Semiparametric Specification Testing

DISCUSSION PAPER #778

by

Miguel A. Delgado[1]
and
Thanasis Stengos[2]

March, 1990

1. Department of Economics, Indiana University, Bloomington, IN 47405

2. Department of Economics, University of Guelph, Guelph, Ontario N1G 2W1; Department of Economics, Queen's University, Kingston, Ontario K7L 3N6

Abstract

We propose a specification test of a parametrically specified model against a weakly specified alternative. The latter is estimated using k nonparametric nearest neighbors (k-NN) in the context of an artificial regression. We derive the asymptotic distribution under the null hypothesis and under a series of local alternatives. Monte carlo simulations suggest that the test is quite powerful although it has a tendency to over-reject under the null hypothesis.

Key Words: k-nearest neighbor regression, artificial nesting, nonnested hypothesis.

1. Introduction

In this note we propose a specification test of a parametrically specified model, as dictated by economic theory, against a weakly specified alternative. Our test is based on an artificial nesting procedure for testing separate regressions, see Davidson and Mackinnon (1981) and Fisher and McAleer (1981). The performance of the alternative "under the truth" of the null forms the basis for a test of the latter. However, our procedure differs from theirs in that we are only interested in the performance of the null. In other words we do not want to make a statement about the validity of the alternative model, since the latter is not derived from some specific economic model. We also avoid the effect that a possibly misspecified parameterization of the alternative hypothesis would have on testing the validity of the null hypothesis. We use k nonparametric nearest neighbors (k-NN) to estimate the alternative model that enters the artificial regression.

In the next section we discuss the nature of the proposed test and we derive its distribution under the null hypothesis and under a series of local alternatives. We then proceed to investigate its small sample properties by means of a small monte carlo. Finally, we conclude.

2. The Specification Test

Suppose we observe a random sample $\{(Y_i, X_i, Z_i), 1 \le i \le n\}$ from the $R x R^p x R^r$ valued random variable $\{Y, X, Z\}$, having finite variance and conditional distribution $\mathbb{F}_{Y|X,Z}$ that is nondegenerate for all X,Z at which it is defined. Let $\mathbb{E}(.)$ denote the mathematical expectation. The researcher faces the following competing hypotheses:

$$H_0: \mathbb{E}(Y|X,Z) = X^T\beta_0 \ ; \ H_A: \mathbb{E}(Y|X,Z) = \mathbb{E}(Y|Z) \qquad (1.1)$$

In other words H_0 is completely parameterized. The alternative hypothesis H_A is only weakly specified. Hence, the functional form of E(Y|Z) does not take a specific parametric form. We can look at the composite hypothesis H_C as the weighted average of H_0 and H_A:

$$H_C:\ \mathbb{E}(Y|X,Z) = X^T\beta_0(1-\theta) + \theta\mathbb{E}(Y|Z) \tag{1.2}$$

Since $\mathbb{E}(Y|Z)$ is unspecified, we can estimate it by using a nonparametric estimator. In this paper we propose to use (k-NN) nonparametric weights. The use of these weights have been introduced in the semiparametric literature by Robinson (1987). In particular, the estimator of E_i, where $E_i = \mathbb{E}(Y_i|Z_i)$, is given by :

$$\hat{E}_i = \sum_j Y_j\omega_{ij}(k) \tag{1.3}$$

where $\omega_{ij}(k)$ are the weights based on the r Z-regressors. For a positive integer k let $C_i(k)$ be constants satisfying $C_i(k)\geq 0$; $C_i(k)=0$, $i>k$; $\sum_{i=1}^{k} C_i(k)=1$

$$\omega_{ij}(k) = 1(i\neq j)\ r_{ij}^{-1} \sum_{\tau=p_{ij}}^{p_{ij}+r_{ij}-1} C_\tau(k) \tag{1.4}$$

where 1(.) is the indicator function and p_{ij} is 1 plus the number of Z's closer to Z_i than Z_j and r_{ij} is 1 plus the number of Z's equally near from Z_i as Z_j. To calculate distance we use the euclidean metric after we standardize by the sample standard deviation. Then we estimate $(\theta,b^T)^T$, where $b = \beta_0(1-\theta)$, by:

$$\begin{bmatrix} \hat{\theta}_n \\ \hat{b}_n \end{bmatrix} = A_n^{-1}\sum_i \begin{bmatrix} \hat{E}_i \\ X_i \end{bmatrix}, \text{ where } A_n = \sum_i \begin{bmatrix} \hat{E}_i & \hat{E}_i X_i^T \\ X_i\hat{E}_i & X_iX_i^T \end{bmatrix}$$

and $\hat{E}_i$ is given by (1.3).

We will derive the asymptotic distribution of $n^{1/2}\hat{\theta}_n$ under H_0. Hence we will be able to test the validity of H_0 by simply testing the significance of θ. Rejection of this simple hypothesis would imply that H_0

is incorrectly specified. Note that in our context, we are not interested in reversing the order of H_0 and H_A to test one against the other as it is traditionally the case with nonnested tests. In our framework H_A is not parametrically specified and hence it is not of specific interest to the researcher. In the event that one rejects the hypothesis that θ is zero, then the researcher might want to parameterize H_0 differently to account for the possible effect of the Z-variables on Y.

In order to derive the asymptotic distribution of $n^{1/2}\hat{\theta}_n$ under H_0, we need the following regularity conditions :

K.1 $\quad \overline{\lim}_{n\to\infty} \max_i C_i(k) < \infty.$

K.2 $\quad nk^{-\nu/2} \longrightarrow 0,\ kn^{-1} \longrightarrow 0$ as $n \longrightarrow \infty$ for $\nu > 2$.

A.1 $\quad \mathbb{E}(XX^T)$ is positive definite (p.d.).

A.2 $\quad \mathrm{var}(Y|X,Y) = \sigma^2 > 0$ a.s.

A.3 $\quad \mathbb{E}\{\mathrm{var}[\mathbb{E}(Y|Z)|X]\} > 0$ a.s.

A.4 $\quad \mathbb{E}\|X\|^{\nu} < \infty,\ \mathbb{E}|Y - X^T\beta_0|^{2\nu/(\nu-2)} < \infty$, with ν defined in K.2

Conditions K1 and K2 are sufficient for the consistency of k-NN weights, see Stone (1977). Robinson (1987) found that it was technically convenient to relate the rate of convergence of k to the moment conditions imposed on the regressors and residuals. Conditions A1 and A3 ensure that the asymptotic variance of $n^{1/2}(\hat{\theta}_n\ \hat{b}_n)^T$ under H_0 is positive definite. Note that A3 rules out situations where $\mathbb{E}(Y|Z)$ only depends on X or Y is independent of X. The moment conditions in A4 are needed for the asymptotic analysis. It seems difficult to relax them without imposing a stronger rate of convergence in K2. These type of moment conditions were also needed in Newey (1989).

The following theorem is proved in the appendix:

Theorem 1: If K.1-K.2, A.1-A.4 hold then if under H_0: $\mathbb{E}(Y|X,Z) = X^T\beta_0$ a.s. we have that

(i) $$n^{1/2}\hat{\theta}_n \xrightarrow{d} N\left[0,\ \sigma^2/\mathbb{E}\{\mathrm{var}[\mathbb{E}(Y|Z)|X]\}\right].$$

(ii) $$\hat{\sigma}^2\left\{n^{-1}\sum_i\left[\hat{E}_i - X_i^T\hat{\beta}\right]^2\right\}^{-1} \xrightarrow{p} \sigma^2/\mathbb{E}\{\mathrm{var}[\mathbb{E}(Y|Z)|X]\}.$$

where $\hat{\beta}= \left[\sum_i X_iX_i^T\right]^{-1}\sum_i X_iY_i$ and $\hat{\sigma}^2= n^{-1}\sum_i\left[Y_i - X_i^T\hat{\beta}\right]^2$.

The implication of the theorem is that the t-statistic on the significance of θ under H_0 in (1.2) has asymptotically a standard normal distribution. The above statistic is similar in spirit to the Davidson and MacKinnon (1981) J-test, where b and θ are estimated jointly.

Below we will derive the asymptotic distribution of $n^{1/2}\hat{\theta}_n$ under a series of local alternatives. Hence, if the null hypothesis is false, the test statistic will have power to reject it. The proof is given in the appendix.

Theorem 2 If K1,K2, A1-A4 hold and $\mathbb{E}(Y_i|X_i,Z_i)=X_i^T\beta_0(1-n^{-1/2}\theta)+n^{-1/2}\theta\mathbb{E}(Y_i|Z_i)$ a.s. then

$$n^{1/2}\hat{\theta}_n \xrightarrow{d} N\left[\theta,\sigma^2/\mathbb{E}\{\mathrm{var}[\mathbb{E}(Y|Z)|X]\}\right]$$

We can easily accommodate the case where H_0: $\mathbb{E}(Y|X,Z) = f(\beta_0,X)$, where f(.) is assumed to be a known parametric function and the case that Y is assumed to be multivariate. Conditional heteroskedasticity under H_0 does not affect the normality result, but the asymptotic variance of the estimator of θ and b will be different. In this case, one can obtain a consistent estimator of the asymptotic variance, see Eicker (1963) and White (1980).

Under heteroskedasticity of unknown form, an asymptotically efficient estimator of θ under H_0 may be obtained by a semiparametric weighted least squares procedure. One can estimate var$\{Y|X,Z\}$ from the squared residuals computed under the null hypothesis, see Carroll (1982), Robinson (1987) or Newey (1987). A pure nonparametric estimator of var$\{Y|X,Z\}$ may alternatively be obtained from nonparametric estimators of $\mathbb{E}\{Y^2|X,Z\}$ and $\mathbb{E}\{Y|X,Z\}$, see Delgado (1989). The method we propose is adaptive in the sense that the asymptotic variance of $n^{1/2}\hat{\theta}_n$ is the same as the one that could have been obtained had the alternative hypothesis H_A been known or fully parameterized.

3. Monte Carlo Simulations

In this section we will investigate the small sample performance of the proposed test statistic by examining its size and power properties in the context of some monte carlo experiments. We take H_0 to be linear as $H_0: Y = \beta_0+X_1\beta_1+X_2\beta_2+ u_0$. We consider two alternative hypotheses H_{1A} and H_{2A} for two sets of experiments, where $H_{1A}: Y = \gamma_0+(\gamma_1 Z_1+\gamma_2 Z_2)^2+ u_1$ and $H_{2A}: Y = \delta_0+\exp\{\delta_1 Z_1+\delta_2 Z_2\} + u_1$. The parameters $\beta_0,\beta_1,\beta_2,\gamma_0,\gamma_1,\gamma_2,\delta_0,\delta_1,\delta_2$ are set to unity. The X's are generated as NID(0,1) variates and the error terms are generated independently of the regressors as $N(0,\sigma^2)$. By choosing different values of σ we control the fit of the data generating process. For instance, under H_0 which is linear a σ of 0.33 corresponds to a squared correlation coefficient between y and the X's of 0.9483, whereas a σ of 7 corresponds to one of 0.0392. We generate the Z's as $Z_i=\lambda X_i+v_i$, where v_i is distributed as NID(0,1), i=1,2. By varying λ we control the correlation coefficient between the Z_i's and the X_i's. When λ is 1, the correlation coefficient between Z_i and X_i is 0.71, whereas when λ is 0.1, the latter is 0.1. We have used two k-NN estimates of $\mathbb{E}(Y|Z)$, one with $k=n^{1/2}$ and the other with $k=n^{2/3}$. We have chosen sample sizes of n=25,100,1000, which correspond to very small and to moderate sizes for real world cross sections. All the programs were written in FORTRAN double precision and they were run on the VAX of Indiana University. The normal variates were generated by the G05DDF routine of NAG-13. The analysis of sample size n=1000 proved to be quite expensive computationally. We only performed 250 replications in that case, whereas for n=25 and n=100 we performed 10000 and 2500 replications respectively. In both the size and the power experiments we consider also as a benchmark the t-statistic on the significance of θ from the regression $Y = X^T b + \theta \mathbb{E}(Y|Z) + u$, where $\mathbb{E}(Y|Z)$ takes the exact value from H_{1A} and H_{2A} respectively. The above t-statistic will outperform our statistic because it uses exact information that is unavailable to the researcher.

Table 1 presents the results of the size experiments. There is a tendency for our statistic to over-reject, although the proportion decreases as the sample size increases for $\lambda=1$, for the different choices of σ. The test performs quite poorly in the case of $\lambda=0.1$, but this is to be expected since the X's and the Z's are nearly orthogonal. In that case the denominator of the t-statistic approaches zero. The same problem exists in the case, when $\mathbb{E}(Y|Z)$ is completely parameterized as with the J-test of Davidson and MacKinnon (1981). The k-NN estimator with $k=n^{1/2}$ performs better than the one with $k=n^{2/3}$. Tables 2 and 3 present the power results under H_{1A} and H_{2A} respectively. Except for the case of very small samples with a large σ or a small λ, the power results seem quite encouraging. Also as the sample size increases the results improve noticeably. In short the monte carlo results are mixed. The test displays a tendency to over-reject under the null, but it also seems to have considerable power.

4. Conclusions

In this note we have proposed a test statistic that tests a parametric formulation of a null hypothesis against a weakly specified alternative. The procedure we follow resembles in spirit the artificial nesting technique of Davidson and MacKinnon (1981). By means of a monte carlo we investigated the small sample properties of this test and we found them to be satisfactory with respect to power, but less so with respect to size.

The test statistic is derived in the context of iid data and the extension to dependent data is left for future research.

Appendix: Proof of the Theorems

Proof of Theorem 1

Below we will present the lemmas that are used in the proof of the theorem. We define throughout as $\tilde{E}_i = \sum_j \mathbb{E}(Y_j|Z_j)\omega_{ij}$, where $\omega_{ij}=\omega_{ij}(k)$. We need the following preliminary lemmas. Since the proofs to these lemmas constitute only minor modifications to the proofs of lemmas 1,7,8 and 9 in Robinson (1987), they are omitted.

Lemma 1: Let $f(.)$ be a Borel function such that $\mathbb{E}|f(Z)|^p < \infty$, for some $p \geq 1$. Then

$$\mathbb{E}\left\{\sum_j |f(Z_j)-f(Z_i)|^p \omega_{ij}\right\} = o(1).$$

Lemma 2 : For any $p \leq \nu$, $\mathbb{E}|\hat{E}_i - \tilde{E}_i|^p = O(k^{-p/2})$.

Lemma 3 : $\max_i |\hat{E}_i - \tilde{E}_i| = O_p(n^{1/\nu} k^{-1/2})$.

We have that

$$\begin{bmatrix} \hat{\theta}_n \\ \hat{b}_n \end{bmatrix} = A_n^{-1} \sum_i \begin{bmatrix} \hat{E}_i \\ X_i \end{bmatrix} \left\{ \begin{bmatrix} E_i & ,X_i^T \end{bmatrix} \begin{bmatrix} \theta \\ b \end{bmatrix} + \varepsilon_i \right\}$$

$$= A_n^{-1} \sum_i \begin{bmatrix} \hat{E}_i \\ X_i \end{bmatrix} \left\{ \begin{bmatrix} (\hat{E}_i & ,X_i^T)+ ([E_i-\hat{E}_i], & 0) \end{bmatrix} \begin{bmatrix} \theta \\ b \end{bmatrix} +\varepsilon_i \right\}$$

$$= \begin{bmatrix} \theta \\ b \end{bmatrix} + A_n^{-1} \sum_i \begin{bmatrix} \hat{E}_i \\ X_i \end{bmatrix} \left\{ \begin{bmatrix} E_i-\hat{E}_i, 0 \end{bmatrix} \begin{bmatrix} \theta \\ b \end{bmatrix} + \varepsilon_i \right\}$$

where $\varepsilon_i = Y_i - \mathbb{E}(Y_i|X_i,Z_i)$. Under $H_0: \theta=0$,

$$\begin{bmatrix} \hat{\theta}_n \\ \hat{b}_n \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} + A_n^{-1} \sum_i \begin{bmatrix} E_i \\ X_i \end{bmatrix} \varepsilon_i$$

Let

$$V_0 = \mathbb{E}\begin{bmatrix} [\mathbb{E}(Y|Z)]^2 & \mathbb{E}(Y|Z)X^T \\ X\mathbb{E}(Y|Z) & XX^T \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\mathbb{E}(Y|Z)^2] & \beta_0^T\mathbb{E}(XX^T) \\ \mathbb{E}(XX^T)\beta_0 & \mathbb{E}(XX^T) \end{bmatrix}$$

We need to prove that

(a.1) $$n^{1/2}\begin{bmatrix} \hat{\theta}_n \\ \hat{b}_n - b \end{bmatrix} \xrightarrow{d} N\left\{ 0,\ \sigma^2 V_0^{-1} \right\}$$

which implies that

$$n^{1/2}\hat{\theta}_n \xrightarrow{d} N\left[0\ ,\ \sigma^2/\mathbb{E}\{\mathrm{var}[\mathbb{E}(Y|Z)|X]\}\right]$$

Note that under H_0 $\mathbb{E}\{\mathrm{var}[\mathbb{E}(Y|Z)|X]\} = \mathbb{E}[\mathbb{E}(Y|Z)^2] - \beta_0^T\mathbb{E}(XX^T)\beta_0$.

We prove (a.1) from

(a.2) $$n^{-1/2}\sum_i \begin{bmatrix} E_i \\ X_i \end{bmatrix}\varepsilon_i \xrightarrow{d} N(\ 0,\ \sigma^2 V_0)$$

(a.3) $$n^{-1} A_n - V_0 = o_p(1)$$

(a.4) $$n^{-1/2}\sum_i (\hat{E}_i - E_i)\varepsilon_i = o_p(1)$$

By A1-A4, $\begin{bmatrix} E_i \\ X_i \end{bmatrix}\varepsilon_i$ are iid with zero mean and finite variance and (a.2) follows from the Lindenberg-Levy Central Limit Theorem.

We conclude (a.3) from,

(a.5) $$n^{-1}\sum_i \left[E_i^2 - \mathbb{E}(E_i^2)\right] = o_p(1)$$

(a.6) $$n^{-1}\sum_i \left[E_i X_i - \mathbb{E}(E_i X_i)\right] = o_p(1)$$

(a.7) $$n^{-1}\sum_i \left[X_i X_i^T - \mathbb{E}(X_i X_i^T)\right] = o_p(1)$$

(a.8) $$n^{-1}\sum_i \left[\hat{E}_i^2 - \tilde{E}_i^2\right] = o_p(1)$$

(a.9) $$n^{-1}\sum_i \left[\tilde{E}_i^2 - E_i^2\right] = o_p(1)$$

(a.10) $$n^{-1}\sum_i \left[(\hat{E}_i - \tilde{E}_i)X_i\right] = o_p(1)$$

(a.11) $$n^{-1}\sum_i \left[(\tilde{E}_i - E_i)X_i\right] = o_p(1)$$

(a.5) to (a.7) follow from the law of large numbers (LLN). In order to prove (a.8) note that its left side is bounded by

$$\max_i |\hat{E}_i - \tilde{E}_i| \; n^{-1}\sum_i |\hat{E}_i - \tilde{E}_i| = o_p(1),$$

using lemmas 2 and 3 above. Also (a.9) follows from lemma 1. The left side of (a.10) is bounded by

$$\left[n^{-1}\sum_i (\hat{E}_i - \tilde{E}_i)^2\right]^{1/2}\left[n^{-1}\sum_i \|X_i\|^2\right]^{1/2} = O_p(k^{-1/2})$$

using Markov's inequality and lemma 2. The left side of (a.11) is bounded by

$$\left[n^{-1}\sum_i (\tilde{E}_i - E_i)^2\right]^{1/2}\left[n^{-1}\sum_i \|X_i\|^2\right]^{1/2} = o_p(1)$$

using lemma 1. Finally we conclude (a.4) from

(a.12) $$n^{-1/2}\sum_i \left[(\hat{E}_i - \tilde{E}_i)\varepsilon_i\right] = o_p(1)$$

(a.13) $$n^{-1/2}\sum_i \left[(\tilde{E}_i - E_i)\varepsilon_i\right] = o_p(1)$$

(a.12) follows from Chebyschev's inequality, noting that

$$\mathbb{E}\left[n^{-1/2}\sum_i\left[(\hat{E}_i-\tilde{E}_i)\varepsilon_i\right]\right]^2=C_1+C_2$$

where
$$\begin{aligned}C_1&=\mathbb{E}\left\{|\hat{E}_1-\tilde{E}_1|^2\ |\varepsilon_1|^2\right\}\\&\le\left\{\mathbb{E}|\hat{E}_1-\tilde{E}_1|^{\nu}\right\}^{2/\nu}\left\{\mathbb{E}|\varepsilon_1|^{2\nu/(\nu-2)}\right\}^{(\nu-2)/\nu}=O(k^{-1})\end{aligned}$$

by Hölder's inequality and lemma 2, and

$$\begin{aligned}C_2&=\mathbb{E}\left[n^{-1}\sum_{i\neq j}\sum(\hat{E}_i-\tilde{E}_i)(\hat{E}_j-\tilde{E}_j)\varepsilon_i\varepsilon_j\right]\\&\le\left[\mathbb{E}\left[n^{-1}\sum_{i\neq j}\sum(\hat{E}_i-\tilde{E}_i)^2\varepsilon_j^2\right]\right]^{1/2}\left[\mathbb{E}\left[n^{-1}\sum_{i\neq j}\sum(\hat{E}_j-\tilde{E}_j)^2\varepsilon_i^2\right]\right]^{1/2}\end{aligned}$$

Each of these two expectations is by Hölder's inequality bounded by

$$\left[\mathbb{E}|\varepsilon_1|^{2\nu/(\nu-2)}\right]^{(\nu-2)/\nu}\left[n\ \mathbb{E}|\hat{E}_1-\tilde{E}_1|^{\nu}\right]^{2/\nu}=O(n^{2/\nu}k^{-1})$$

using lemma 2. Also (a.13) follows from Chebyschev's inequality, since by lemma 1,

$$\begin{aligned}\mathbb{E}\left[|n^{-1/2}\sum_i(\tilde{E}_i-E_i)\varepsilon_i|^2\right]&=\mathbb{E}\left[|\tilde{E}_1-E_1|^2\varepsilon_1^2\right]\\&=\sigma^2\mathbb{E}|\tilde{E}_1-E_1|^2=o(1).\end{aligned}$$

It is straightforward to show that $\hat{\sigma}^2-\sigma^2=o_p(1)$ under H_0. Hence part (ii) of the theorem follows from

(a.14) $$n^{-1}\sum_i(\hat{E}_i^2-E_i^2)=o_p(1)$$

(a.15) $$n^{-1}\sum_i\left[\hat{\beta}^TX_iX_i^T\hat{\beta}-\beta_0\mathbb{E}(XX^T)\beta_0\right]=o_p(1)$$

(a.16) $$n^{-1}\sum_i \left[\hat{E}_i X_i^T\hat{\beta} - \beta_0 \mathbb{E}(XX^T)\beta_0\right] = o_p(1)$$

(a.14) follows from (a.8) and (a.9) above, whereas (a.15) follows from the LLN noting that $\hat{\beta}$ is $n^{1/2}$-consistent under H_0. Using this fact and (a.10) and (a.11) we obtain (a.16). ■

Proof of Theorem 2

Let $b_0 = \beta_0(1-n^{-1/2}\theta)$. From the proof of Theorem 1,

$$\begin{bmatrix} n^{1/2}\hat{\theta}_n - \theta \\ n^{1/2}(\hat{b}_n - b_0) \end{bmatrix} = [n^{-1}A_n]^{-1} n^{-1/2}\sum_i \begin{bmatrix} \hat{E}_i \\ X_i \end{bmatrix} \left\{ \begin{bmatrix} \hat{E}_i - E_i, 0 \end{bmatrix} \begin{bmatrix} \theta n^{-1/2} \\ b_0 \end{bmatrix} + \varepsilon_i \right\}$$

Given the proof of Theorem 1, for the proof of theorem 2 it suffices to prove that

(b.1) $$n^{-1}\sum_i \hat{E}_i\left[\hat{E}_i - E_i\right] = o_p(1)$$

(b.2) $$n^{-1}\sum_i X_i\left[\hat{E}_i - E_i\right] = o_p(1)$$

(b.1) follows from

(b.3) $$n^{-1}\sum_i \hat{E}_i\left[\hat{E}_i - \tilde{E}_i\right] = o_p(1)$$

(b.4) $$n^{-1}\sum_i \hat{E}_i\left[\tilde{E}_i - E_i\right] = o_p(1)$$

We prove (b.3) noting that its left side is bounded by

$$\max_i |\hat{E}_i - \tilde{E}_i| \; n^{-1}\sum_i \hat{E}_i = o_p(1)$$

by lemma 1,2,3 and the LLN. (b.4) is bounded by

$$\left[n^{-1}\sum_i \hat{E}_i^2\right]^{1/2}\left[n^{-1}\sum_i [\hat{E}_i - E_i]^2\right]^{1/2} = o_p(1)$$

by lemma 1, (a.7), (a.8) and (a.11). The proof of (b.2) is identical to the proof of (b.1).■

References

Carroll, R.J. (1982) "Adapting for Heteroskedasticity in Linear Models" Annals of Statistics, 10, 1224-1233.

Davidson, R. and J.G. MacKinnon (1981) "Several Model Specification Tests in the Presence of Alternative Hypotheses" Econometrica 49, 781-793.

Delgado, M. (1989) "An Alternative Asymptotically Efficient Semiparametric Generalized Least Squares Estimator" manuscript, London School of Economics.

Eicker, F. (1963) "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions" The Annals of Mathematical Statistics 34, 447-456.

Fisher, G.R. and M. McAleer (1981) "Alternative Procedures and Associated Tests of Significance for Non-Nested Hypotheses" Journal of Econometrics, 16, 103-119.

Newey, W.K. (1987) "Efficient Estimation of Models with Conditional Moment Restrictions" manuscript, Princeton University.

Newey, W.K. (1989) "Efficient Instrumental Variable Estimation of Nonlinear Models" Econometrica (forthcoming).

Robinson, P.M. (1987) "Asymptotic efficient estimation in the Presence of Heteroskedasticity of Unknown Form", Econometrica 55, 875-891.

Stone, C.J. (1977) "Consistent Nonparametric Regression" Annals of Statistics, 5, 595-645.

White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity" Econometrica 48, 817-838.

Table 1*

Size Results: Proportion of Rejections when H_0 is true

n =25; Number of Replications = 10000;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.3886	0.2610	0.4861	0.3715	0.5533	0.4406
	0.5376	0.4262	0.6458	0.5476	0.7009	0.6200
	0.1596	0.0631	0.1599	0.0644	0.1805	0.0729
	0.2034	0.0924	0.2112	0.1046	0.2241	0.1083
σ = 2	0.4330	0.3130	0.4626	0.3479	0.4849	0.3661
	0.6369	0.5413	0.6518	0.5568	0.6710	0.5828
	0.1621	0.0660	0.1661	0.0687	0.1858	0.0781
	0.2046	0.0945	0.2142	0.1089	0.2277	0.1157
σ = 7	0.4083	0.2683	0.3778	0.2506	0.3705	0.2410
	0.6330	0.5222	0.5736	0.4600	0.5519	0.4441
	0.1805	0.0797	0.2085	0.1053	0.2379	0.1274
	0.2298	0.1168	0.2689	0.1609	0.2937	0.1783

n =100; Number of Replications = 2500;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.2688	0.1460	0.4652	0.3316	0.6156	0.5056
	0.3968	0.2548	0.5740	0.4572	0.7332	0.6500
	0.0852	0.0244	0.0936	0.0236	0.1020	0.0280
	0.1404	0.0504	0.1496	0.0484	0.1484	0.0552
σ = 2	0.3132	0.1956	0.4088	0.2936	0.4960	0.3716
	0.4660	0.3496	0.5688	0.4680	0.6632	0.5668
	0.0856	0.0248	0.0956	0.0240	0.1036	0.0288
	0.1416	0.0504	0.1540	0.0504	0.1524	0.0564
σ = 7	0.2668	0.1556	0.2860	0.1744	0.2832	0.1708
	0.4464	0.3328	0.4240	0.3100	0.4304	0.3224
	0.0852	0.0240	0.0996	0.0272	0.1140	0.0348
	0.1448	0.0560	0.1708	0.0620	0.1640	0.0700

*The first row corresponds to the number of rejections, under H_0, when $\mathbb{E}(Y|Z)$ is estimated by k-NN, where $k = n^{1/2}$. The second uses $k = n^{2/3}$. The third row corresponds to the benchmark t-statistic, when the added regressor is given by $1+\exp\{Z_1+Z_2\}$ and the fourth when the added regressor is given by $1+(Z_1+Z_2)^2$, i.e when $\mathbb{E}(Y|Z)$ is perfectly known.

Table 1 (continued)

n =1000; Number of Replications = 250;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.2000	0.0720	0.4800	0.3320	0.7200	0.6560
	0.2520	0.1080	0.5240	0.3520	0.8320	0.7960
	0.0520	0.0120	0.0720	0.0280	0.0560	0.0360
	0.1480	0.0600	0.1360	0.0520	0.1560	0.0640
σ = 2	0.2440	0.1120	0.4640	0.3320	0.6800	0.5800
	0.2920	0.1640	0.5160	0.4040	0.7800	0.7160
	0.0520	0.0120	0.0720	0.0280	0.0600	0.0360
	0.1480	0.0560	0.1400	0.0520	0.1600	0.0600
σ = 7	0.2280	0.1200	0.3040	0.2040	0.3720	0.2240
	0.3000	0.2080	0.4120	0.3080	0.5520	0.4240
	0.0520	0.0120	0.0760	0.0280	0.0760	0.0280
	0.1480	0.0560	0.1360	0.0520	0.1560	0.0600

Table 2*

Power Results: Proportion of Rejections when H_{1A} is true

n =25; Number of Replications = 10000;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.9830	0.9645	0.9818	0.9614	0.9809	0.9572
	0.9418	0.9022	0.9422	0.9039	0.9420	0.9013
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 2	0.8885	0.8140	0.7251	0.5948	0.6541	0.5137
	0.8212	0.7490	0.6887	0.5823	0.6359	0.5166
	0.9998	0.9993	0.9952	0.9862	0.9910	0.9716
σ = 7	0.3701	0.2311	0.3143	0.1882	0.3090	0.1826
	0.4772	0.3522	0.4628	0.3381	0.4436	0.3266
	0.8933	0.7747	0.7047	0.5297	0.6514	0.4724

n =100; Number of Replications = 2500;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 2	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 7	0.9428	0.8884	0.5556	0.4120	0.4536	0.2892
	0.9536	0.9112	0.6000	0.4680	0.4972	0.3528
	1.0000	1.0000	0.9908	0.9648	0.9716	0.9068

*The first row corresponds to the number of rejections, under H_{1A}, when $\mathbb{E}(Y|Z)$ is estimated by k-NN, where $k = n^{1/2}$. The second uses $k = n^{2/3}$. The third row corresponds to the benchmark t-statistic, when the added regressor is given by $1+(Z_1+Z_2)^2$, i.e when $\mathbb{E}(Y|Z)$ is perfectly known.

Table 2 (continued)

n =1000; Number of Replications = 250;

	$\lambda = 1$		$\lambda = 0.43$		$\lambda = 0.1$	
	5%	1%	5%	1%	5%	1%
$\sigma = 0.33$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma = 2$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma = 7$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 3*

Power Results: Proportion of Rejections when H_{2A} is true

n =25; Number of Replications = 10000;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.8896	0.8241	0.9857	0.9655	0.9940	0.9844
	0.8439	0.7653	0.9841	0.9603	0.9959	0.9864
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 2	0.8194	0.7194	0.8483	0.7501	0.8430	0.7304
	0.7734	0.6664	0.8615	0.7671	0.8621	0.7584
	0.9990	0.9966	0.9938	0.9850	0.9911	0.9776
σ = 7	0.5189	0.3912	0.4106	0.2649	0.3757	0.2395
	0.5537	0.4310	0.4925	0.3589	0.4735	0.3359
	0.9402	0.8861	0.8288	0.7140	0.7904	0.6553

n =100; Number of Replications = 2500;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.9960	0.9992	1.0000	1.0000	1.0000	0.9996
	0.9980	0.9980	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 2	0.9996	0.9992	1.0000	1.0000	1.0000	0.9996
	0.9980	0.9976	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 7	0.9900	0.9808	0.8920	0.8276	0.8176	0.7116
	0.9932	0.9840	0.9128	0.8684	0.8508	0.7772
	1.0000	1.0000	0.9972	0.9952	0.9964	0.9844

*The first row corresponds to the number of rejections, under H_{2A}, when $\mathbb{E}(Y|Z)$ is estimated by k-NN, where $k = n^{1/2}$. The second uses $k = n^{2/3}$. The third row corresponds to the benchmark t-statistic, when the added regressor is given by $1+\exp\{Z_1+Z_2\}$, i.e $\mathbb{E}(Y|Z)$ is perfectly known.

Table 3 (continued)

n =1000; Number of Replications = 250;

	$\lambda = 1$		$\lambda = 0.43$		$\lambda = 0.1$	
	5%	1%	5%	1%	5%	1%
$\sigma = 0.33$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma = 2$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma = 7$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000