NONLINEAR REGRESSION WITH DISCRETE EXPLANATORY VARIABLES, WITH AN APPLICATION TO THE EARNINGS FUNCTION

H.J. Bierens
J. Hartog

Research memorandum 1988-3 Febr.'88

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM
NONLINEAR REGRESSION WITH DISCRETE EXPLANATORY VARIABLES, WITH AN APPLICATION TO THE EARNINGS FUNCTION: *)

MATHEMATICAL APPENDIX **) 

by Herman J. Bierens 1) and Joop Hartog 2) 

1) Department of Econometrics 
   Free University 
   P.O. Box 7161 
   1007 MC Amsterdam 

2) Department of Economics 
   University of Amsterdam 
   Jodenbreestraat 23 
   1011 NH Amsterdam 

*) Forthcoming in the Journal of Econometrics

**) This separate appendix contains the proofs of the theorems in the published article.
APPENDIX: Mathematical Proofs

Proof of Theorem 1. Let $\theta$ be a random drawing from $F$. Then by hypothesis of the theorem the distribution of the random variable $\theta'(x_1 - x_2)$, where $x_1 \in X$ and $x_2 \in X$ are non-random, is atomless for $x_1 \neq x_2$. Consequently we have

$$P(\theta'(x_1 - x_2) = 0) = \begin{cases} 
1 & \text{if } x_1 = x_2, \\
0 & \text{if } x_1 \neq x_2.
\end{cases}$$

From (A1) and the countability of $X$ we now conclude

$$(A2) \quad P(\theta'x_1 = \theta'x_2 \text{ for some } (x_1, x_2) \in X \times X \text{ with } x_1 \neq x_2) \\ \leq \sum_{x_1 \in X, x_2 \in X, x_1 \neq x_2} P(\theta'(x_1 - x_2) = 0) = 0$$

This proves the theorem. Q.E.D.

Proof of Theorem 4. The strong consistency results in Theorem 4 follow straightforwardly from Kolmogorov's strong law of large numbers and Theorem 2.2.5 of Bierens (1981). For proving asymptotic normality, observe that by the central limit theorem

$$(A3) \quad \sqrt{n} \left( \sum_{j=1}^{n} x_j x_j' y_j \right) \xrightarrow{d} N_k \left(0, \beta_0 \right),$$

and that by Kolmogorov's strong law of large numbers,

$$(A4) \quad \frac{1}{n} \sum_{j=1}^{n} x_j x_j' \rightarrow E x_j x_j' \quad \text{a.s.},$$

The asymptotic normality result follows now from Theorem 2.2.14 in Bierens (1981). A similar proof can be found in White (1980). Q.E.D.

Proof of Theorem 5. Since $\theta_0$ is a linear separator its components are non-zero, possibly except the components corresponding with nonvarying components of $X$. Therefore the functions $N_1(\theta)$ are continuously differentiable in a neighborhood of $\theta_0$, and so is $z(x, \theta)$ for each $x \in X$. Using
Theorem 2.3.3. of Bierens (1981) it is now not hard to verify that for some compact neighborhood $S_0$ of $\theta_0$,

\begin{equation}
(A5) \quad \sup_{|z| \leq 1} \sup_{\theta \in S_0} |\psi_{z}(z|\theta) - \tilde{\psi}_{z}(z|\theta)| \to 0 \quad \text{a.s.}
\end{equation}

for $z=0,1,2,\ldots$, where similarly to (12).

\begin{equation}
(A6) \quad E_{\psi_{x_1}}(z(x_j,\theta)|\theta) = \begin{cases} 1 & \text{if } r_1 = r_2, \\ 0 & \text{if } r_1 \neq r_2. \end{cases}
\end{equation}

Defining

\begin{equation}
(A7) \quad \tilde{\gamma}_{z}(\theta) = \mathbb{E} y \tilde{\psi}_{z}(z(x_j,\theta)|\theta)
\end{equation}

it follows from (A5) that

\begin{equation}
(A8) \quad \sup_{\theta \in S_0} |\tilde{\gamma}_{z}(\theta) - \tilde{\gamma}_{2}(\theta)| \to 0 \quad \text{a.s.}
\end{equation}

Since by Theorem 4, $\hat{\theta} \to \theta_0$ a.s., the theorem under review now easily follows from (A6) and (A8) and Theorem 2.2.5 of Bierens (1981).

Q.E.D.

For proving Theorems 6 and 7 we need the following lemma's.

**Lemma A1.** Let $u$ be a random variable in $\mathbb{R}$, satisfying $E|u| < \infty$ and let $z$ be a random variable in a bounded subset $Z$ of $\mathbb{R}$. Then $P(E(u|z) = 0) < 1$ if and only if for some $\delta > 0$, $E_u^{\tau Z} \neq 0$ for all $\tau \in (-\delta,0) \cup (0,\delta)$.

**Proof.** Lemma A1 follows straightforwardly from the Proof of Theorem 2 of Bierens (1982).

Q.E.D.

**Lemma A2.** Let the conditions of Lemma A1 be satisfied. Let

\begin{equation}
(A9) \quad T = \{\tau \in \mathbb{R} : E_u^{\tau Z} = 0\}.
\end{equation}
If \( P(E(u|z) = 0) < 1 \) then \( T \) is countable and any bounded subset of \( T \) is finite.

Proof: Let \( \tau_0 \in T \). From lemma A1 it follows that there exists a \( \delta > 0 \) such that

\[
E \mu E^{\tau_0} e^{\tau z} \neq 0 \quad \text{for all } \tau \in (-\delta, 0) \cup (0, \delta),
\]

hence

\[
E \mu E^{\tau z} \neq 0 \quad \text{for all } \tau \in (-\delta+\tau_0, \tau_0) \cup (\tau_0, \tau_0+\delta).
\]

Obviously (A11) implies that for every \( \tau_0 \in T \)

\[
\inf_{\tau \in T, \tau \neq \tau_0} |\tau - \tau_0| > 0,
\]

which in its turn implies that Lemma A2 holds.

Q.E.D.

Proof of Theorem 6. First we note that we may replace \( \tilde{z}(x_j, \theta^*) \) in (14), (15) and (1.7) by \( \theta^* x_j \). However, using \( \tilde{z} \) has the advantage that \( \tau \) then becomes independent of the scale of \( \theta^* x_j \).

Now let

\[
\bar{z}(x_j, \theta^*) = \frac{2 \theta^* x_j - \max_{x \in X} \theta^* x - \min_{x \in X} \theta^* x}{\max_{x \in X} \theta^* x - \min_{x \in X} \theta^* x}.
\]

Since \( X \) is finite and contains only points with positive probability mass, we have

\[
\lim_{n \to \infty} P(\{x_1, \ldots, x_n\} \supseteq X) = 1,
\]

hence

\[
\lim_{n \to \infty} P(\tilde{z}(x_j, \theta^*) = \bar{z}(x_j, \theta^*) \text{ for } j = 1, \ldots, n) = 1.
\]

Therefore we may replace \( \tilde{z} \) by \( \bar{z} \) without loss of generality.

Assume that (18) holds. Then
(A16) \[ E(y_j | x_j) = g(x_j) = \sum_{l=0}^{m-1} \gamma_l(\theta_0) \psi_l(z(x_j, \theta_0) | \theta_0) \quad \text{a.s.} \]

so that

\[
(A17) \quad (1/\sqrt{n}) \sum_{j=1}^{n} (y_j - \hat{g}_m(x_j | \theta)) e^{i \bar{z}(x_j, \theta^*)} = (1/\sqrt{n}) \sum_{j=1}^{n} u_j \psi_{l_0}(z(x_j, \theta_0) | \theta_0) e^{i \bar{z}(x_j, \theta^*)}
\]

\[
- (1/\sqrt{n}) \sum_{j=1}^{n} \sum_{l \geq 0} (\hat{\gamma}_l(\theta) - \gamma_l(\theta_0)) \psi_l(z(x_j, \theta_0) | \theta_0) e^{i \bar{z}(x_j, \theta^*)}
\]

\[
= \tilde{c}_1(\tau, \theta^*) - \tilde{c}_2(\tau, \theta^*) - \tilde{c}_3(\tau, \theta^*), \text{ say.}
\]

Observe from (9) that for \( \tau = 0, 1, 2, \ldots \)

\[
(A18) \quad \tilde{\gamma}_l(\theta_0) - \gamma_l(\theta_0) = \frac{1}{n} \sum_{j=1}^{n} u_j \psi_l(z(x_j, \theta_0) | \theta_0),
\]

hence

\[
(A19) \quad \tilde{c}_2(\tau, \theta^*) = \sum_{l=0}^{m-1} \left( \frac{1}{n} \sum_{j=1}^{n} u_j \psi_l(z(x_j, \theta_0) | \theta_0) \right) e^{i \bar{z}(x_j, \theta^*)}
\]

Denoting

\[
(A20) \quad \tilde{c}_2(\tau, \theta^*) = \sum_{l=0}^{m-1} \left( \frac{1}{n} \sum_{j=1}^{n} u_j \bar{\psi}_l(z(x_j, \theta_0) | \theta_0) E(\bar{\psi}_l(z(x_j, \theta_0) | \theta_0) e^{i \bar{z}(x_j, \theta^*)}) \right),
\]

where \( \bar{\psi}_l(x | \theta_0) \) is the probability limit of \( \psi_l(x | \theta_0) \), it is easy to verify that

\[
(A21) \quad \lim_{n \to \infty} \{ \tilde{c}_2(\tau, \theta^*) - \tilde{c}_2(\tau, \theta^*) \} = 0.
\]

Next, observe that by the mean value theorem there exists a random vector \( \delta(\tau, \theta^*) \) satisfying

\[
(A22) \quad ||\delta(\tau, \theta^*) - \theta_0|| \leq ||\delta - \theta_0|| \quad \text{a.s.}
\]
and
\( (A23) \quad \tilde{c}_3(x, \theta^*) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left( \hat{g}_m(x_j | \delta) - \hat{g}_m(x_j | \theta_0) \right) \tau z(x_j, \theta^*) \)

\( = \sqrt{n}(\delta - \theta_0)' \frac{1}{n} \sum_{j=1}^{n} (\partial / \partial \theta^t) \hat{g}_m(x_j | \delta) \tau z(x_j, \theta^*) \).

Moreover, from \((A22)\) it follows
\( (A24) \quad \text{plim}_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (\partial / \partial \theta^t) \hat{g}_m(x_j | \delta) \tau z(x_j, \theta^*) \)

\( = \text{plim}_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (\partial / \partial \theta^t) \hat{g}_m(x_j | \theta_0) \tau z(x_j, \theta^*) \)

\( = \tilde{\xi}_m(\tau | \theta^*), \text{ say.} \)

Denoting
\( (A25) \quad \tilde{c}_3(x, \theta^*) = \sqrt{n}(\delta - \theta_0)' \tilde{\xi}_m(\tau | \theta^*) \)

we thus have
\( (A26) \quad \text{plim}_{n \to \infty} \{ \tilde{c}_3(x, \theta^*) - \tilde{c}_3(x, \theta^*) \} = 0. \)

Furthermore, observe that
\( (A27) \quad \sqrt{n}(\delta - \theta_0) = \left( \frac{1}{n} \sum_{j=1}^{n} x_j x_j^t \right)^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j (y_j - x_j \theta_0) \)

and that by \((A4)\),
\( (A28) \quad \text{plim}_{n \to \infty} \{ \sqrt{n}(\delta - \theta_0) - (Ex_j x_j^t)^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j (y_j - x_j \theta_0) \} = 0. \)

Thus denoting
\( (A29) \quad \tilde{\tilde{c}}_3(x, \theta^*) = \tilde{\xi}_m(\tau, \theta^*) (Ex_j x_j^t)^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j (y_j - x_j \theta_0) \)

we have
\( (A30) \quad \text{plim}_{n \to \infty} \{ \tilde{\tilde{c}}_3(x, \theta^*) - \tilde{\tilde{c}}_3(x, \theta^*) \} = 0. \)
From (A15), (A17), (A21), (A26) and (A30) we now obtain

\[(A31)\quad \text{plim} \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (y_j - \hat{\theta}_{m}(x_j|\theta)) \right\} = 0 ,\]

where

\[(A32)\quad \hat{d}(\tau|\theta^*) = \tilde{c}_1(\tau,\theta^*) - \tilde{c}_2(\tau,\theta^*) - \tilde{c}_3(\tau,\theta^*)
= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} u_j \tilde{\tau}(x_j,\theta^*)
- \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{l=0}^{m-1} \tilde{\Psi}_l(z(x_j,\theta_0)|\theta_0).E(\tilde{\psi}_l(z(x_j,\theta_0)|\theta_0)e^{\tilde{\tau}(x_j,\theta^*)})
- \tilde{\xi}_m(\tau|\theta^*)(E x_jx_j')^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j(y_j-x_j')e^{\tilde{\tau}(x_j,\theta^*)}\]

with

\[(A33)\quad \tilde{\rho}_{j,m}(\tau|\theta^*) = e^{\tilde{\tau}(x_j,\theta^*)} - \sum_{l=0}^{m-1} \tilde{\Psi}_l(z(x_j,\theta_0)|\theta_0).E(\tilde{\psi}_l(z(x_j,\theta_0)|\theta_0)e^{\tilde{\tau}(x_j,\theta^*)})\]

Realizing that the terms in (A32) are i.i.d. with zero mean and variance

\[(A34)\quad \tilde{s}_j^2(\tau|\theta^*) = E[u_j \tilde{\rho}_{j,m}(\tau|\theta^*) - \tilde{\xi}_m(\tau|\theta^*)'(E x_jx_j')^{-1}x_j(y_j-x_j')\theta_0)]^2
= E[u_j^2 \tilde{\rho}_{j,m}^2(\tau|\theta^*) - 2E(u_j \tilde{\rho}_{j,m}(\tau|\theta^*)\tilde{\xi}_m(\tau|\theta^*)'(E x_jx_j')^{-1}x_j(y_j-x_j')\theta_0)]
+ \tilde{\xi}_m(\tau|\theta^*)'(E x_jx_j')^{-1}(E(y_j-x_j')^2x_jx_j')(E x_jx_j')^{-1} \tilde{\xi}_m(\tau|\theta^*)
\]

we have by the central limit theorem

\[(A35)\quad \tilde{d}_m(\tau|\theta^*) \rightarrow N(0, \tilde{s}_m^2(\tau|\theta^*)) \text{ in distr.}\]

We leave it to the reader to verify that

\[(A36)\quad \tilde{s}_m^2(\tau|\theta^*) = \tilde{s}_m^2(\tau|\theta^*) \text{ a.s.}\]

and that \(\tilde{s}_m^2(\tau|\theta^*) > 0\) for \(\tau \neq 0\). Combining (A31), (A35) and (A36), part
(19) of Theorem 6 follows.

Next, assume that (18) fails to hold. Then by (A15),

$$\frac{1}{n} \sum_{j=1}^{n} (y_j - \hat{\theta}_m(x_j, \delta)) e^{-\mathcal{L}(x_j, \theta^*)}$$

$$\rightarrow E(y_j - \sum_{j=0}^{m-1} \gamma_j(\theta_0) \psi_j(z(x_j, \theta_0) | \theta_0)) e^{-\mathcal{L}(x_j, \theta^*)} \text{ a.s.}$$

Moreover, it is not hard to verify that also now $\bar{s}_m^2(\tau|\theta^*)$ converges a.s. to a limit $\bar{s}_m^2(\tau|\theta^*)$, say, which is positive for $\tau \neq 0$. Thus part (20a) follows straightforwardly from (A37) and Lemma's A1 and A2.

Finally, the conclusion that we may substitute $\hat{\theta}$ for $\theta^*$ follows from the fact that by Theorem 4

$$\lim \{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (y_j - \hat{\theta}_m(x_j, \delta)) e^{-\mathcal{L}(x_j, \delta)}$$

$$- \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (y_j - \hat{\theta}_m(x_j, \delta)) e^{-\mathcal{L}(x_j, \theta_0)} \} = 0$$

provided (18) is satisfied. Proving (A38) is not too hard and therefore left to the reader.

Q.E.D.

Proof of Theorem 7. The result (19) is equivalent with

$$\lim_{n \to \infty} E e^{i \tau \hat{\theta}_m} (\tau|\theta^*) = e^{-\frac{1}{2} \tau^2} \text{ for every } \tau \in \mathbb{R}$$

If $\tau$ and $\theta^*$ are random and independent from the data-generating process then we have similarly,

$$E(e^{i \tau \hat{\theta}_m} (\tau|\theta^*) | \tau, \theta^*) \rightarrow e^{-\frac{1}{2} t^2} \text{ a.s.}$$

Hence by bounded convergence,

$$E e^{i \tau \hat{\theta}_m} (\tau|\theta^*) = E[E(e^{i \tau \hat{\theta}_m} (\tau|\theta^*) | \tau, \theta^*)] \rightarrow e^{-\frac{1}{2} t^2}$$

which proves that (19) carries over if $\theta^*$ and $\tau$ are random.

Now suppose that (18) fails to hold. Lemma A2 implies that (20a) hold for $\tau \in \mathbb{R}\setminus T$, where $T$ is a countable subset of $\mathbb{R}$. But since $\tau$ is now continuously distributed we have
Moreover, Theorem 1 implies that $\theta^*$ is a.s. a linear separator. Therefore (20a) also holds for the random $t$ and $\theta^*$ involved. Q.E.D.

Proof of Theorem 8. From the mean value theorem it follows

\[(A43) \quad \hat{\gamma}_k(\delta) - \hat{\gamma}_k(\theta_0) = [(\delta/\theta)\hat{\gamma}_k(\delta)] (\delta - \theta_0),\]

where $\delta$ is a mean value satisfying $||\delta - \theta_0|| \leq ||\delta - \theta_0||$. From this result we see

\[(A44) \quad \text{plim} \left\{ \frac{1}{n} \begin{bmatrix} \hat{\gamma}_0(\delta) \\ \vdots \\ \hat{\gamma}_{m-1}(\delta) \end{bmatrix} - \begin{bmatrix} \hat{\gamma}_0(\theta_0) \\ \vdots \\ \hat{\gamma}_{m-1}(\theta_0) \end{bmatrix} \right\} = \bar{\Gamma}_m / n (\delta - \theta_0) = 0,\]

where

\[(A45) \quad \bar{\Gamma}_m = \text{plim} \Gamma_m.\]

Moreover, from (9) and (10) it follows that

\[(A46) \quad \hat{\gamma}_k(\theta_0) - \gamma_k(\theta_0) = - \frac{1}{n} \Sigma_j^m \Sigma_j^m u_j \psi(x_j, \theta_0)|\theta_0).\]

Combining (A44) and (A46) and using (A28) we see that

\[(A47) \quad \text{plim} \left\{ \frac{1}{n} \begin{bmatrix} \hat{\gamma}_0(\delta) \\ \vdots \\ \hat{\gamma}_{m-1}(\delta) \end{bmatrix} - \begin{bmatrix} \hat{\gamma}_0(\theta_0) \\ \vdots \\ \hat{\gamma}_{m-1}(\theta_0) \end{bmatrix} \right\} - \left\{ \bar{\Gamma}_m (E x_j x_j')^{-1} \frac{1}{\sqrt{n}} \Sigma_j^m x_j (\gamma_j - x_j' \theta_0) + \frac{1}{\sqrt{n}} \Sigma_j^m u_j \psi(x_j, \theta_0) \right\} = 0\]

where

\[(A48) \quad \psi_{j,m} = \begin{bmatrix} \psi_0(z(x_j, \theta_0)|\theta_0) \\ \vdots \\ \psi_{m-1}(z(x_j, \theta_0)|\theta_0) \end{bmatrix}\]

But the random vectors
are independent with zero mean vector and variance matrix

\[ (A50) \quad E d_j d_j' = \bar{\Gamma}_m \Omega \bar{\Gamma}_m + \bar{\Gamma}_m (\text{Ex}_j x_j^t)^{-1} E(u_j^2 x_j \psi'_j, \psi_j) + E(u_j^2 \psi'_j, \psi_j) \]

Denoting

\[ (A51) \quad \bar{\Delta}_m = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} E u_j^2 \psi'_j, \psi_j \]

\[ (A52) \quad \bar{\Sigma}_m = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} (u_j^2 \psi'_j, \psi_j) (\text{Ex}_j x_j^t)^{-1} \]

\[ (A53) \quad \Lambda_m = \bar{\Gamma}_m \Omega \bar{\Gamma}_m + \bar{\Delta}_m \bar{\Gamma}_m + \bar{\Gamma}_m \bar{\Sigma}_m + \bar{\Delta}_m \]

we thus have by the central limit theorem

\[ (A54) \quad \frac{1}{\sqrt{n}} \sum_{j=1}^{n} d_j \to N_m (0, \Lambda_m) \text{ in distr.} \]

Combining (A47) and (A54), the first part of Theorem 8 follows.

We leave it to the reader to verify the second part.

Q.E.D.

Proof of Theorem 9: Similarly to the proof of Theorem 7.