothing in $\hat{m}(w)$. In applications we prefer to adopt an exploratory attitude to the smoothing and testing problems and it is in this context that we suggest the use of $\hat{T}_a$.

4. AN EXAMPLE

We consider a data set of $n = 2888$ women under the age of 50 at examination, with predictor variable x = long-term log daily saturated fat intake and response y = incidence of breast cancer. The infrequency of breast cancer (37 cases) makes the analysis of these data inherently difficult.

Controversy exists regarding the role of saturated fat in the development of breast cancer. The position that fat is protective is suggested by Jones, et al. (1987). However, although the test statistic (1.2) has the right sign, it is not statistically significant. We explore the effect of measurement error in x on the analysis of these data.

Saturated fat is not the only predictor of breast cancer, for example age at menarche and age at examination are related to breast cancer. These and other factors are correlated with saturated fat and thus the interpretation of a significant effect due to saturated fat in a univariate analysis is subject to debate. We undertake a univariate analysis of these data for illustrative purposes, fully aware of the inherent limitations of this approach.

The problem of concern in this paper is measurement error. Instead of observing x exactly, we observe w, a 24-hour recall measurement of log saturated fat intake. It is known that 24-hour recall measurements of dietary variables can exhibit considerable error, sometimes orders of magnitude larger than is standard in the physical sciences. Discussion of this issue is given by Armstrong, et al. (1989), Beaton, et al. (1979), Jones, et al. (1987), Liu, et al. (1978), Willett, et al. (1985) and Wu, et al. (1986).

The usual device for understanding the size and nature of the measurement error is a validation study, i.e., some of the study participants are examined over a longer period of time and their daily measurements related to the long term averages. Sometimes these data are taken from a different population, so they yield no reliable information about the distribution of w or x given w. However, it is reasonable to assume that the measurement error distribution carries over from the validation study, thus yielding the information we need about σ_z^2. We do not have validation data specific to our our data set, but it seems reasonable to use the validation results of Willett, et al. (1985). They do not transform fat intake, using seven-day diet record measurements and finding that the correlation between two weekly measurements is approximately 0.55. If we assume normal measurement error in (1.4), then their data suggest that for their study, $\sigma_z = 0.34$. Their seven day diet records differ from our 24-hour recall measurements and should be more precise. It is not clear how to reconcile these figures, but as a reasonable guess we suppose that the seven day measurement error standard deviations are 60% of the 24-hour standard deviation, and we use $\sigma_z = .55$; Because of the uncertainties associated with this estimate, as well as others discussed previously, the following analysis is illustrative only. Our intention is to show that the techniques we have developed provide a means of studying the effects of measurement error on conclusions drawn from a statistical analysis.

We first ran a logistic regression ignoring measurement error, and found a parameter estimate $\hat{\alpha}_{1u} = -.40$ with estimated standard error $se(\hat{\alpha}_{1u}) = .23$ and significance level 0.08. The sign of the estimate and the lack of statistical significance is in line with the study by Jones, et al. (1987). Additionally, the usual test (1.2) had value -1.76, with significance level 0.08.

The corrected estimator of Section 2 and its estimated standard error are $\hat{\alpha}_{1,c}(\sigma_z) = -.61$ and $se\{\hat{\alpha}_{1,c}(\sigma_z)\} = .24$. As expected, the corrected estimates adjust for attenuation, but now the significance level is 0.01.

We now turn to the methods of Section 3. Good estimates of f_w and $f_w^{(1)}$ can be obtained here because of the large size of the data set.

In Figure 1 we plot an ordinary kernel estimate of the density of the observed data, using as our kernel the standard normal density function with bandwidth $\lambda = 0.4$. This estimate does not exactly satisfy the conditions of Theorem 2, but we have employed other kernels with similar results, see Section 3. Conclusions were relatively insensitive to choice of bandwidth in a wide range. The solid line is the full data set of 2888 observations, while the dashed line is for the 37 cases of breast cancer. The latter is meant to be illustrative only, as density estimation with such a small data set is problematic. The dashed line seems slightly shifted to the left, in keeping with a hypothesis that fat is protective. In Figure 2, we plot the approximation (3.2) to $E(x|w)$, again with the solid line being the full data and the dashed line those with breast cancer. We used the larger bandwidth λ $= 0.6$ because of the need to estimate the derivative of f_w. The noticable curvature here is in part influenced by a few points in the left tail of the distribution of w. This left tail is rather long, with $.5^{th}$ and 1^{st} percentiles 3.7 and 3.1 standard deviations from the mean respectively. The direction of the curvature works to make the usual test less significant. Our approximate test statistic has value -1.50 and significance level 0.13.

Both the usual and the approximate score tests are influenced by the long left tail in w. For example, if we eliminate all data points below the 1st percentile of w, then the usual test takes the value -2.12, while the approximate test takes the value -2.01. These fairly large changes indicate the lack of robustness of score tests.

5. CONCLUSIONS

We have shown that the use of corrected estimates in Wald tests yields no loss in local power for testing a null hypothesis when compared to the usual test which ignores measurement error. Improved power can be obtained by our approximate tests, which also yield an estimate of the regression function $E(x|w)$.

ACKNOWLEDGEMENT

Our research was supported by the National Science Foundation, the Air Force Office of Scientific Research, and the National Institutes of Health. We thank S. Green, A. Schatzkin and Y. Wax for helpful discussions.

REFERENCES

Amemiya, Y. and Fuller, W. A. (1988). Estimation for the nonlinear functional relationship. *Ann. Statist.*, 16, 147-160.

Armstrong, B., Howe, G. and Whittemore, A. S. (1989). Correcting for measurement error in a nutrition study. *Statistics in Medicine*, to appear.

Beaton, G. H., Milner, J. and Little, J. A. (1979). Sources of variation in 24-hour dietary recall data: implications for nutrition study design and interpretation. *American Journal of Clinical Nutrition*, 32, 2546-2559.

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

Fuller, W. A. (1987). *Measurement error models.* Wiley, New York.

Hasabelnaby, N., Fuller, W. A. and Ware, J. (1989). A measurement error model for the Six Cities data. *Statistics in Medicine*, to appear.

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, K., Stamler, J., Dyer, A., McKeever, J. and McKeever, P. (1978). Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *J. Chron. Dis.*, 31, 399-418.

Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika*, 72, 583-592.

Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.*, 13, 1335-1351.

Tosteson, T., Stefanski, L. A. and Schafer, D. W. (1989). A measurement error model for binary and ordinal regression. *Statistics in Medicine*, to appear.

Tosteson, T. and Tsiatis, A. (1988). The asymptotic relative efficiency of score tests in the generalized linear model. *Biometrika*, 75, 507–514.

Whittemore, A. S. and Keller, J. B. (1988). Approximations for errors in variables regression. *J. Amer. Statist. Assoc.*, 83, 1057–1066.

Willett, W. C., Sampson, L., Stampfer, M. J., Rosner, B., Bain, C., Witschi, J., Hennekens, C. H. and Speizer, F. E. (1985). Reproducibility and validity of a semiquantitative food frequency questionairre. *Amer. J. Epid.*, 122, 51-65.

Wu, M. L., Whittemore, A. S. and Jung, D. L. (1986). Errors in reported dietary intakes. *Amer. J. Epid.*, 124, 826-835.

APPENDIX A: PROOF OF THEOREM 1

Throughout this appendix, we reserve the symbol $\theta_0 = (\alpha_0 \quad 0)^t$ for the null hypothesis value. When matrices such as A, B, C, V, etc. are used, we mean $A(\theta_0)$, etc.

First consider the usual Wald statistic. Let $\theta(\sigma_z, \Delta)$ be the solution to (2.1) under the local alternative $\alpha_1 = \Delta/n^{1/2}$. Assume that for some $\Lambda(\Delta) = \{\Lambda_1(\Delta), \quad \Lambda_2(\Delta)\}^t$,

$$n^{1/2}\{\theta(\sigma_z, \Delta) - \theta_0\} \longrightarrow \Lambda(\Delta), \tag{A.1}$$

with $\Lambda(0) = 0$. It follows that under these local alternatives

$$U \Rightarrow \mathcal{N}\left(\Lambda_2(\Delta)/\left\{e^t A^{-1} B A^{-1} e\right\}^{1/2}, 1\right).$$

Since $\Lambda_2(0) = 0$, U attains its nominal level. A detailed examination shows that under the null hypothesis, $e^t A^{-1} B A^{-1} e = E\ell^2(\alpha_0)/[\{E\ell^{(1)}(\alpha_0)\}^2 Var(w)]$, thereby verifying (2.2).

Now consider the corrected Wald statistic. Let $\theta_c(\sigma_z, \Delta)$ be defined as in (2.4), but with $\theta(\sigma_z)$ replaced by $\theta(\sigma_z, \Delta)$, see (A.1). Studying the structure of (2.5), it suffices to show that under the hypothesis,

$$e^t V(\theta_0) e = E\ell^2(\alpha_0)\left\{1 + \sigma_z^2/Var(w)\right\}^2 / \left[\left\{E\ell^{(1)}(\alpha_0)\right\}^2 Var(w)\right] \qquad (A.3)$$

and

$$n^{1/2} e^t \left\{\theta_c(\sigma_z, \Delta) - \theta_c(\sigma_z, 0)\right\} \longrightarrow \Lambda_2(\Delta)\left\{1 + \sigma_z^2/Var(w)\right\}. \qquad (A.4)$$

Verifying (A.3) procedes as follows. Under the hypothesis, $C(\theta_0) = 2E\ell^{(1)}(\alpha_0)\Omega$. Since $\Omega\theta_0 = (0,\ 0)^t$, $C(\theta_0)\theta_0 = (0,\ 0)^t$ and also in Stefanski's notation, $D = G = 0$. Dropping the dependence on $\theta(\sigma_z)$, Stefanski's $H = I + .5\sigma_z^2 A^{-1} C$, and $V = EWW^t$, where $W = -\ell(\alpha_0) H A^{-1} X_v$. Thus, $V = E\ell^2(\alpha_0) H A^{-1} H^t / E\ell^{(1)}(\alpha_0)$, from which (A.3) is immediate.

Verifying (A.4) is cumbersome. Note that $A(\theta)$ and $C(\theta)$ are continuous functions of θ. Writing $\theta_0 = (\alpha_0,\ \ 0)^t$, (A.4) follows if we can show that

$$e^t \left\{I + .5\sigma_z^2 A^{-1}(\theta_0) C(\theta_0)\right\} \Lambda(\Delta) = \Lambda_2(\Delta)\left\{1 + \sigma_z^2/Var(w)\right\}; \qquad (A.5)$$

$$n^{1/2} e^t A^{-1}(\theta_0)\left[C\left\{\theta(\sigma_z, \Delta)\right\} - C(\theta_0)\right]\theta_0 \longrightarrow 0; \qquad (A.6)$$

$$n^{1/2} e^t \left[A^{-1}\left\{\theta\left(\sigma_z, \Delta\right)\right\} - A^{-1}(\theta_0)\right] C(\theta_0)\theta_0 \longrightarrow 0. \qquad (A.7)$$

Equation (A.5) is readily verified, and (A.6)-(A.7) follow since $C(\theta)\theta_0 = (0,\ 0)^t$ for all θ. This verifies (2.6) and completes the proof of Theorem 1.

APPENDIX B: PROOF OF THEOREM 2

Lemma B.1: Define the identically distributed random variables

$$a_{i,n} = I(w_i \in \mathcal{C})\left\{\frac{\hat{f}_{w,i}^{(1)}(w_i)}{\hat{f}_{w,i}(w_i) + \eta_n} - \frac{f_w^{(1)}(w_i)}{f_w(w_i) + \eta_n}\right\}.$$

Then $\hat{S}_a - S_a = o_p(1)$ if $Ea_{1,n}^2 \rightarrow 0$.

Proof of Lemma B.1: Let $E(y) = \mu_y$. Write $\hat{S}_a - S_a = A_{1,n} + A_{2,n}$, where

$$A_{1,n} = -\sigma_z^2 n^{-1/2} \sum_{i=1}^{n} I(w_i \in \mathcal{C})(y_i - \bar{y})\, \frac{f_w^{(1)}(w_i)}{f_w(w_i)}\, \frac{\eta_n}{f_w(w_i) + \eta_n}.$$

In $A_{1,n}$, replace $y_i - \bar{y}$ by $(y_i - \mu_y) + (\mu_y - \bar{y})$, forming the two terms $A_{1,n,1} + A_{1,n,2}$. To show that $A_{1,n,1} = o_p(1)$, note that it has expectation 0 and variance given by

$$Var(A_{1,n,1}) \le \sigma_z^4 Var(y) E\left\{\frac{f_w^{(1)}(w)}{f_w(w)}\, \frac{\eta_n}{f_w(w) + \eta_n}\right\}^2 \qquad (B.1)$$

This converges to zero by (3.4) and dominated convergence. Since $\bar{y} - \mu_y = \mathcal{O}_p(n^{-1/2})$, $A_{1,n,2} = o_p(1)$ if

$$n^{-1}\sum_{i=1}^{n} I(w_i \in \mathcal{C})\ \frac{f_w^{(1)}(w_i)}{f_w(w_i)}\ \frac{\eta_n}{f_w(w_i)+\eta_n} = o_p(1),$$

which follows since, applying Cauchy-Schwarz, the individual terms have absolute values whose expectations are bounded by a constant times the square root of the right hand side of (B.1). We now turn to $A_{2,n}$, writing it as $A_{2,n} = A_{2,n,1} + \hat{\sigma}_z^2(A_{2,n,2} + A_{2,n,3})$, where

$$A_{2,n,1} = (\hat{\sigma}_z^2 - \sigma_z^2)n^{-1/2}\sum_{i=1}^{n} I(w_i \in \mathcal{C})(y_i - \bar{y})f_w^{(1)}(w_i)/\{f_w(w_i)+\eta_n\};$$

$$A_{2,n,2} = n^{-1/2}\sum_{i=1}^{n} I(w_i \in \mathcal{C})(y_i - \mu_y)\left\{ \frac{\hat{f}_{w,i}^{(1)}(w_i)}{\hat{f}_{w,i}(w_i)+\eta_n} - \frac{f_w^{(1)}(w_i)}{f_w(w_i)+\eta_n} \right\};$$

$$A_{2,n,3} = n^{-1/2}\sum_{i=1}^{n} I(w_i \in \mathcal{C})(\mu_y - \bar{y})\left\{ \frac{\hat{f}_{w,i}^{(1)}(w_i)}{\hat{f}_{w,i}(w_i)+\eta_n} - \frac{f_w^{(1)}(w_i)}{f_w(w_i)+\eta_n} \right\};$$

Since $\hat{\sigma}_z^2 - \sigma_z^2 = o_p(1)$, $A_{2,n,1} = o_p(1)$ by (3.4) and a simple calculation analagous to that involving (B.1). Since $\mu_y - \bar{y} = \mathcal{O}_p(n^{-1/2})$, $A_{2,n,3} = o_p(1)$ as long as

$$n^{-1}\sum_{i=1}^{n} a_{i,n} = o_p(1). \qquad (B.2)$$

Of course, (B.2) follows from the assumption that $Ea_{1,n}^2 \to 0$. Finally, $A_{2,n,2}$ has mean zero and variance $Var(y)Ea_{1,n}^2 \to 0$. This completes the proof of Lemma B.1. • •

Define $h_n(v) = E\hat{f}_{w,1}(v)$ and let $h_n^{(1)}$ be its derivative. Write $a_{1,n} = \sum_1^4 \mathcal{L}_j(w_1)$, where

$$\mathcal{L}_1(v) = I(v \in \mathcal{C})\left\{h_n^{(1)}(v) - f_w^{(1)}(v)\right\}/\left\{f_w(v)+\eta_n\right\};$$

$$\mathcal{L}_2(v) = I(v \in \mathcal{C})h_n^{(1)}(v)\Big[\ \{h_n(v)+\eta_n\}^{-1} - \ \{f_w(v)+\eta_n\}^{-1}\Big];$$

$$\mathcal{L}_3(v) = I(v \in \mathcal{C})\left\{\hat{f}_{w,1}^{(1)}(v) - h_n^{(1)}(v)\right\}/\left\{h_n(v)+\eta_n\right\};$$

$$\mathcal{L}_4(v) = I(v \in \mathcal{C})\hat{f}_{w,1}^{(1)}(v)\left\{h_n(v) - \hat{f}_{w,1}(v)\right\}\ \{\hat{f}_{w,1}(v)+\eta_n\}^{-1}\ \{h_n(v)+\eta_n\}^{-1}.$$

The proof of Theorem 2 is accomplished by showing that $E\mathcal{L}_j^2(w_1) \to 0$, j=1,2,3,4, and hence that $Ea_{1,n}^2 \to 0$, plus an appeal to Lemma B.1. We make use of the inequalities

$$\mathcal{L}_2^2(v) \le (c/\lambda)^2 I(v \in \mathcal{C})\left[\left\{h_n(v) - f_w(v)\right\}/f_w(v)\right]^2; \qquad (B.3)$$

$$\mathcal{L}_4^2(v) \leq (c/\lambda)^2 I(v \in \mathcal{C}) \left[\left\{ h_n(v) - \hat{f}_{w,1}(v) \right\} / h_n(v) \right]^2, \qquad (B.4)$$

which hold when $\eta_n = 0$ and $|K^{(1)}(a)/K(a)| \leq c$. • •

Proof of Theorem 2, Case 1: Recall that f_w has three bounded and continuous derivatives. By standard results of kernel density estimation, there is a constant c_* such that for all v,

$$|h_n^{(1)}(v) - f_w^{(1)}(v)| \leq c_* \lambda^2; \qquad (B.5)$$

$$|h_n(v) - f_w(v)| \leq c_* \lambda^2; \qquad (B.6)$$

$$E\left\{ \hat{f}_{w,1}^{(1)}(v) - h_n^{(1)}(v) \right\}^2 \leq c_*(n\lambda^3)^{-1}; \qquad (B.7)$$

$$E\left\{ \hat{f}_{w,1}(v) - h_n(v) \right\}^2 \leq c_*(n\lambda)^{-1}. \qquad (B.8)$$

From (B.5), $E\mathcal{L}_1^2(w_1) = \mathcal{O}\{(\lambda/\eta_n)^2\} = o(1)$. From (B.6), since $h_n^{(1)}$ is bounded, $E\mathcal{L}_2^2(w_1) = \mathcal{O}\{(\lambda/\eta_n)^2\} = o(1)$. From (B.7), $E\mathcal{L}_3^2(w_1) = \mathcal{O}\{(n\lambda^3\eta_n^2)^{-1}\} = o(1)$. Finally, we consider $\mathcal{L}_4$. We have that

$$\left\{ E|\mathcal{L}_4(w_1)| \right\}^2 \leq \eta_n^{-4} E\left\{ \hat{f}_{w,1}^{(1)}(w_1) \right\}^2 E\left\{ h_n(w_1) - \hat{f}_{w,1}(w_1) \right\}^2.$$

From (B.7), noting that $h_n^{(1)}$ is bounded and possibly by increasing c_*, $E\{\hat{f}_{w,1}^{(1)}(w_1)\}^2 \leq c_*\{1 + (n\lambda^3)^{-1}\}$, so that from (B.8), for some c_{**}, $\left\{ E|\mathcal{L}_4(w_1)| \right\}^2 \leq c_{**}\eta_n^{-4}(n\lambda)^{-1}\{1 + (n\lambda^3)^{-1}\}$. • •

Proof of Theorem 2, Case 2: Inequalities (B.3)–(B.4) hold. Since $\mathcal{C}$ is compact and f_w and h_n are strictly positive on $\mathcal{C}$, the proof follows the same outline as in Case 1, as we require now only bias and variance properties of kernel density estimates and their derivatives. • •

Proof of Theorem 2, Case 3: From a first order Taylor series, we have that $|\mathcal{L}_1(v)| \leq \lambda \int |z| H_n(z, v) K(z) dz$. Since K is a density, it follows that $\mathcal{L}_1^2(v) \leq \lambda^2 \int z^2 H_n^2(z, v) K(z) dz$, so that by (3.7), $E\mathcal{L}_1^2(w_1) \to 0$. Similarly, employing (B.3), we have that $\mathcal{L}_2^2(v) \leq (\lambda c/2)^2 \int z^4 H_n^2(z, v) K(v) dv$, so that by (3.7), $E\mathcal{L}_1^2(w_1) \to 0$. To complete the proof, we must show that $E\{\mathcal{L}_3(w_1) + \mathcal{L}_4(w_1)\}^2 \to 0$. Note that $\mathcal{L}_3(v) + \mathcal{L}_4(v) = \mathcal{L}_5(v)$, where $\mathcal{L}_5^2(v) = \left[\{\hat{f}_{w,1}^{(1)}(v)/\hat{f}_{w,1}(v)\} - \{h_n^{(1)}(v)/h_n(v)\} \right]^2$. By the assumptions on K, $\mathcal{L}_5^2(v) \leq (2c/\lambda)^2$. We have that $(1/4)\mathcal{L}_5^2(v) \leq \min\{(c/\lambda)^2, \mathcal{L}_3^2(v)\} + \min\{(c/\lambda)^2, \mathcal{L}_4^2(v)\}$. Given $w_1 = v$, we have that

$$E\mathcal{L}_3^2(v) \leq (n\lambda^3)^{-1} \int \left\{ K^{(1)}(z) \right\}^2 f(v + \lambda z) dz / h_n^2(v) \leq c^2 (n\lambda^3)^{-1} R_n(v),$$

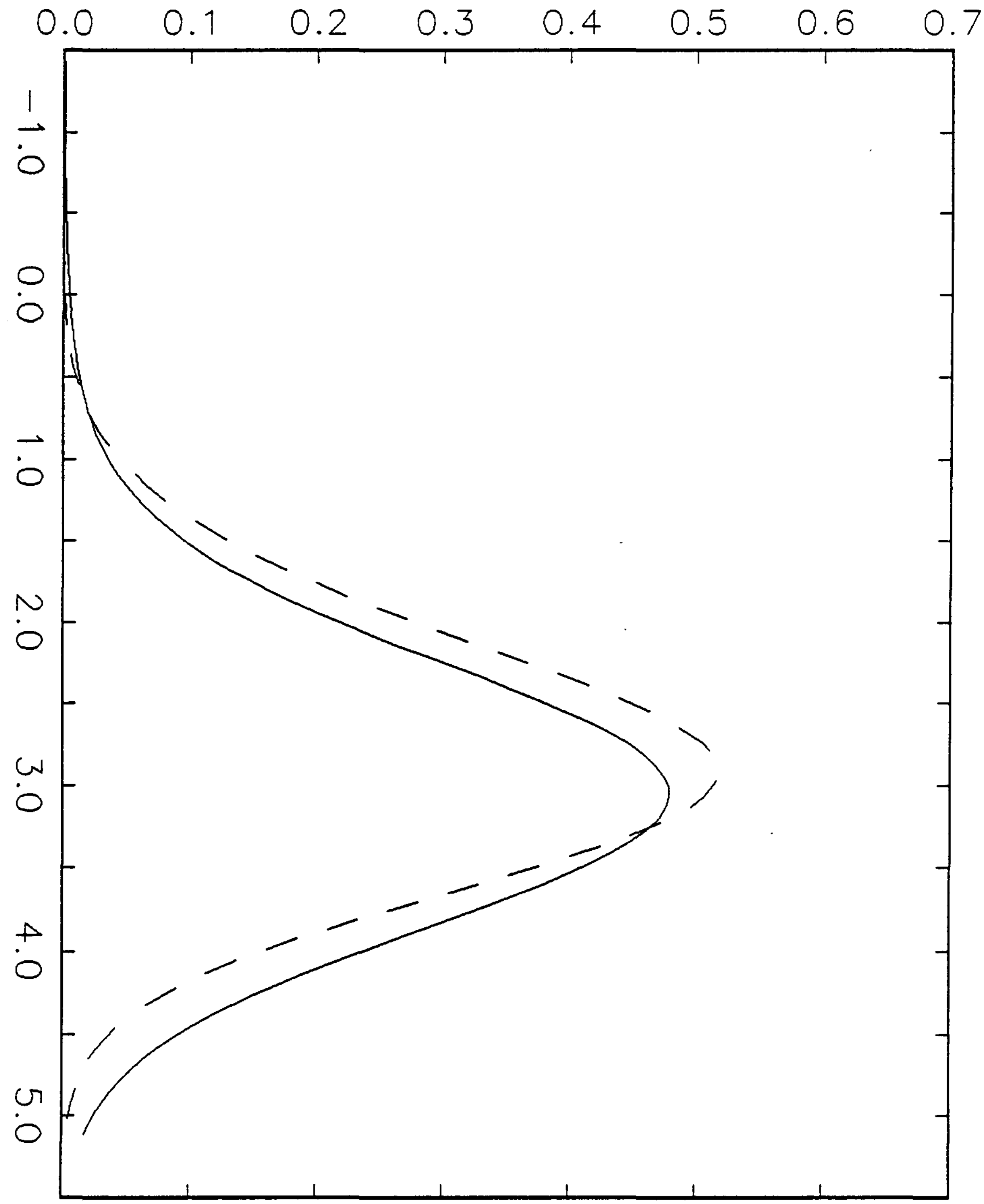
the last step following since $|K^{(1)}(v)/K(v)| \leq c$. Similarly, from (B.4), $E\mathcal{L}_4^2(v) \leq c^2(n\lambda^3)^{-1}R_n(v)$. In light of (3.6), it follows that

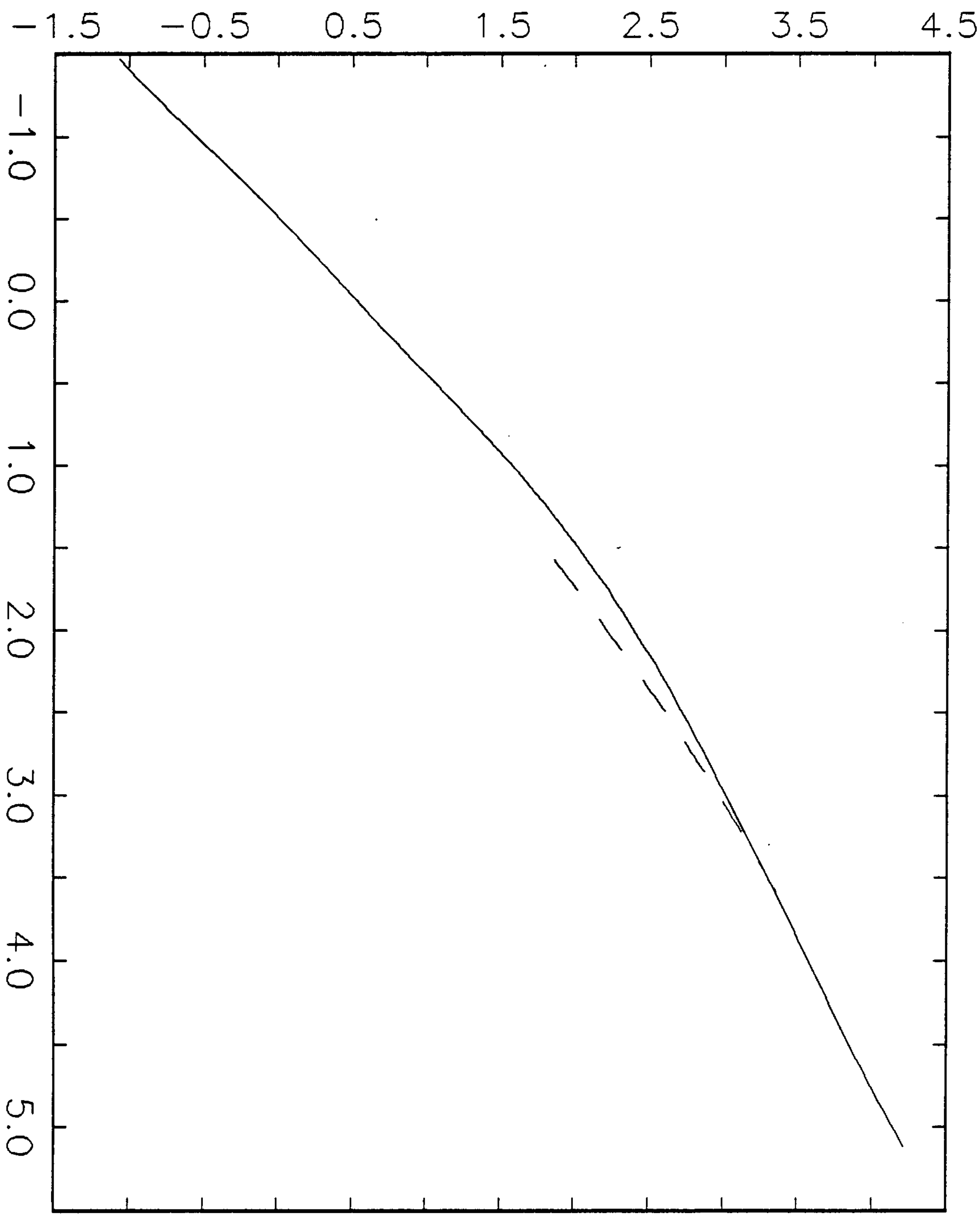
$$E\min\left\{(c/\lambda)^2, \mathcal{L}_5^2(w_1)\right\} \leq c^2(n\lambda^3)^{-1}E\min\{n\lambda, R_n(w_1)\} \to 0.$$

FIGURES

Figure 1: A plot of density estimates with bandwidth $\lambda = 0.4$. The solid line is the density estimate based on all the data, while the dashed line is the estimate for the 37 cases of breast cancer.

Figure 2: A plot of the approximation (1.5) with bandwidth $\lambda = 0.6$. The solid line is the estimate for all the data, while the dashed line is for the 37 cases of breast cancer. The latter is restricted to the range of the data.


0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
−1.0 0.0 1.0 2.0 3.0 4.0 5.0



−1.5
−0.5
0.5
1.5
2.5
3.5
4.5
−1.0
0.0
1.0
2.0
3.0
4.0
5.0