

CONSISTENT ESTIMATION OF THE PARAMETERS AND ERROR
DENSITY IN A CENSORED REGRESSION MODEL

by

Douglas Nychka and A. Ronald Gallant

Institute of Statistics Mimeograph Series No. 1642

June 1984

Consistent Estimation of the Parameters and Error
Density in a Censored Regression Model

by

Douglas Nychka and A. Ronald Gallant
Department of Statistics
North Carolina State University

1. Introduction

The underlying model in this discussion has the form:

$$\begin{aligned}y_1 &= f_1(X, \theta) + U_1 \\y_2 &= f_2(X, \theta) + U_2\end{aligned}\tag{1.1}$$

where $(U_1, U_2) = U$ is a mean zero random vector with probability density function h . f_1 and f_2 are specified, continuous functions of the exogeneous variable X and the parameters θ . θ and h are unknown.

The interest in this paper concerns estimating θ and h when y is not observed directly. We consider the data, z , where

$$\begin{aligned}z_1 &= y_1 I_{[0, \infty)}(y_2) \\z_2 &= I_{(-\infty, 0]}(y_2)\end{aligned}\tag{1.2}$$

Thus, y_1 is only observed conditional on the event $\{y_2 > 0\}$. Also, note that y_2 is never observed explicitly; only its sign is known. This particular observational model has a variety of applications in economics, psychology and education. One important example arises in labor economics for evaluating training programs where the second equation represents a selection rule such as voluntary selection or selection by program administrators.

Using this observational model, we propose consistent estimates for θ and h based on the maximum likelihood criterion. In order to define these estimates we first give the probability density for z and state the assumptions made on θ and h .

Let μ denote Lebesgue measure and δ_v be a measure that gives the point v unit mass. Then the joint distribution of (z_1, z_2) is absolutely continuous with respect to the product measure $(\mu + \delta_0) \times (\delta_0 + \delta_1)$. Using a conditioning argument, the log density of z with respect to this dominating measure is

$$g_{\theta, h}(z, X) = z_2 \log \left\{ \int_{-f_2(X, \theta)}^{\infty} h(z_1 - f_1(X, \theta), w) dw \right\} I_{\mathbb{R} - \{0\}}(z_1) \quad (1.3)$$

$$+ (1 - z_2) \log \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{-f_2(X, \theta)} h(w_1, w_2) dw_2 dw_1 \right\}$$

Given n independent observations the log-likelihood has the form:

$$\ell_n(\theta, h) = \sum_{t=1}^n g_{\theta, h}(z_t, X_t) .$$

We will assume that $\theta \in \Theta$, $h \in H$ where Θ is a compact metric space (usually a closed and bounded set in \mathbb{R}^k) and H is a bounded subset of an $(m+1)$ order Sobolev space on \mathbb{R}^2 . Finally let ϕ be a positive weight function, p^K a polynomial of degree K in \mathbb{R}^2 and $F_K = \{f: f = (p^K \phi)^2\}$. Set

$$F = \bigcup_{K=1}^{\infty} F_K \text{ and } H_k = F_k \cap H . \quad (1.4a)$$

We will assume that $K = K_n \rightarrow \infty$ as $n \rightarrow \infty$. Then the estimates $(\hat{\theta}_n, \hat{h}_n) \in \Theta \times H_{K_n}$ satisfy:

$$\ell_n(\hat{\theta}_n, \hat{h}_n) = \max_{(\theta, h) \in \Theta \times H_{K_n}} \ell_n(\theta, h) .$$

In order to define H we first present some notation concerning function spaces.

Define the differential operator for functions on \mathbb{R}^2 as

$$D^P f = \frac{\partial^{P_1}}{\partial v_1^{P_1}} \frac{\partial^{P_2}}{\partial v_2^{P_2}} f \text{ where } P = (P_1, P_2) \text{ and } |P| = P_1 + P_2 .$$

We will use the Sobolev Spaces (m^{th} order):

$$W^{m,q}(A) = \{f: D^P(f) \in L^q(A) \forall |P| \leq m\} \quad (1.4b)$$

for $m > 0$. Note that $W^{0,2}(A) = L^2(A)$ and in general $W^{m,2}(A)$ is a Hilbert Space with respect to the inner product:

$$\langle f, g \rangle_{W^{m,2}(A)} = \sum_{|P| \leq m} \langle D^P f, D^P g \rangle_{L^2(A)}$$

(see Adams, 1976). To simplify notation, in this paper when $A = \mathbb{R}^2$ we will often omit the domain in (1.4b): $W^{m,2} \equiv W^{m,2}(\mathbb{R}^2)$.

For $R > 0$ let

$$B_R = \{f \in W^{m+1,2}(\mathbb{R}^2) : \|f\|_{W^{m+1,2}} \leq R\}$$

and let $\psi_L, \psi_H \in B_R$ such that $0 < \psi_L \leq \psi_H$ and $v_1 v_2 \psi_H(v)$ is integrable.

Then we have,

$$H = \{h \in B_R : \int_{\mathbb{R}^2} h \, dv = 1, \int v_1 v_2 h(v) \, dv = 0\}$$

$$|D^P \psi_L| \leq |D^P h| \leq |D^P \psi_H| \text{ for } |P| \leq m, \sqrt{h} \in W^{m+1,2}.$$

Thus H is a set of density functions with zero mean and whose tail behavior lies between the extremes of ψ_L and ψ_H . Moreover, the restriction of $h \in B_R$ is an implicit assumption on the smoothness of h . This constraint limits the amount of oscillation in h and its derivative. The requirement that $\sqrt{h} \in W^{m+1,2}$ is necessary to insure that the weighted polynomials in F can be used to approximate functions in H . (see Lemma 4.2)

Let θ^* and h^* denote the true values of θ and h . Under conditions A-C (see Section 2) which insure that the parameters are identifiable and assuming conditions I-III which concern the Cesaro summability of $\{U_k, X_k\}_{K_1=1, \infty}$ and the tail behavior of the densities in H (see Section 3). We prove:

Theorem 1.1

Suppose $\theta^* \in \Theta$ and $h^* \in \overline{\bigcup_{k=1}^{\infty} H_k}$ where the closure is taken with respect to $W^{m,2}(\mathbb{R}^2)$. If $m \geq 2$ then $\hat{\theta}_n \rightarrow \theta^*$ a.s. in Θ

$$\text{and } \hat{h}_n \rightarrow h^* \text{ a.s. in } W^{m,2}(\mathbb{R}^2) . \quad (1.4c)$$

We note that the convergence with respect to m^{th} order Sobolev norm is quite strong. For example (1.4b) implies that

$$\sup_{U \in \mathbb{R}^2} |(D^P \hat{h}_n)(U) - (D^P h)(U)| \rightarrow 0 \text{ a.s.}$$

for all P such that $0 \leq |P| \leq m-2$. In general, if Λ is any continuous functional on $W^{m,2}(\mathbb{R}^2)$ then $|\Lambda(\hat{h}_n) - \Lambda(h^*)| \rightarrow 0$ a.s. .

The main difficulty in interpreting this theorem is relating $\overline{\bigcup_{k=1}^{\infty} H_k}$ to H . Although in Section 4 we show that $\bigcup_{k=1}^{\infty} F_k$ is a dense set of functions in H , this is not enough to guarantee that $\bigcup_{k=1}^{\infty} H_k$ will also be dense. (The problem arises in taking the intersection: $H_k = F_k \cap H$ and in general, $\overline{\bigcup_{k=1}^{\infty} H_k}$ may not be equal to H .) In order to obtain consistency for all functions in H we consider a slightly different estimate. Let \bar{H} be the closure of H in $W^{m,2}$ and let $Q: W^{m,2} \rightarrow \bar{H}$ be the projection operator such that

$$Q(g) = h \Leftrightarrow \|g-h\|_{W^{m,2}} = \min_{f \in \bar{H}} \|f-h\|_{W^{m,2}} . \quad (1.5)$$

Thus, Q in this case is the best approximation to f by a density satisfying the constraints in H . Since \bar{H} is a closed, convex set in $W^{m,2}$ such an operator is well-defined (see Rudin, 1974, p. 83).

Now define the estimates $\bar{\theta}_n, \bar{h}_n$ such that $\bar{\theta}_n \in \Theta$, $\bar{h}_n = Q(g)$ for some $g \in F_{K_n}$ and

$$\ell_n(\bar{\theta}_n, \bar{h}_n) = \max_{f \in F_{K_n}, \theta \in \Theta} \ell_n(\theta, Q(f)) .$$

Using this estimate we prove.

Theorem 1.2

Suppose $F \subseteq W^{m+1,2}(\mathbb{R}^2)$ and $m \geq 2$. If $h^* \in H$ and $\theta^* \in \Theta$ then $\bar{\theta}_n \rightarrow \theta^*$ a.s. in Θ and $\bar{h}_n \rightarrow h^*$ a.s. in $W^{m,2}$.

The next section gives conditions under which θ, h are identifiable and the expected log-likelihood has a unique maximum. Section 3 proves Theorem 1.1 and the following section proves Theorem 1.2.

2. Identifiability of the Parameters

Let X denote the space containing the exogenous variable x and let ν be a measure on X that is the weak limit of the empirical distributions of $\{X_t\}_{t=1, \infty}$.

Identifiability conditions:

Let $\theta, \theta' \in \Theta$.

A) If

$$f_1(x, \theta) = f_1(x, \theta')$$

$$f_2(x, \theta) = f_2(x, \theta') \quad \text{a.e. } \nu$$

then $\theta = \theta'$.

B) If $f_2(x, \theta) = \Phi(f_2(x, \theta'))$ a.e. ν for some monotone transformation Φ then $f_2(x, \theta) = f_2(x, \theta')$.

C) Let B be any rectangle in \mathbb{R}^2 and let $\theta \in \Theta$.

$$\text{If } F = \{x \in X : \{f_1(x, \theta), f_2(x, \theta)\} \in B\} \text{ then } \nu(F) > 0.$$

The later two conditions may appear unusual and require some comment. In B), the invariance of f_2 under monotone functions is necessary because y_2 is never observed directly, only its sign is known.

To illustrate the problem encountered in this situation suppose $\Theta, X \subseteq \mathbb{R}$ and $f(x, \theta) = x\theta$. Then,

$$P(y_2 < 0 | x) = \int_{-\infty}^{\infty} \int_{-\infty}^{-x\theta} h(u) du .$$

However, if we make the change of variables $w = cu$ where c is some arbitrary value

$$\begin{aligned}
 P(y_2 < |x) &= \int_{-\infty}^{\infty} \int_{-\infty}^{-x\theta c} \frac{1}{c} h(w) dw \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{-x\theta'} h'(w) dw
 \end{aligned}$$

where $\theta' = \theta c$, $h' = \frac{1}{c} h$.

Thus, the parameters of the model can be varied without changing the probability distribution of z_2 .

Condition C insures that the distribution of the exogenous variables is rich enough so that $(f_1(x, \theta), f_2(x, \theta))$ will trace out the support of $h(u)$. Without this requirement, the error density might only be identifiable on a region smaller than its support. If $\chi \subseteq \mathbb{R}^{\ell}$ and ν is dominated by Lebesgue measure then condition C) implies that for fixed $\theta \in \Theta$ the map $\{f_1(x, \theta), f_2(x, \theta)\}: X \rightarrow \mathbb{R}^2$ is onto i.e. for any $u \in \mathbb{R}^2$ there is an $x \in X$ such that $(f_1(x, \theta), f_2(x, \theta)) = u$.

The following two theorems address the identifiability and uniqueness of the maximum likelihood estimates of the parameters and error density. For technical reasons that will be clearer in Section 3 (see Lemmas 3.1 and 3.2) we state these theorems for $h \in \bar{H}$, where the closure of H is taken with respect to $W^{m,2}$. Although elements in the closure will not be contained in B_R they will still be densities with zero mean.

Theorem 2.1

Suppose conditions A-C hold. Let $h, h' \in \bar{H}$, $\theta, \theta' \in \Theta$.

If

$$g_{h, \theta}(z, x) = g_{h', \theta'}(z, x) \tag{2.1}$$

a.e. for $z \in \mathbb{R} \times \{0, 1\}$, $x \in \chi$

then $h = h'$ and $\theta = \theta'$.

Proof

Using the fact that log is monotone (2.1) implies

$$\begin{aligned} \int_{-f_2(x,\theta)}^{\infty} h(z_1 - f_1(x,\theta), w) dw \\ = \int_{-f_2(x,\theta')}^{\infty} h'(z_1 - f_1(x,\theta'), w) dw \end{aligned} \quad (2.2)$$

and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{-f_2(x,\theta)} h(w_1, w_2) dw_1 dw_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{-f_2(x,\theta')} h'(w_1, w_2) dw_1 dw_2 . \quad (2.3)$$

Note that (2.3) can be rewritten as

$$H_2(-f_2(x,\theta)) = H_2'(-f_2(x,\theta')) \quad (2.4)$$

where H_2 and H_2' are the marginal distribution functions for the second variable in h and h' . Thus we have

$$(H_2'^{-1} \circ H_2)(-f_2(x,\theta)) = -f_2(x,\theta') \quad \text{a.e. } x \in X$$

since $H_2'^{-1} \circ H_2$ is a monotone transformation by B, $f_2(x,\theta) = f_2(x,\theta')$. Now referring to (2.2) by C we can conclude that these integrals are equal for any interval. Thus

$$h(z_1 - f_1(x,\theta), w_2) = h'(z_1 - f_1(x,\theta'), w_2)$$

a.e. $(z_1, w_2) \in \mathbb{R}^2$ or

$$h(z_1 - f_1(x,\theta), w_2) = h'(z_1 - f_1(x,\theta) + \delta, w_2)$$

where $\delta = f_1(x,\theta) - f_1(x,\theta')$. Thus, h, h' can only differ by a shift in location.

However, since both densities have zero mean $\delta = 0$ or $f_1(x,\theta) = f_1(x,\theta')$.

Therefore $h = h'$ and by A, $\theta = \theta'$.

QED

Next we show that the expected value of the log-likelihood of z has a unique maximum at the true parameters.

Theorem 2.2

Let h^* and θ^* denote the true values of the parameters and $P_{\theta^*, h^*}(\cdot | x)$ the conditional probability measure for z given x . Assume $v(X) < \infty$.

If $\theta^* \in \Theta$, $h^* \in \bar{H}$ then

$$s(\theta, h) = \int_X \int_{\mathbb{R} \times \{0,1\}} g_{\theta, h}(z, x) dP_{\theta^*, h^*}(z|x) dv(x) \quad (2.5)$$

achieves a unique maximum at $\theta = \theta^*$ and $h = h^*$.

Proof: Since $-\log$ is strictly convex by the usual application of Jensen's inequality:

$$\begin{aligned} & -\log[E(\exp\{g_{\theta, h}(z, x) - g_{\theta^*, h^*}(z, x)\} | X)] \\ & \leq E(g_{\theta, h}(z, x) - g_{\theta^*, h^*}(z, x) | X) \end{aligned}$$

which gives,

$$\begin{aligned} 0 & \leq \int_{\mathbb{R} \times \{0,1\}} g_{\theta^*, h^*}(z, x) dP_{\theta^*, h^*}(z|x) - \\ & \int_{\mathbb{R} \times \{0,1\}} g_{\theta, h}(z, x) dP_{\theta^*, h^*}(z|x) \end{aligned}$$

$$\theta \in \Theta, h \in H \text{ and a.e. on } X$$

with equality holding only if $g_{\theta, h} = g_{\theta^*, h^*}$.

Thus $S(\theta, h) \leq S(\theta^*, h^*) \forall \theta \in \Theta, h \in \bar{H}$. Moreover, when this maximum is attained from the remarks above and Theorem 2.1,

$$\theta = \theta^*, h = h^* .$$

QED

3. Consistency of estimates

We first state the necessary assumptions for Theorem 1.1.

Condition I (Ceasaro Summability)

If $b(u, x)$ is a continuous function on $\mathbb{R}^2 \times \psi$ then

$$\frac{1}{n} \sum_{t=1}^n b(u_t, x_t) \rightarrow \int_X \int_{\mathbb{R}^2} b(u, x) h(u) du dv(x)$$

Condition II (Continuity of likelihood)

$$\sup_{u_1 \in \mathbb{R}} \int_{\mathbb{R}} \psi_H(u_1, w) dw = J < \infty$$

Let

$$g_{\theta, h}^{(1)}(u_1, x) = \log \left\{ \int_{-f_2(x, \theta)}^{\infty} h(u_1 - f_1(x, \theta), w) dw \right\}$$

$$g_{\theta, h}^{(2)}(x) = \log \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{-f_2(x, \theta)} h(w_1, w_2) dw_1 dw_2 \right\}$$

$$b(u_1, x) = \sup_{\theta \in \Theta} \left\{ |g_{\theta, \psi_L}^{(1)}(u, x)|, |g_{\theta, \psi_L}^{(2)}(x)|, \log(J) \right\}$$

$$\text{and } M(x) = \sup_{u_2 \in \mathbb{R}} \int b(u_1, x) \psi_U(u_1, u_2) du_1$$

Condition III (Continuity of expected log-likelihood)

$$\int_{\chi} M(x) dv(x) < \infty$$

The main idea behind the proof of Theorem 1.1 is to first prove consistency for the estimate $\tilde{\theta}_n, \tilde{h}_n$ where

$$\ell_n(\tilde{\theta}_n, \tilde{h}_n) = \max_{\theta \in \Theta, h \in \bar{H}} \ell_n(\theta, h) . \quad (3.1a)$$

Note that the only difference between $(\tilde{\theta}_n, \tilde{h}_n)$ and $(\hat{\theta}_n, \hat{h}_n)$ is that the maximization in (3.1a) is defined over all of \bar{H} rather than H_K . The next step is to show that the difference between these two estimates converges almost surely to zero. Hence, the consistency of the original estimates follows. The next lemma gives a compactness property for H . This result and Lemmas 3.2-3.4 will be used to establish the consistency of $(\tilde{\theta}_n, \tilde{h}_n)$.

Lemma 3.1

H is precompact in $W^{m,2}(\mathbb{R}^2)$

Proof: Let $\{h_n\}_{n=1, \infty} \subseteq H$. We will show that there exists a subsequence of $\{h_n\}$ that is a Cauchy sequence. Since $W^{m,2}$ is a complete space this subsequence must have a limit which lies in the closure of H . Hence, H is precompact.

$$\text{Let } A_J: \{v \in \mathbb{R}^2: \|v\|_{\mathbb{R}^2} \leq J\} \quad (3.1b)$$

and let Γ_J be the operator that restricts a function with domain on \mathbb{R}^2 to one with domain on A_J . By Theorem 6.2 I, IV (Adams, 1975) $\Gamma_J: W_0^{m+1,2}(\mathbb{R}^2) \rightarrow W^{m,2}(A_J)$ is a compact embedding, where $W_0^{m+1,2}(A)$ is the completion of $C_0^\infty(A)$ with respect to the norm for $W^{m+1,2}(A)$. Moreover, when $A = \mathbb{R}^2$, $W_0^{m+1,2} = W^{m+1,2}$ and thus (Adams, 1976, p. 56) the embedding from $W^{m+1,2}$ into $W^{m,2}$ is also compact. $H \subseteq B_R$ and thus is a bounded set in $W^{m+1,2}$. Therefore, H is pre-compact in $W^{m,2}(A_J)$. Using this result we can extract a subsequence; $\{h_K^{(1)}\} \subseteq \{h_n\}$ that converges to a limit in $\overline{\Gamma_1(H)}$. Similarly, a convergent subsequence on $W^{m,2}(A_2)$ can now be extracted from $\{h_K^{(1)}\}$. In general, we can choose $\{h_K^{(J)}\}$ such that $\{h_K^{(J)}\} \subseteq \{h_K^{(J-1)}\}$ and $\{h_K^{(J)}\}$ has a limit in $W^{m,2}(A_J)$.

Set $g_j = h_j^{(j)}$. Note that this sequence is obtained from the diagonal entries when these subsequences are arranged in a table. Clearly, $\{g_j\} \subseteq \{h_n\}$ and the proof will be completed by showing that $\{g_j\}$ is Cauchy. Let $\alpha_J \in W^{m+1,2}$ with $|D^p \alpha_J| < 1$ $|p| \leq m$ and $\alpha_J(u) = \begin{matrix} 0 & u \in A_{J-1} \\ 1 & u \in A_J^C \end{matrix}$

Take $\epsilon > 0$. For $0 < J \leq j_1 \leq j_2 < \infty$

$$\begin{aligned} \|g_{j_1} - g_{j_2}\|_{W^{m,2}} &\leq \|\Gamma_J(g_{j_1} - g_{j_2})\|_{W^{m,2}(A_J)} \\ &\quad + \|\alpha_J(g_{j_1} - g_{j_2})\|_{W^{m,2}} \end{aligned}$$

From the definition of ψ_M it is straightforward to show

$$\|\alpha_J(g_{j_1} - g_{j_2})\|_{W^{m,2}} \leq 2 \|\alpha_J \psi_M\|_{W^{m,2}} \quad (3.1c)$$

Since $\psi_M \in W^{m+1,2}(\mathbb{R}^2)$, by the dominated convergence theorem $\|\alpha_J \psi_M\|_{W^{m,2}} \rightarrow 0$ as $J \rightarrow \infty$. We will choose J such that (3.1c) is bounded by $\epsilon/2$. Now $\{h^J\}$ is a Cauchy sequence in $W^{m,2}(A_J)$ and by construction $\Gamma_J(g_j) \in \{h_K^J\}$ provided $j > J$.

Thus there is an $M < \infty$ such that $M \geq J$ and $j_1, j_2 \geq M$ implies that the first term on the RHS of (3.1c) is bounded by $\epsilon/2$. Thus (3.1c) is bounded by ϵ for $j_1, j_2 \geq M$ and therefore $\{g_j\}$ is Cauchy sequence. QED

Let $S_n(\theta, h) = \frac{1}{n} \ell_n(\theta, h)$ and take $S(\theta, h)$ as defined in (2.5).

Lemma 3.2 (Gallant, 1984)

Suppose $\Theta \times \bar{H}$ is compact in the usual product topology, $S(\theta, h)$ has a unique maximum at (θ^*, h^*) and S is continuous on $\Theta \times \bar{H}$. If

$$\sup_{\theta, h \in \Theta \times \bar{H}} |S_n(\theta, h) - S(\theta, h)| \rightarrow 0 \quad \text{a.s.} \quad (3.2)$$

then $(\tilde{\theta}_n, \tilde{h}_n) \rightarrow (\theta^*, h^*)$ in the product topology on $\Theta \times \bar{H}$.

Proof. Since $\Theta \times \bar{H}$ is compact the sequence $(\tilde{\theta}_n, \tilde{h}_n)$ will have at least one limit point. Suppose (θ_0, h_0) is such a limit point and $(\tilde{\theta}_m, \tilde{h}_m)$ is a subsequence converging to it. From the definition of the maximum likelihood estimate

$$S_m(\tilde{\theta}_m, \tilde{h}_m) \geq S_m(\theta^*, h^*)$$

and by the assumption of uniform convergence in (3.2),

$S(\theta_0, h_0) \geq S(\theta^*, h^*)$. Since (θ^*, h^*) yields a unique maximum, we can conclude that $(\theta_0, h_0) = (\theta^*, h^*)$. Therefore the sequence of estimates has only one limit point at (θ^*, h^*) .

Lemma 3.3

Under conditions II and III

a) the functionals $g_{\theta, h}^{(1)}(u, x)$ and $g_{\theta, h}^{(2)}(x)$ are continuous on

$$\mathbb{R}^2 \times X \times \Theta \times \bar{H}$$

and

b) $S(\theta, h)$ is continuous on $\Theta \times \bar{H}$.

Proof: Suppose $\{(X_n, U_n, \theta_n, h_n)\}_{n=1, \infty}$ is a sequence converging to (x, u, θ, h)

$$\text{let } \phi_n(w) = \int_{(-f_2(x_n, \theta_n), \infty)} h_n(u, w) \, du .$$

Using the facts that convergence in Sobolev norm with $m \geq 2$ implies the uniform convergence of a function pointwise (see Adams, 1976, p. 97-98) and that f_2 is a continuous function of x and θ we have

$$\phi_n(w) \rightarrow \int_{(-f_2(x, \theta), \infty)} h(u, w) \, du .$$

Since $\phi_n \leq \psi_U$ by condition II we can apply the dominated convergence theorem to conclude that

$$\int \phi_n(w) \rightarrow \int_{-f_2(x, \theta)}^{\infty} h(u, w) \, dw .$$

Finally, noting that the log is a continuous function we have

$g_{\theta_n, h_n}^{(1)}(u, x_n) \rightarrow g_{\theta, h}^{(1)}(u, x)$ and thus this functional is continuous. The continuity of $g^{(2)}$ is proved in a similar manner.

b) From the results in a) it is clear that for fixed u and x $g_{\theta, h}(u, x)$ is a continuous function on $\Theta \times H$.

Also, from condition II

$$\sup_{\substack{\theta \in \Theta \\ h \in H}} |g_{\theta, h}(u, x)| \leq b(u, x)$$

with $\int_{\mathcal{X}} \int_{\mathbb{R}^2} b(u, x) h(u) \, du \, dv(x) < \infty$.

If (θ_n, h_n) is a sequence converging to (θ, h) then by the dominated convergence theorem

$$S_{\theta_n, h_n} \rightarrow S_{\theta, h} \text{ and thus the continuity in b) follows.}$$

Lemma 3.4

Under conditions I-III

$$\sup_{(\theta, h) \in \Theta \times \bar{H}} |S_n(\theta, h) - S(\theta, h)| \rightarrow 0 \text{ a.s.}$$

If $g_{\theta, h}(u, x)$ were continuous in all of its arguments then this result would follow by a Uniform Strong Law of Large Numbers, such as Theorem 1 in (Gallant, 1982). However, because of the indicator functions in the log-likelihood some additional work is required. The idea behind this proof is to approximate $g_{\theta, h}$ by a continuous likelihood and then show that the difference is negligible.

Proof: Let $0 \leq \chi^\varepsilon(z) \leq 1$ be a continuous function such that

$$\chi^\varepsilon(z) = 0 \quad z < -\varepsilon \text{ and}$$

$$\chi^\varepsilon(z) = 1 \quad z > \varepsilon$$

Let

$$g_{\theta, h}^\varepsilon(u, x) = \chi^\varepsilon(u + f_2(x, \theta^*)) g_{\theta, h}^{(1)}(u, x) + \{1 - \chi^\varepsilon(u + f_2(x, \theta^*))\} g_{\theta, h}^2(x)$$

$$S_n^\varepsilon(\theta, h) = \frac{1}{n} \sum_{w=1}^n g_{\theta, h}^\varepsilon(u, x),$$

and

$$S^\varepsilon(\theta, h) = \int_{\mathbb{R}^2} \int_{\mathcal{X}} g_{\theta, h}^\varepsilon(u, x) h(u) du dv(x).$$

Thus $S_n^\varepsilon(\theta, h)$ is identical to $S_n(\theta, h)$, the original log-likelihood, except for a continuous modification of the indicator functions close to 0. Adding and subtracting this modified likelihood gives

$$\begin{aligned} |S_n(\theta, h) - S(\theta, h)| &\leq |S_n^\varepsilon(\theta, h) - S^\varepsilon(\theta, h)| \\ &+ |S^\varepsilon(\theta, h) - S_n(\theta, h)| \end{aligned} \quad (3.3)$$

Now we argue that each term on the RHS of (3.3) converges to zero uniformly in $\Theta \times \bar{H}$.

By construction, $|g_{\theta,h}^\varepsilon(u,x)| \leq b(u,x)$ and is continuous on $\mathbb{R}^2 \times \mathcal{X} \times \Theta \times F$. Thus, by Theorem 1, (Gallant, 1982) the first term in (3.3) converges uniformly to zero.

Considering the second term in (3.3) and applying condition III gives

$$\begin{aligned} \sup_{\theta,h} \int_{\Theta \times F} |S_{\theta,h}^\varepsilon - S_{\theta,h}| &\leq \int_{\mathcal{X}} \int_{\mathbb{R}^2} \mathbb{I}_{(-\varepsilon,\varepsilon)}(u_2 + f_2(x,\theta^*)) b(u,x) h^*(u) du dv(x) \\ &\leq \int_{\mathcal{X}} \int_{\mathbb{R}} \mathbb{I}_{(-\varepsilon,\varepsilon)}(u_2 + f_2(x,\theta^*)) M(x) du_2 dv(x) \\ &\leq 2\varepsilon \int_{\mathcal{X}} M(x) dv(x) . \end{aligned}$$

Since by assumption $\int_{\mathcal{X}} M(x) dv(x) < \infty$ we see then this term converges to zero uniformly as $\varepsilon \rightarrow 0$.

Let $0 \leq \phi^\varepsilon \leq 1$ be a continuous, function which is one on the interval $[-\varepsilon,\varepsilon]$ but zero outside the interval $[-2\varepsilon,2\varepsilon]$. Considering the last term in (3.3)

$$|S_n(\theta,h) - S_n^\varepsilon(\theta,h)| \leq \frac{1}{n} \sum_{k=1}^n \phi^\varepsilon(u_2 + f_2(x_k, \theta^*)) b(u_k, x_k)$$

Once again, applying Theorem 1, (Gallant, 1982) the RHS of (3.4) converges uniformly to

$$\int_{\mathcal{X}} \int_{\mathbb{R}^2} b(u,x) \phi^\varepsilon(u_2 + f_2(x,\theta)) h^*(u) du dv(x) \tag{3.5}$$

and by using the same arguments given above this quantity is uniformly bounded by

$$4 \varepsilon \int M(x) dv(x) .$$

Therefore for $\varepsilon > 0$ there is an N such that for $n \geq N$

$$\begin{aligned} \sup_{(\theta,h) \in \Theta \times F} |S_n(\theta,h) - S_n^\varepsilon(\theta,h)| \\ \leq \varepsilon + 4 \varepsilon \int M(x) dv(x) . \end{aligned}$$

Combining the results for the separate terms in (3.3) proves the lemma. QED

Before giving the proof of Theorem 1.1 we need to introduce the following projection operator:

Let $P_K: W^{m,2} \rightarrow \bar{H}$ such that

$$P_k(h) = g \Leftrightarrow \|h-g\|_{W^{m,2}} = \min_{f \in H_k} \|f-h\|_{W^{m,2}}$$

For $h \in W^{m,2}$, $\|P_k(h) - h\|_{W^{m,2}}$ will be non-increasing in K and if $h \in \overline{\bigcup_{k=1}^{\infty} H_k}$ then

$$\|P_K(h) - h\|_{W^{m,2}} \rightarrow 0 \text{ a.s. } K \rightarrow \infty. \quad (3.6a)$$

Finally, we claim that if $h_n \rightarrow h$ in $W^{m,2}$ and $h \in \overline{\bigcup_{k=1}^{\infty} H_k}$ then

$$P_K(h_n) \rightarrow h \text{ as } n, K \rightarrow \infty. \quad (3.6b)$$

To see this we have

$$\begin{aligned} \|P_k(h_n) - h\|_{W^{m,2}} &\leq \min_{f \in H_k} \|f-h_n\|_{W^{m,2}} + \|h_n-h\|_{W^{m,2}} \\ &\leq \min_{f \in H_k} \|f-h\|_{W^{m,2}} + 2\|h_n-h\|_{W^{m,2}} \\ &\leq \|P_K(h)-h\|_{W^{m,2}} + 2\|h_n-h\|_{W^{m,2}} \end{aligned} \quad (3.7)$$

where both terms on the RHS of (3.6) will converge to zero.

Proof of Theorem 1.1

We first argue that $(\tilde{\theta}_n, \tilde{h}_n)$ are consistent estimates.

By Lemma 3.1 \bar{H} will be compact therefore, so is $\Theta \times \bar{H}$. Under the identifiability conditions A-C by Lemma 2.2 $S(\theta^*, h^*)$ is the unique maximum of $S(\theta, h)$ over $\Theta \times \bar{H}$ and by Lemma 3.3b S is continuous on $\Theta \times \bar{H}$. Finally, by Lemma 3.3 we can conclude that (3.2) holds. Having satisfied the assumptions of Lemma 3.1 we have:

$$\tilde{\theta}_n \rightarrow \theta^* \text{ a.s. in } \Theta \quad (3.8)$$

$$\tilde{h}_n \rightarrow h^* \text{ a.s. in } W^{m,2}(\mathbb{R}^2)$$

Now we show that $\hat{\theta}_n$ and \hat{h}_n are consistent.

By the definition of $(\hat{\theta}_n, \hat{h}_n)$

$$S_n(\hat{\theta}_n, \hat{h}_n) \geq S_n(\tilde{\theta}_n, P_{K_n}(\tilde{h}_n))$$

From the properties of P_K given above, the consistency of $(\tilde{\theta}_n, \tilde{h}_n)$ and Lemma 3.4.

$$\liminf S_n(\hat{\theta}_n, \hat{h}_n) \geq S(\theta^*, h^*)$$

Now, using arguments similar to those in the proof of Lemma 3.2, it is straightforward to show that $(\hat{\theta}_n, \hat{h}_n)$ must have a single limit point at (θ^*, h^*) .

QED

4. Consistency of $(\bar{\theta}_n, \bar{h}_n)$

Let \mathcal{P} denote the class of polynomials on

\mathbb{R}^2 and let

$$\mathcal{V} = \{f: f = p\phi, p \in \mathcal{P}\}$$

Lemma 4.1

If ϕ^2 has a moment generating function and $\mathcal{V} \subseteq W^{S,2}(\mathbb{R}^2)$ then

a) \mathcal{V} is dense in $W^{S,2}(\mathbb{R}^2)$

b) \mathcal{H} is contained in the closure of \mathcal{F} with respect to the norm for $W^{m,2}(\mathbb{R}^2)$

Proof

a) The first statement of the lemma is equivalent to the following condition:

$$\text{If } h \in W^{S,2} \text{ and } \langle f, h \rangle_{W^{S,2}} = 0 \tag{4.1}$$

for all $f \in \mathcal{V}$ then $h = 0$.

First assume that (4.1) holds for $h \in C_0^\infty$. Using the fact that the adjoint of D^P in L^2 for functions in C_0^∞ is $(-1)^{|P|} D^P$

$$\begin{aligned} 0 &= \langle f, h \rangle_{W^{S,2}} = \sum_{|P| \leq S} \langle D^P f, D^P h \rangle_{L^2} \\ &= \sum_{|P| \leq S} (-1)^{|P|} \langle f, D^P \cdot D^P h \rangle_{L^2} = \langle f, g \rangle_{L^2} \end{aligned}$$

where $g = \sum_{|P| \leq S} (-1)^{|P|} (D^P \cdot D^P)(h)$.

Setting $f = p\phi$ and rewriting (4.2) gives

$$(p, g/\phi)_{L^2(\mathbb{R}^2, \phi^2)} = 0 \quad \forall p \in \mathcal{P}$$

From Gallant (1980), we know that the polynomials are dense in the weighted L^2 space $L^2(\mathbb{R}^2, \phi^2)$. Thus $g/\phi = 0$ and since g has compact support and is continuous $g \equiv 0$.

Now $\langle h, g \rangle_{L^2(\mathbb{R}^2)} = 0$ and from (4.2) we have $\langle h, h \rangle_{W^{S,2}(\mathbb{R}^2)} = 0$ which implies $h \equiv 0$.

Thus we have demonstrated that (4.1) holds for all $h \in C_0^\infty(\mathbb{R}^2)$. Since $C_0^\infty(\mathbb{R}^2)$ is dense in $W^{S,2}(\mathbb{R}^2)$ by continuous extension (4.1) holds for all $h \in W^{S,2}$ and the result follows

b) Since $h \in H$ by assumption $\sqrt{h} \in W^{m+1,2}$. From part a) there is a sequence $\{f_n\} \in V$ such that (4.1b) $f_n \rightarrow \sqrt{h}$ in $W^{m+1,2}$.

It remains to show that $f_n^2 \rightarrow h$ in $W^{m,2}$. Repeated application of the chain rule yields the formula

$$\begin{aligned} \frac{\partial^{P_1}}{\partial U_1^{P_1}} \frac{\partial^{P_2}}{\partial U_2^{P_2}} (\alpha\beta) &= \sum_{i=0}^{P_1} \sum_{j=0}^{P_2} \binom{P_1}{i} \binom{P_2}{j} \frac{\partial^i}{\partial U_1^i} \frac{\partial^j}{\partial U_2^j} (\alpha) \cdot \frac{\partial^{P_1-i}}{\partial U_1^{P_1-i}} \frac{\partial^{P_2-j}}{\partial U_2^{P_2-j}} (\beta) \end{aligned} \quad (4.1c)$$

Or in operator notation:

$$D^P(\alpha\beta) = \sum_{0 \leq |q| \leq |p|} \binom{P_1}{q_1} \binom{P_2}{q_2} D^q \alpha D^{P-q} \beta$$

Since the binomial coefficients are bounded for $|p| \leq m$ there is a $c < \infty$ such that

$$|D^P(\alpha\beta)|^2 \leq c \sum_{0 \leq |q| \leq |p|} |D^q \alpha|^2 |D^{P-q} \beta|^2 \quad (4.1d)$$

$$\leq c(w_1)(w_2)$$

where $w_1 = \sum_{0 \leq |q| \leq |p|} |D^q(\alpha)|^2$, $w_2 = \sum_{0 \leq |q| \leq |p|} |D^q(\beta)|^2$

Thus

$$\begin{aligned} \|D^P(\alpha\beta)\|_{L^2} &\leq c \|w_1 w_2\|_{L^2} \\ &\leq c \|w_1^2\|_{L^2} \|w_2^2\|_{L^2} \end{aligned}$$

By expanding w_1^2 and applying the Cauchy Schwartz inequality to the cross products it is straightforward to verify that there is and $C' < \infty$ independent of w_1 such that

$$\|w_1^2\|_{L^2}^2 \leq C' \sum_{0 \leq |q| \leq |p|} \|D^q \alpha\|_{L^4}^4 \quad (4.1e)$$

or

$$\leq C' \|\alpha\|_{W^{|p|,4}}^4$$

Clearly 4.1e will also hold for w_2 and therefore it follows that there is and $M < \infty$ such that

$$\|\alpha\beta\|_{W^{m,2}}^2 \leq M \|\alpha\|_{W^{m,4}}^4 \|\beta\|_{W^{m,4}}^4 \quad (4.1f)$$

Let $\alpha = f_n + \sqrt{h}$ and $\beta = f_n - \sqrt{h}$. By Theorem 5.4 (5) (Adams, 1976) the imbedding $W^{m+1,2}(\mathbb{R}^2) \rightarrow W^{m,4}(\mathbb{R}^2)$ is continuous. Hence $\|f_n + \sqrt{h}\|_{W^{m,4}}^4$ will be bounded for sufficiently large n and $\|f_n - \sqrt{h}\|_{W^{m,4}}^4 \rightarrow 0$. Thus by 4.1f

$$\|\alpha\beta\|_{W^{m,2}}^2 = \|h - f_n^2\|_{W^{m,2}}^2 \rightarrow 0. \quad \text{QED}$$

Before proving Theorem 1.2 we need to give some properties of two projection operators.

Let $T_k: W^{m,2} \rightarrow F_k$ denote the projection operator such that

$$T_k(h) = g \Leftrightarrow g \in F_k, \quad \|g-h\|_{W^{m,2}} = \min_{f \in F_k} \|f-h\|_{W^{m,2}}$$

Now if we replace P_k by T_k and H_k by F_k in (3.6) - (3.7) the same relations will hold. In particular, if $h \in H \subseteq F$ and $h_n \rightarrow h$ then

$$T_k(h_n) \rightarrow h \quad \text{as } n, k \rightarrow \infty.$$

Also we will need the fact that Q (see (1.5)) is a continuous operator.

Suppose $\{h_n\} \subseteq W^{m,2}$ converges to $h \in W^{m,2}$. By Lemma 3.1 \bar{H} is compact and there is a subsequence $\{Q(h_{n_k})\}$ with a limit $\rho \in \bar{H}$. Since

$$\|Q(h_{n_k}) - h\|_{W^{m,2}} \leq \|Q(h) - h\|_{W^{m,2}} + 2\|h_{n_k} - h\|_{W^{m,2}}$$

it follows that for any $\varepsilon > 0$

$$\|\rho-h\|_{W^{m,2}} \leq \|Q(h)-h\|_{W^{m,2}} + \varepsilon$$

By definition $\|Q(h)-h\|_{W^{m,2}} \geq \|\rho-h\|_{W^{m,2}}$. Thus $\|Q(h)-h\|_{W^{m,2}} = \|\rho-h\|_{W^{m,2}}$ and by the uniqueness of the projection, $Q(h) = \rho$. Therefore, $\{Q(h_n)\}$ has only one limit point at $Q(h)$ and $Q(h_n) \rightarrow Q(h)$.

Proof of Theorem 1.2

From the definition of $\bar{\theta}_n, \bar{h}_n$

$$S_n(\bar{\theta}_n, (Q \cdot T_k)(\bar{h}_n)) \leq S_n(\bar{\theta}_n, \bar{h}_n) \quad (4.2)$$

Also, from section 3 we know $(\tilde{\theta}_n, \tilde{h}_n) \rightarrow (\theta^*, h^*)$ a.s.

Since Q and T_k are continuous, $(Q \circ T_k)$ will also be continuous and $(Q \circ T_k)(\tilde{h}_n) \rightarrow h^*$. Therefore, by Lemma 3.4 (4.2) implies

$$S(\theta^*, h^*) \leq \liminf S_n(\bar{\theta}_n, \bar{h}_n) .$$

Using arguments similar to those given in the proof of Lemma 3.2 one can show that $\{(\bar{\theta}_n, \bar{h}_n)\}$ must have a single limit point at θ^*, h^* .

QED

References

- Adams, R. A. (1975). Sobolev Spaces. Academic Press, New York.
- Gallant, A. R. (1980). "Explicit Estimators of Parametric Functions in Nonlinear Regression," JASA 75(369): 182-193.
- Gallant, A. R. (1982). "Nonlinear Statistical Methods, Chapter 3. A Unified Asymptotic Theory of Nonlinear Statistical Models," Institute of Statistics Mimeograph Series No. 1617, North Carolina State University.
- Rudin, W. (1974). Real and Complex Analysis. McGraw Hill, Inc., New York.