Queen's Economics Department Working Paper No. 783

Semiparametric Specification Testing of Nonlinear Models

Miguel A. Delgado Thanasis Stengos

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2013

Semiparametric Specification Testing
of Nonlinear Models

DISCUSSION PAPER #783

by

Miguel A. Delgado*
Economics Dept., Indiana University
&
Thanasis Stengos
Economics Dept., University of Guelph

*This work has been carried out while the first author was visiting the London School of Economics in the Summer of 1990. Financial support from the Economic and Social Research Council (ESRC), reference number: B00232156 is gratefully acknowledged.

Abstract

We propose a specification test of a parametrically specified nonlinear model against a weakly specified alternative. We generalize a similar test procedure proposed by Delgado and Stengos (1990) to test the specification of a linear model. We estimate the alternative model by using k nonparametric nearest neighbors (k-NN) in the context of an artificial regression. We derive the asymptotic distribution of the test statistic under the null hypothesis and under a series of local alternatives. Monte Carlo simulations suggest that the test has good power and size characteristics.

1. INTRODUCTION

In a recent paper, Delgado and Stengos (1990), henceforth DS, propose a specification test of a linear model against a weakly specified alternative. We generalize their procedure to allow for a nonlinear parametric formulation of the null hypothesis. The proposed test as in the case of the DS test, is based on an artificial nesting procedure for testing separate regressions, see Davidson and Mackinnon (1981) and Fisher and MacAleer (1981). Recently, Wooldridge (1989) proposes a test which also allows for nonparametric alternatives using a different methodology and relying on "sieve" estimators for the alternative model. In this paper as in DS we use k-nonparametric nearest neighbors (k-NN) to estimate the alternative model.

In the next section we discuss the nature of the proposed test. We then proceed to investigate the small sample properties of the test by means of a small Monte Carlo. In the next section we apply the test to an empirical example from labor economics and demonstrate its usefulness with real world data. Finally we conclude.

2.THE SPECIFICATION TEST

Suppose we have independent observations $\{(Y_i,X_i,Z_i),1\le i\le n\}$ from the $\mathbb{R}x\mathbb{R}^r x\mathbb{R}^q$ 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:

$$H_0\text{: } \mathbb{E}(Y|X,Z) = \mathbb{E}(Y|X) = f(\beta_0,X) \text{ ; } H_A\text{:}\mathbb{E}(Y|X,Z) = \mathbb{E}(Y|Z) \qquad (2.1)$$

In other words H_0 is completely parameterized and f: $\mathbb{R}^p$x $\mathbb{R}^r \longrightarrow \mathbb{R}$ is a known function of a vector of parameters and a set of regressors. We focus our attention on the composite hypothesis :

$$H_C\colon\ \mathbb{E}(Y|X,Z) = (1-\delta_0)f(\beta_0,X) + \delta_0\mathbb{E}(Y|Z) \tag{2.2}$$

It is always possible to reparameterize the composite hypothesis as follows:

$$H_c\colon\ \mathbb{E}(Y|X,Z) = g(\theta_0,X) + \delta_0\mathbb{E}(Y|Z) \tag{2.3}$$

where θ_0 is a p'x1 vector of unknown parameters and $(1-\delta_0)f(\beta_0,X) \triangleq g(\theta_0,X)$.

Since $\mathbb{E}(Y|Z)$ is unspecified, we can estimate it by nonparametric regression. In this paper, as in DS, we 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 $E_i = \mathbb{E}(Y_i|Z_i)$, is given by :

$$\hat{E}_i = \sum_j Y_j \omega_{ij}(k) \tag{2.4}$$

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 $C_i(k)\geq 0$; $C_i(k)=0$, $i>k$; $\sum_i^k C_i(k)=1$

$$\omega_{ij}(k) = 1(i\neq j)\, r_{ij}^{-1} \sum_{\tau=p_{ij}}^{p_{ij}+r_{ij}-1} C_\tau(k) \tag{2.5}$$

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.

If we define $\mu=(\delta,\theta^T)^T$ and $\mu_0=(\delta_0,\theta_0^T)^T$, then we proceed to estimate μ by,

$$\hat{\mu}_n = \begin{bmatrix} \hat{\delta}_n \\ \hat{\theta}_n \end{bmatrix} = \underset{M}{\operatorname{argmin}}\ Q_n(\mu,\hat{E})$$

where $\mu \in M$ and $Q_n(\mu,A) = n^{-1}\sum_i\{Y_i - g(\theta,X_i)-\delta A_i\}^2$.

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

A.1 $\mathbb{E}(Y|X,Z) = g(\theta_0,X) + \delta_0\mathbb{E}(Y|Z)$ a.s. where $g\colon \mathbb{R}^{p'} \times \mathbb{R}^{r} \longrightarrow \mathbb{R}$ is uniformly continuous on Θ where θ_0 is an interior point of Θ, a compact subset of $\mathbb{R}^{p'}$. Also δ_0 is an interior point of Δ which is a compact subset of $\mathbb{R}$.

A.2 For all $\varepsilon>0$, there exists a $\xi > 0$ such that

$$\inf_{\|\mu-\mu_0\|\geq \varepsilon} \mathbb{E}\{g(\theta,X) - g(\theta_0,X) + (\delta - \delta_0)\mathbb{E}(Y|Z)\}^2 \geq \xi.$$

A.3 For $\nu > 2$,

$$\text{(i)} \quad \sup_{\beta} |f(\beta,X)| \leq m_0(x) \text{ where } \mathbb{E}|m_0(x)|^{\nu} < \infty,$$

$$\text{(ii)} \quad \mathbb{E}|Y-f(\beta_0,X)|^{\nu} < \infty,$$

$$\text{(iii)} \quad \mathbb{E}|Y-\mathbb{E}(Y|Z)|^{\nu} < \infty.$$

A.4 $\mathrm{Var}(Y|X,Z) = \sigma^2 > 0$ a.s.

Condition A1 describes the model. Note that under A1 $M=\{\Delta,\Theta\}$ is also compact and $(\delta_0,\theta_0^T)^T$ is an interior point of M. Condition A2 is the typical identifiability condition in nonlinear regression (see e.g. Jennrich 1969). Note that condition A2 rules out situations where the same set of regressors is in X and Z. In this case $E(Y|Z)= f(\beta_0, X)$ under the composite hypothesis and condition A2 does not hold for any $\delta\neq \delta_0$ as far as $\beta= \beta_0$. That is δ_0 is not identified under the composite hypothesis. Condition A3 establishes the moment conditions on the conditional expectation of the dependent variable under the null hypothesis and the moment conditions of the disturbances under the null and under the alternative hypotheses respectively. The following two conditions state the rate of convergence of the number of nearest neighbors, which is related to the moment conditions in A3.

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

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

<u>Theorem 1:</u> If K.1-K.2, A.1-A.4 hold, then $\hat{\mu}_n - \mu_0 = o_p(1)$.

Proof.- See appendix.

In order to derive the asymptotic distribution of $n^{1/2}(\hat{\mu}_n - \mu_0)$ we need to introduce the following notation :

$$f_i(\beta)=f(\beta,X_i),\ f_i=f_i(\beta_0),\ \mathring{f}_i(\beta)=\partial f(\beta,X_i)/\partial\beta,\ \mathring{f}_i = \mathring{f}_i(\beta_0),\ g_i(\theta)=g(\theta,X_i),$$

$$g_i=f_i(\theta_0),\ \mathring{g}_i(\theta)=\partial g(\theta,X_i)/\partial\theta,\ \mathring{g}_i=\mathring{g}_i(\theta_0),\ \ddot{g}_i(\theta)=\partial^2 g(\theta,X_i)/\partial\theta\partial\theta^T,\ \ddot{g}_i= \ddot{g}_i(\theta_0),$$

$$\ddot{f}_i(\beta)=\partial^2 f(\beta,X_i)/\partial\beta\partial\beta^T,\ \ddot{f}_i= \ddot{f}_i(\beta_0),\ V_0= \mathbb{E}(\mathring{f}_i\mathring{f}_i^T),$$

$$B_0 = \begin{bmatrix} \mathbb{E}[\mathbb{E}(Y|Z)^2] & \mathbb{E}[\mathring{f}_i^T f_i] \\ \mathbb{E}[\mathring{f}_i f_i] & V_0 \end{bmatrix}$$

We also need to introduce some additional assumptions on the distribution of (Y, X, Z) under H_0,

A.5 There exists a neighborhood of β_0, $\mathbb{N}_0$, such that,

$$\sup_{\mathbb{N}_0}\|\mathring{f}_i(\beta)\| \le m_1(X_i),\ \text{where}\ \mathbb{E}|m_1(X_i)|^{\nu}< \infty,$$

$$\sup_{\mathbb{N}_0}\|\ddot{f}_i(\beta)\| \le m_2(X_i),\ \text{where}\ \mathbb{E}|m_2(X_i)|^{\nu}< \infty.$$

A.6 B_0 is positive definite (p.d).

Condition A5 states the moment conditions for the first and second derivatives of f(.,.). This condition was also assumed by Newey (1989). Condition A5 and A6 ensures that the asymptotic covariance matrix of $n^{1/2}(\hat{\mu}_n-\mu_0)$ is positive definite. Note that when $f_i(\beta)=X_i^T\beta$, a necessary and sufficient condition for A5 to hold is that $\mathbb{E}\{\mathrm{var}[\mathbb{E}(Y|Z)|X]\} > 0$ under H_0, which was assumed by DS. Let define,

$$\hat{B}_n = n^{-1}\sum_i \begin{bmatrix} \hat{E}_i^2 & \hat{E}_i\, \mathring{g}_i(\hat{\theta}_n)^T \\ \hat{E}_i\, \mathring{g}_i(\hat{\theta}_n) & \mathring{g}_i(\hat{\theta}_n)\mathring{g}_i(\hat{\theta}_n)^T \end{bmatrix}$$

$$\hat{\sigma}^2 = (1/n-p) \sum_i\{ Y_i- g(\hat{\theta}_n,X_i) - \hat{\delta}_n\hat{E}_i\}^2.$$

We are now in position to derive the asymptotic distribution of $n^{1/2}(\hat{\mu}_n-\mu_0)$

under H_0.

Theorem 2 : If K1, K2, A1-A6 hold and H_0: $\delta_0 = 0$,

$$n^{1/2}\begin{bmatrix} \hat{\delta}_n \\ \hat{\theta}_n - \theta_0 \end{bmatrix} \xrightarrow{d} N(0, \sigma^2 B_0^{-1}),$$

$$\hat{\sigma}^2 \hat{B}_n - \sigma^2 B_0 = o_p(1).$$

Proof.- See appendix.

Note that the above result implies that under H_0,

$$n^{1/2}\hat{\delta}_n \xrightarrow{d} N(0, \sigma^2 / \Big[\mathbb{E}\{var[\mathbb{E}(Y|Z)|X]\} + \{\mathbb{E}[f_i^2] - \mathbb{E}[f_i \mathring{f}_i^T]\ V_0^{-1}\mathbb{E}[\mathring{f}_i f_i]\}\Big].$$

The first diagonal component of $\hat{\sigma}^2\hat{B}_n^{-1}$ is the estimate of the asymptotic covariance matrix of $n^{1/2}\hat{\delta}_n$ under H_0. Note that once the conditional expectation estimates $\hat{E}_i$ have been computed, one may use standard software to compute the t-ratio to test H_0:δ=0.

Under conditional heteroskedasticity of unknown form, the covariance matrix is different, but valid t-ratios can be obtained by estimating the asymptotic covariance matrix of $n^{1/2}(\hat{\mu} - \mu_0)$ under the null by the Eicker-White (Eicker 1963 and White 1982) estimate. Let assume that $Var(Y|X,Z) = \sigma^2(X,Z)$ is a function of unknown form. Let define,

$$C_0 = B_0^{-1}\ D_0\ B_0^{-1}$$
$$\hat{C}_n = \hat{B}_n^{-1}\ \hat{D}_n\ \hat{B}_n^{-1}$$

where,

$$D_0 = \begin{bmatrix} \mathbb{E}[\mathbb{E}(Y|Z)^2\sigma^2(X,Z)] & \mathbb{E}[\mathring{f}_i^T f_i \sigma^2(X,Z)] \\ \mathbb{E}[\mathring{f}_i f_i \sigma^2(X,Z)] & \mathbb{E}[\mathring{f}_i \mathring{f}_i^T \sigma^2(X,Z)] \end{bmatrix}$$

$$\hat{D}_n = n^{-1}\sum_i \begin{bmatrix} \hat{E}_i^2 e_i^2 & \hat{E}_i\ \mathring{g}_i(\hat{\theta}_n)^T e_i^2 \\ \hat{E}_i \mathring{g}_i(\hat{\theta}_n) e_i^2 & \mathring{g}_i(\hat{\theta}_n)\mathring{g}_i(\hat{\theta}_n)^T e_i^2 \end{bmatrix}$$

where $e_i = Y_i - g_i(\hat{\theta}_n) - \hat{\delta}_n\ \hat{E}_i$.

Theorem 3: If K1, K2, A1-A3,A5,A6 hold, D_0 nonsingular and when H_0: $\delta_0 = 0$,

$$n^{1/2}\begin{bmatrix} \hat{\delta}_n \\ \hat{\theta}_n - \theta_0 \end{bmatrix} \xrightarrow{d} N(0, \sigma^2 C_0),$$

$$\hat{C}_n - C_0 = o_p(1).$$

Proof.- See appendix.

Given Theorem 1 and 2 it is straightforward to deduce the asymptotic distribution of $n^{1/2}\hat{\delta}_n$ under a series of local alternatives.

Theorem 3: If K1, K2, A1-A6 holds, then when $\delta_0 = n^{1/2}\ \gamma_0$, where γ_0 is a constant,

$$n^{1/2}\hat{\delta}_n \xrightarrow{d} N(\delta_0,\sigma^2/\Big[\mathbb{E}\{var[\mathbb{E}(Y|Z)|X]\} + \{\mathbb{E}[f_i^2] - \mathbb{E}[f_i\mathring{f}_i^T]\ V_0^{-1}\mathbb{E}[\mathring{f}_i f_i]\}\Big].$$

Proof.- See appendix.

Hence, if the null hypothesis is false, the test statistic will have power to reject it. A similar result was also obtained by DS. Note that, in the linear case, $\{\mathbb{E}[f_i^2] - \mathbb{E}[f_i\mathring{f}_i^T]\ V_0^{-1}\mathbb{E}[\mathring{f}_i f_i]\} = 0$. Under conditional heteroskedasticity of unknown form, the asymptotic variance of $n^{1/2}\hat{\delta}_n$ is different. In this case Theorem 3 follows but the asymptotic variance is the first diagonal component of C_0 which can be consistently estimated by the first diagonal component of $\hat{C}_n$. Once the conditional expectation under the alternative is estimated using k-nn, then one can use any package that allows for nonlinear least squares estimation to carry out the test.

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. The structure of the experiments is similar to DS for comparison purposes. 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=25,100,500, 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 IBM 3081 of the University of Guelph. The normal variates were generated by the GGNQF routine of IMSL. For n=25 we performed 10000 and for n=100 and 500 we performed 1000 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, as σ increases. This is to be expected, since large values of σ represent a lot of noise and a poor fit for the data generating process. As the sample size increases the actual size tends to its asymptotic value, and the difference between the proposed test and the benchmark unfeasible one becomes smaller. 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 suggest that the proposed test performs adequately with respect to its size and power characteristics.

3. EMPIRICAL EXAMPLE

In this empirical example we estimate a wage equation from a sample of Canadian microdata from the 1981 Survey of Work History conducted by Statistics Canada in January of 1982 in conjunction with the monthly Labor Force Survey. The sample consists of 1541 unionized male workers in Ontario from 20 to 54 years of age, employed in the non agricultural sectors throughout the year 1981 with a single employer. The definitions and descriptive statistics of all the variables that enter the analysis are given in table 4.

To simplify the estimation in most wage studies one assumes a semi-logarithmic specification for the wage equation in question. The implicit assumption in this case is that the errors enter multiplicatively in the original exponential formulation. In the present study we will estimate the wage equation for the unionized workers in our sample in its original exponential form, assuming implicitly the presence of an additive error structure.

We will distinguish between a null hypothesis that takes the process of unionization as exogenous an a model that assumes an endogenous selection process that assigns workers to the union sector. The latter is parametrically unspecified and it includes a certain distinguishing feature from the null hypothesis model.

In the literature the effect of endogenous unionization on the wage rate is typically modelled in the context of an endogenous switching model whereby one postulates a sample selection criterion function that assigns workers into the union and nonunion sectors. Furthermore, one assumes joint normality of the errors of the wage equations in their semi-logarithmic form and the error of the sample selection rule. That leads to the inclusion of the inverse Mills ratio variables as additional variables in the original wage equations in order to correct for selectivity bias. The selectivity correction variables are themselves functions of the parameters that enter the sample selection rule. The model is then estimated by Maximum Likelihood or by the simpler two stage procedure suggested by Heckman (1976) and Lee (1978).

In the present application we will not assume a specific distribution for the errors. We will only assume that there is a selection process that assigns workers into the union sector and consequently affects the wage rate. The null hypothesis takes the form

$$H_0 : E(Y_i|X_i,Z_i) = \exp\{X_i^T\beta_0\} \tag{3.1}$$

whereas the alternative is left in its conditional expectation form as

$$H_1 : E(Y_i|X_i,Z_i) = E(Y_i|Z_i) \tag{3.2}$$

The composite model is constructed as

$$H_c : E(Y_i|X_i,Z_i) = (1-\delta_0)\exp\{X_i^T\beta_0\} + \delta_0 E(Y_i|Z_i) \tag{3.4}$$

The X's and the Z's overlap as they both include the typical demographic, human capital and job specific variables that enter wage equations. The variable that distinguishes the Z's from the X's is the degree of unionization in the industry as measured by the percentage of unionized workers in the industry that the worker in question is employed. The estimation results of the composite model are presented in table 5. The estimated parameters all have the expected signs and they are statistically significant. The standard errors are computed using the Eicker/White heteroskedasticity robust variance-covariance matrix. Skill as measured by a weighted average of the occupational skill requirements of the industry and years of tenure with the firm have a positive impact on the wage rate, although job tenure has a declining marginal effect as seen by the negative sign of the squared tenure variable. Also older workers earn more than younger ones and public sector employees earn more than their private sector counterparts. The number of other wage earners in the industry tends to increase the opportunities available for alternative employment for an individual worker and consequently has a positive effect on wages. The Z's include all of the variables that enter the null model except for the number of wage earners in the industry variable. In addition they include the degree of unionization variable. All the computations, except for $E(Y|Z)$ using k-NN, were carried out in TSP.

The results suggest that the null model should be rejected. One course of action is to augment the null hypothesis to include the degree of unionization variable that suggests the presence of an endogenous sample selection rule. Alternatively, one can simply postulate a particular parametric form for the errors and proceed to estimate the wage equation by including the corresponding sample selectivity correction variable. The latter course will be computationally demanding given the nonlinear structure of the model and the nonlinear nature of the selectivity correction variable itself.

4. CONCLUSIONS

In this paper we have proposed a test statistic that tests a nonlinear parametric formulation of a null hypothesis against a weakly specified alternative. It constitutes a generalization of the test proposed by DS. 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 and size. We also applied the test to an example from labor economics where we analyzed the specification of a wage equation for unionized workers with Canadian data. 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

Henceforth, $\tilde{E}_i = \sum_j E_j \omega_{ij}$ where $\omega_{ij} = \omega_{ij}(k)$. We need the following lemmas. Lemma 1 is Stone (1977) Proposition 1 and lemmas 1 and 2 are minor modifications of Robinson (1987) lemmas, 7, 8 and 9.

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

$$\mathbb{E}\left\{\sum_j |f(Z_j) - f(Z_1)|^p \omega_{1j}\right\} = o_p(1)$$

Lemma 2. For any $p \leq \upsilon$, $\mathbb{E}|\hat{E}_1 - \tilde{E}_1|^p = O(k^{-p/2})$.

Lemma 3.- $\max_i |\hat{E}_1 - \tilde{E}_1| = o_p(n^{1/\upsilon} k^{-1/2})$.

Proof of Theorem 1

Using A3 and standard consistency results,(see Amemiya 1985 Chapter 4) and the machinery adopted in DS, the consistency of $\hat{\mu}_n$ follows from

$$\sup_M |Q_n(\mu,\hat{E}) - \mathbb{E}\{g(\theta,X) - g(\theta_0,X) + (\delta - \delta_0)\mathbb{E}(Y|Z)\}^2 - \sigma^2| = o_p(1) \qquad (a.1)$$

Let us define $h_i(A,\mu) = \left[g(\theta,X_i) - g(\theta_0,X_i) + (\delta - \delta_0)A_i\right]$ and $u_i(A) = Y_i - g(\theta_0,X_i) - \delta_0 A_i$. Then we can write

$$Q_n(\mu,\hat{E}) = n^{-1}\textstyle\sum_i\{u_i(\hat{E})^2 + h_i(\hat{E},\mu)^2 - 2u_i(\hat{E}_i)h_i(\hat{E},\mu)\}.$$

(a.1) follows from

$$n^{-1}\textstyle\sum_i u_i(\hat{E})^2 - \sigma^2 = o_p(1), \qquad (a.2)$$

$$\sup_M |n^{-1}\textstyle\sum_i h_i(\hat{E},\mu)^2 - \mathbb{E}[h_i(E,\mu)^2]| = o_p(1), \qquad (a.3)$$

$$\sup_M |n^{-1}\textstyle\sum_i [h_i(\hat{E},\mu)\ u_i(\hat{E})| = o_p(1). \qquad (a.4)$$

We conclude (a.2) from,

$$n^{-1}\textstyle\sum_i u_i(E)^2 - \sigma^2 = o_p(1), \qquad (a.5)$$

$$n^{-1}\textstyle\sum_i (Y_i - g(\theta_0,X_i)(\hat{E}_i - E_i) = o_p(1) \qquad (a.6)$$

$$n^{-1}\textstyle\sum_i (\hat{E}_i^2 - E_i^2) = o_p(1) \qquad (a.7)$$

Noting that,

$$\mathbb{E}|u_i(E)|^{\nu} \leq K\{ \mathbb{E}|Y_i - f(\beta_0,X_i)|^{\nu} + \mathbb{E}|Y_i - \mathbb{E}(Y_i|Z_i)|^{\nu} \} < \infty \tag{a.8}$$

by A3, (where henceforth K is a generic constant), (a.5) follows by the Law of Large Numbers (LLN). Now note that,

$$\mathbb{E}|Y - (1-\delta_0)f(\beta_0,X)|^{\nu} \leq \mathbb{E}|Y - f(\beta_0,X)|^{\nu} + K\mathbb{E}|m_0(x)|^{\nu} < \infty \tag{a.9}$$

$$\mathbb{E}|Y|^{\nu} \leq \mathbb{E}|Y - f(\beta_0,X)|^{\nu} + K\mathbb{E}|m_0(x)|^{\nu} < \infty. \tag{a.10}$$

by A3. (a.6) follows, after applying Cauchy's inequality and A3, from,

$$n^{-1}\sum_i(\hat{E}_i - E_i)^2 \leq n^{-1}\sum_i(\hat{E}_i - \tilde{E}_i)^2 + n^{-1}\sum_i(\tilde{E}_i - E_i)^2 = o_p(1) \tag{a.11}$$

by Lemma 1 and 2. (a.7) follows from,

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

by (a.11) and Lemma 1. We obtain (a.3) from (a.11) and,

$$\sup_M|n^{-1}\sum_i h_i(E,\mu)^2 - \mathbb{E}[h_i(E,\mu)^2]| = o_p(1), \tag{a.12}$$

$$\sup_M |n^{-1}\sum_i[g(\theta,X_i) - g(\theta_0,X_i)](\delta - \delta_0)(\hat{E}_i - E_i)| = o_p(1). \tag{a.13}$$

(a.12) follows from the uniform LLN; (a.13) is bounded by

$$\left[n^{-1}\sum_i m_0(X_i)^2\right]^{1/2}\left[n^{-1}\sum_i(\hat{E}_i - E_i)^2\right]^{1/2} = o_p(1),$$

by A3 and (a.11). Finally (a.4) follows from,

$$\sup_M|n^{-1}\sum_i[h_i(\hat{E},\mu)\ u_i(E)| = o_p(1), \tag{a.14}$$

$$\sup_M|n^{-1}\sum_i[h_i(\hat{E},\mu)(\delta - \delta_0)(\hat{E}_i - E_i)| = o_p(1). \tag{a.15}$$

(a.14) follows from,

$$\sup_M |n^{-1}\sum_i[g(\theta,X_i) - g(\theta_0,X_i)]u_i(E)| = o_p(1), \tag{a.16}$$

$$|n^{-1}\sum_i(\hat{E}_i - E_i)u_i(E)| = o_p(1), \tag{a.17}$$

$$|n^{-1}\sum_i E_i u_i(E)| = o_p(1). \tag{a.18}$$

(a.16) follows from the LLN, (a.17) follows from (a.11). Also (a.18) follows from the LLN. Finally (a.15) is bounded by

$$\left[n^{-1}\sum_i\{m_0(X_i)^2 + \hat{E}_i^2\}\right]^{1/2}\left[n^{-1}\sum_i(\hat{E}_i - E_i)^2\right]^{1/2} = o_p(1)$$

by A.3, (a.7), (a.11) and Lemma 1.

Let introduce some additional notation:

$$u_i(\mu,A)= Y_i - g(\theta,X_i) - \delta A_i \text{ and } u_i = Y_i - f(\beta_0,X_i) - \delta_0 E_i$$

Proof of theorem 2

In view of theorem 1 and using a standard mean value theorem (mvt) argument, it suffices to prove that,

$$n^{-1/2}\sum_i\begin{bmatrix}\hat{E}_i u_i\\ \mathring{f}_i u_i\end{bmatrix}\left\{n^{-1}\sum_i\begin{bmatrix}\hat{E}_i^2 & \hat{E}_i\, g_i(\bar{\theta}_n)\\ \hat{E}_i\mathring{g}_i(\bar{\theta}_n) & \mathring{g}_i(\bar{\theta}_n)\mathring{g}_i(\bar{\theta}_n)^T + \ddot{g}_i(\bar{\theta}_n)u_i(\bar{\mu}_n,\hat{E})\end{bmatrix}\right\}$$

$$\xrightarrow{d} N(0,\sigma^2B_0^{-1})$$

for any $\bar{\mu}_n = (\bar{\delta}_n, \bar{\theta}_n^T)^T$ such that $\bar{\mu}_n - (0, \theta_0^T)^T = o_p(1)$. Then, it suffices to prove that

$$n^{-1/2}\sum_i\begin{bmatrix}E_iu_i\\ \mathring{f}_iu_i\end{bmatrix} \xrightarrow{d} N(0,\sigma^2B_0) \tag{b.1}$$

$$n^{-1/2}\sum_i(\hat{E}_i - E_i)\, u_i = o_p(1) \tag{b.2}$$

$$n^{-1}\sum_i(\hat{E}_i^2 - E_i^2) = o_p(1) \tag{b.3}$$

$$n^{-1}\sum_i(\hat{E}_i\mathring{g}_i(\bar{\theta}_n) - E_i\mathring{g}_i) = o_p(1) \tag{b.4}$$

$$n^{-1}\sum_i[\mathring{g}_i(\bar{\theta}_n)\mathring{g}_i(\bar{\theta}_n)^T - \mathring{g}_i\mathring{g}_i^T] = o_p(1) \tag{b.5}$$

$$n^{-1}\sum_i[\ddot{g}_i(\bar{\theta}_n)u_i(\bar{\mu}_n,\hat{E}) - \ddot{g}_iu_i] = o_p(1) \tag{b.6}$$

$$n^{-1}\sum_i\begin{bmatrix}E_i^2 & E_i\mathring{g}_i^T\\ E_i\mathring{g}_i & \mathring{g}_i\mathring{g}_i + \ddot{g}_iu_i\end{bmatrix} - B_0 = o_p(1) \tag{b.7}$$

(b.1) follows from the Lindenber-Levy Central Limit Theorem. Noting that

$$\mathbb{E}|n^{-1/2}\sum_i(\hat{E}_i - E_i)\, u_i|^2 = C_1 + C_2 + C_3$$

where

$$C_1 = \mathbb{E}\left\{|\hat{E}_1 - \tilde{E}_1|^2 \ |U_1|^2\right\},\ C_2 = \mathbb{E}\left\{n^{-1}\sum_{i \neq j}\sum (\hat{E}_i - \tilde{E}_i)\ (\hat{E}_j - \tilde{E}_j)\ U_i\ U_j\right\} \text{ and}$$

$$C_3 = \mathbb{E}|n^{-1/2} \textstyle\sum_i (\tilde{E}_i - E_i)\ U_i|^2,$$

Then, (b.2) follows from Markov's inequality since, by Hölder's inequality,

$$C_1 \leq \left\{\mathbb{E}|\hat{E}_1 - \tilde{E}_1|^{\upsilon}\right\}^{2/\upsilon} \left\{\mathbb{E}|U_1|^{2\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon} = O(k^{-1})$$

by Lemma 2; by Cauchy and Hölder's inequalities,

$$C_2 \leq \mathbb{E}\left\{n^{-1} \sum_{i \neq j}\sum (\hat{E}_i - \tilde{E}_i)^2 U_j^2\right\} \leq \left\{n\ \mathbb{E}|\hat{E}_1 - \tilde{E}_1|^{\upsilon}\right\}^{2/\upsilon} \left\{\mathbb{E}|U_1|^{2\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon}$$

$$= O(n^{2/\upsilon}k^{-1})$$

by Lemma 2; by Cauchy's inequality,

$$C_2 \leq \left\{\mathbb{E}|\tilde{E}_1 - E_1|^{\upsilon}\right\}^{2/\upsilon} \left\{\mathbb{E}|U_1|^{2\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon} = o_p(1)$$

by Lemma 1. (b.3) follows from (a.7), (b.4) from

$$n^{-1}\textstyle\sum_i E_i(\mathring{g}_i(\bar{\theta}_n) - \mathring{g}_i) = o_p(1) \tag{b.8}$$

$$n^{-1}\textstyle\sum_i \mathring{g}_i(\bar{\theta}_n)[\hat{E}_i - E_i] = o_p(1) \tag{b.9}$$

(b.8) and (b.9) follow from Serfling (1980), Lemma 7.2.1.A (henceforth S), A4 and (a.11). Further, (b.5) follows from S and (b.6) follows from,

$$n^{-1}\textstyle\sum_i(\ddot{g}_i(\bar{\theta}_n) - \ddot{g}_i)u_i = o_p(1) \tag{b.10}$$

$$n^{-1}\textstyle\sum_i \ddot{g}_i(\bar{\theta}_n)[u_i(\bar{\mu},\hat{E}) - u_i] = o_p(1) \tag{b.11}$$

(b.10) follows from S and (b.11) follows from

$$n^{-1}\textstyle\sum_i \ddot{g}_i(\bar{\theta}_n)[g_i(\bar{\theta}_n) - g_i] = o_p(1) \tag{b.12}$$

$$n^{-1}\textstyle\sum_i \ddot{g}_i(\bar{\theta}_n)\bar{\delta}_n\hat{E}_i = o_p(1) \tag{b.13}$$

(b.12) and (b.13) follow from S and, the consistency of $\bar{\mu}_n$, Lemma 1 and 2.

Proof of Theorem 3.

The consistency of $\hat{\mu}_n$ follows form,

$$\sup_M |Q_n(\mu,\hat{E}) - \left\{\mathbb{E}[g(\theta,X) - g(\theta_0,X) + (\delta - \delta_0)\mathbb{E}(Y|Z)]^2 - \sigma^2(X,Z)\right\}| = o_p(1)$$

it follows form (a.3), (a.4) and

$$n^{-1}\sum_i\{u_i(\hat{E})^2 - \sigma^2(X,Z)\} = o_p(1),$$

which follows using the same arguments as in the proof to (a.2). Then the Theorem follows from,

$$n^{-1/2}\sum_i \begin{bmatrix} E_i u_i \\ \mathring{f}_i u_i \end{bmatrix} \xrightarrow{d} N(0,D_0)$$

and (b.2)-(b.7). In order to prove the consistency of $\hat{C}_n$, it is not necessary to assume that $\delta_0 = 0$. In fact, we prove that $\hat{C}_n - C_0 = o_p(1)$ for any δ_0. Let define follows from (b.2)-(b.7) and, $\hat{D}_n - D_0 = o_p(1)$ which follows from,

$$n^{-1}\sum_i \hat{E}_i^2 e_i^2 - \mathbb{E}(E_1\sigma_1^2) = o_p(1), \tag{d.1}$$

$$n^{-1}\sum_i \hat{E}_i \mathring{g}(\bar{\theta}_n) e_i^2 - \mathbb{E}(E_1\mathring{f}_1\sigma_1^2) = o_p(1), \tag{d.2}$$

$$n^{-1}\sum_i \mathring{g}(\bar{\theta}_n)\mathring{g}(\bar{\theta}_n)^T e_i^2 - E(\mathring{f}_1\mathring{f}_1^T\sigma_1^2) = o_p(1). \tag{d.3}$$

(d.1) follows from,

$$n^{-1}\sum_i E_i^2\sigma_i^2 - \mathbb{E}(E_1\sigma_1^2) = o_p(1), \tag{d.4}$$

$$n^{-1}\sum_i E_i^2(u_i^2 - \sigma_i^2) = o_p(1), \tag{d.5}$$

$$n^{-1}\sum_i E_i^2\,(e_i^2 - u_i^2) = o_p(1), \tag{d.6}$$

$$n^{-1}\sum_i (\hat{E}_i^2 - E_i^2)\, e_i^2 = o_p(1). \tag{d.7}$$

(d.4) and (a.5) follow form the LLN. In order to prove (d.6), note that,

$$u_i(\hat{E})^2 + h_i(\hat{E},\hat{\mu}_n)^2 - 2u_i(\hat{E}_i)h_i(\hat{E},\hat{\mu}_n),$$

then, (d.6) follows form,

$$n^{-1}\sum_i E_i^2\,(u_i(\hat{E})^2 - u_i^2) = o_p(1), \tag{d.8}$$

$$n^{-1}\sum_i E_i^2\, h_i(\hat{E},\hat{\mu}_n)^2 = o_p(1), \tag{d.9}$$

$$n^{-1}\sum_i E_i^2\, h_i(\hat{E},\hat{\mu}_n)\; u_i(\hat{E}) = o_p(1). \tag{d.10}$$

The left side of (d.8) is equal to,

$$\delta_0\, n^{-1}\sum_i [\hat{E}_i - E_i]^2 - 2\,\delta_0\, n^{-1}\sum_i [\hat{E}_i - E_i]\, E_i^2\, u_i(\hat{E})$$

$$\le \left\{n^{-1}\sum_i |\hat{E}_i^2 - E_i^2|^{\upsilon/2}\right\}^{2/\upsilon} \left\{n^{-1}\sum_i |E_i|^{2\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon}$$

$$\times \left\{n^{-1}\sum_i |\hat{E}_i^2 - E_i^2|^{\upsilon}\right\}^{1/\upsilon} \left\{n^{-1}\sum_i |u_i(\hat{E})|^{\upsilon}\right\}^{1/\upsilon} \left\{n^{-1}\sum_i |E_i|^{2\upsilon/(\upsilon-2)}\right\}^{(\upsilon-2)/\upsilon}$$

$= o_p(1)$,

by Hölder and Cauchy's inequalities, Lemma 1 and 2; and (d.10) and (d.11) follow by S and Lemma 1 and 2. The proof of (d.2) and (d.2) uses very similar arguments to those employed in the proof of (d.1).

Proof of Theorem 4:

Using the same arguments as in Theorem 1 $\hat{\delta}_n = o_p(1)$. Then using a mean value theorem argument, the theorem follows by similar arguments as in Theorem 2.

References

Amemiya, T. (1985) Advanced Econometrics Harvard University Press, Cambridge, Massachusetts.

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

Delgado, M. and T. Stengos (1990) "Semiparametric Specification Testing" Preprint.

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.

Heckman, J.J. (1976) "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models" Annals of Economic and Social Measurement 5, 475-492.

Jennrich, R.I. (1969) "Asymptotic Properties of Non-Linear Least Squares Estimators" The Annals of Mathematical Statistics 40, 633-643.

Lee, L.F. (1978) "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables" International Economic Review 19, 415-433.

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.

Serfling R.J. (1980) Approximation Theorems of Mathematical Statistics John Wiley, New York.

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

Wooldridge, J.M. (1989) "Some Results on Specification Testing Against Nonparametric Alternatives" . Preprint.

Table 1*

Size Results: Number 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	0.0932	0.0314	0.1015	0.0350	0.1105	0.0410
	0.1076	0.0246	0.1061	0.0344	0.2728	0.1534
	0.0770	0.0240	0.0802	0.0239	0.0807	0.0235
	0.0803	0.0246	0.0802	0.0241	0.0794	0.0242
σ = 2	0.2361	0.1261	0.2057	0.1039	0.1993	0.1003
	0.3186	0.1932	0.2807	0.1610	0.2728	0.1534
	0.0771	0.0242	0.0797	0.0239	0.0808	0.0235
	0.0808	0.0247	0.0800	0.0241	0.0797	0.0242
σ = 7	0.3096	0.1861	0.2543	0.1380	0.2289	0.1248
	0.4427	0.3018	0.3468	0.2185	0.3207	0.1950
	0.0780	0.0244	0.0799	0.0240	0.0810	0.0236
	0.0806	0.0247	0.0799	0.0241	0.0795	0.0242

n =100; Number of Replications = 1000;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.0640	0.0140	0.0700	0.0180	0.0750	0.0140
	0.0640	0.0120	0.0670	0.0210	0.0670	0.0180
	0.0570	0.0100	0.0580	0.0160	0.0650	0.0150
	0.0570	0.0060	0.0570	0.0150	0.0580	0.0170
σ = 2	0.1370	0.0550	0.1410	0.0450	0.1540	0.0700
	0.1600	0.0630	0.1720	0.0850	0.1730	0.0760
	0.0570	0.0100	0.0580	0.0160	0.0650	0.0150
	0.0570	0.0060	0.0570	0.0150	0.0580	0.0170
σ = 7	0.1820	0.0960	0.1650	0.0870	0.1870	0.0860
	0.3040	0.1610	0.2190	0.1180	0.2190	0.1150
	0.0570	0.0120	0.0580	0.0160	0.0580	0.0150
	0.0570	0.0600	0.0570	0.0150	0.0580	0.0170

*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 =500; Number of Replications = 1000;

	λ = 1		λ = 0.43		λ = 0.1	
	5%	1%	5%	1%	5%	1%
σ = 0.33	0.0580	0.0100	0.0610	0.0120	0.0620	0.0140
	0.0570	0.0120	0.0620	0.0130	0.0670	0.0180
	0.0510	0.0100	0.0520	0.0160	0.0650	0.0120
	0.0520	0.0100	0.0580	0.0150	0.0580	0.0170
σ = 2	0.0710	0.0210	0.0710	0.0250	0.0720	0.0230
	0.0810	0.0310	0.0730	0.0210	0.0730	0.0210
	0.0510	0.0100	0.0510	0.0100	0.0600	0.0150
	0.0510	0.0120	0.0520	0.0100	0.0580	0.0170
σ = 7	0.0910	0.0300	0.0930	0.0450	0.0900	0.0390
	0.9300	0.0310	0.0900	0.0450	0.0970	0.0420
	0.0510	0.0100	0.0580	0.0160	0.0600	0.0150
	0.0530	0.0120	0.0570	0.0150	0.0610	0.0170

Table 2*

Power Results: Number 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	0.9495	0.8899	0.9447	0.8850	0.9394	0.8787
	0.7824	0.6515	0.7975	0.6666	0.8028	0.6760
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 2	0.7908	0.6711	0.5936	0.4508	0.5252	0.3800
	0.5892	0.4333	0.4295	0.2792	0.3770	0.2361
	0.9999	0.9997	0.9942	0.9850	0.9862	0.9694
σ = 7	0.2457	0.1288	0.2066	0.1036	0.2038	0.1056
	0.2225	0.1119	0.2441	0.1349	0.2477	0.1356
	0.8570	0.7428	0.5808	0.3963	0.4914	0.3130

n =100; Number of Replications = 1000;

	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	0.9990
	1.0000	1.0000	1.0000	1.0000	0.9980	0.9980
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
σ = 7	0.9160	0.8590	0.5150	0.3430	0.4020	0.2510
	0.8950	0.8230	0.4710	0.3430	0.3480	0.2060
	1.0000	1.0000	0.9790	0.9370	0.9490	0.8640

*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: Number of Rejections when H_{2A} is true

n =25; Number of Replications = 10000;

	$\lambda = 1$		$\lambda = 0.43$		$\lambda = 0.1$	
	5%	1%	5%	1%	5%	1%
$\sigma = 0.33$	0.7365	0.6233	0.9493	0.8957	0.9850	0.9553
	0.5769	0.4340	0.9027	0.8158	0.9637	0.9055
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma = 2$	0.6381	0.5036	0.7583	0.6297	0.7724	0.6470
	0.4715	0.3293	0.6925	0.5372	0.7258	0.5814
	0.9986	0.9956	0.9930	0.9828	0.9887	0.9755
$\sigma = 7$	0.3332	0.2082	0.2831	0.1645	0.2720	0.1539
	0.2524	0.1340	0.2759	0.1556	0.2796	0.1607
	0.9127	0.8635	0.7398	0.6283	0.6748	0.5484

n =100; Number of Replications = 1000;

	$\lambda = 1$		$\lambda = 0.43$		$\lambda = 0.1$	
	5%	1%	5%	1%	5%	1%
$\sigma = 0.33$	0.9890	0.9750	0.9990	0.9990	1.0000	1.0000
	0.9890	0.9550	1.0000	0.9980	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma = 2$	0.9840	0.9760	0.9990	0.9990	1.0000	0.9990
	0.9690	0.9520	0.9990	0.9980	1.0000	1.0000
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\sigma = 7$	0.9710	0.9480	0.8710	0.7840	0.7910	0.6910
	0.9510	0.9140	0.8640	0.7810	0.7950	0.6840
	1.0000	1.0000	0.9990	0.9950	0.9910	0.9820

*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 =500; Number of Replications = 1000;

	$\lambda = 1$		$\lambda = 0.43$		$\lambda = 0.1$	
	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 4

Variables: Names, Definitions and Descriptive Summary Statistics.

Variable Name	Variable Definition	Sample Mean	St. Dev.
WAGE	Hourly wage in Canadian $.	9.8449	4.3868
HEAD	Family status dummy variable; =1 if person is head of family, =0 otherwise.	0.8922	0.3101
AGE2	Age dummy variable, =1 if person is 35-44 years old, =0 otherwise.	0.29851	0.4577
AGE3	Age dummy variable, =1 if person is 45-54 years old, =0 otherwise	0.2329	0.4228
OCCSK	The skill mix is the employment weighted average of the GED and SVP scores of all the occupations employed in the respondent's industry	8.4719	2.6416
PES1	Percentage of wage earners in the industry	0.6569	0.2618
PRIV	Private sector dummy variable, =1 if employed in private sector, =0 otherwise	0.7904	0.4072
FTEN	Number of years with the same employer.	11.2336	9.2837
SQFTEN	Square of FTEN	212.3264	325.2961
INDPU	Percentage of unionized worker in the industry	0.5706	0.2844
FTIME	Employment status dummy variable, =1, if in full time employment,	0.9675	0.1772

Table 5

Estimates of the wage equation H_c model.

Dependent Variable : Hourly Wage

Independent Variables	Estimates	Standard Errors
PRIV	-0.0433	0.0817
AGE2	0.1217	0.0423
AGE3	0.1346	0.0430
PES1	0.2627	0.0757
OCCSK	0.0614	0.0080
FTIME	0.2717	0.1504
FTEN	0.0174	0.0055
SQFTEN	-0.0002	0.0001
HEAD	0.2151	0.0594
THETA	0.3015	0.1152