AISM (2006) 58: 379–387
DOI 10.1007/s10463-005-0021-9

Yannis G. Yatracos

On consistent statistical procedures in regression

Received: 17 December 2003 / Revised: 25 January 2005 / Published online: 2 February 2006

Abstract Necessary and sufficient conditions are provided for the existence of consistent statistical procedures in regression models with random predictors under various error assumptions.

Keywords Consistency · Orthogonality of measures · Random predictors · Regression

1 Introduction

Regression estimation is a central theme in statistics, and in one part of the related literature least squares estimates are proposed and sufficient conditions are provided which guarantee their weak or strong consistency and asymptotic normality. The sufficient conditions pertain to the structures of the errors, the design matrix and the parameter space.

In linear and non-linear parametric regression, conditions for consistency and asymptotic normality of the least squares estimates have been provided among others by Eicker (1963), Jennrich (1969), Malinvaud (1970), Lai and Robbins (1977), Lai et al. (1978), Richardson and Bhattacharyya (1986). In linear models, Drygas (1976) and Wu (1980) gave conditions to characterize consistent directions of least squares estimates, that were shown to be equivalent (Massam, 1987). In parametric non-linear regression, Wu (1981) provided asymptotic theory of the least squares estimates which included necessary conditions for consistent estimation. In most of these results, the predictors are assumed to be non-random. M-estimates in the regression context have been studied among others by Bustos (1982),

Y.G. Yatracos
Department of Statistics and Applied Probability, The National University of Singapore,
6 Science Drive 2, Singapore 117546, Singapore
E-mail: yatracos@stat.nus.edu.sg

Pötscher and Prucha (1986), Jurečkova (1989), Liese and Vajda (1994) and Koul (1996).

In infinite dimensional linear models, the least squares estimates fail often to be consistent due to the underlying structure of the problem, in particular either the behavior of the sequence of the regression predictors or the size of the parameter space. This fact motivated Li (1984) to provide necessary and sufficient conditions pertinent to the non-random predictors for consistent estimation of bounded linear functionals, and Van de Geer and Wegkamp (1996) to provide necessary and sufficient conditions for consistent estimation in nonparametric models with random predictors, based on δ-entropy conditions on the parameter space.

Numerous regression problems involve random predictors, and it is undoubtedly useful to obtain conditions of various nature that guarantee the existence of consistent statistical procedures. In this work, the results in Li (1984) for infinite dimensional linear models are extended to random design. Necessary and sufficient conditions are also provided for the existence of consistent tests of simple hypotheses in regression, in terms of the Hellinger distances of the conditional distributions of the response variables given the predictors at the hypotheses, when neither the errors nor the predictors are identically distributed. These conditions are also necessary for the existence of consistent estimates. When both the predictors and the errors are identically distributed, the errors are independent and have finite Fisher information, it is seen that Wu (1981) necessary condition for consistent estimation holds also for random predictors.

Central in the proofs are the classic results by Kakutani (1948), Kraft (1955) and Shepp (1965) that guarantee consistency in parametric and nonparametric problems using infinite product measures. Their use in regression is motivated by the fact that a regression problem can be seen conditionally on the values of the predictors as a combination of several density estimation problems (e.g. see Yatracos, 1988).

2 Notation, definitions and the tools

The existence of consistent tests for simple hypotheses is directly related to orthogonality of two probability measures on infinite product spaces. The Hellinger distance and the Hellinger affinity are used below to characterize orthogonality of probability measures.

Definition 2.1 *For densities f, g, with respect to a $\sigma-$finite measure λ on the space $\mathcal{X}$, their Hellinger distance $H(f, g)$ is defined as*

$$H^2(f, g) = \int_{\mathcal{X}} (f^{1/2}(x) - g^{1/2}(x))^2 \lambda(\mathrm{d}x).$$

The affinity of f, g is

$$\rho(f, g) = \int_{\mathcal{X}} f^{1/2}(x) g^{1/2}(x) \lambda(\mathrm{d}x).$$

For the Hellinger distance and the affinity it holds:

$$H^2(f,g) = 2(1-\rho(f,g)) \le 2, \quad \rho(f,g) = 1 - .5H^2(f,g); \tag{1}$$

$H^2(f,g) = 2$ if and only if $\rho(f,g) = 0$, that is, if f and g are orthogonal ($f(x)g(x) = 0, \lambda - a.s.$). In the sequel $\lambda(\mathrm{d}x)$ is the Lebesgue measure $\mathrm{d}x$. More about the properties of $H(f,g)$ and its use in Statistics can be found in Le Cam and Yang (1990).

For the random vectors $\mathbf{X}_1, \dots, \mathbf{X}_n, \dots$, it is assumed that $\mathbf{X}_i$ takes values in a measurable space $(\mathcal{X}_i, \mathcal{A}_i)$,

$$(\mathcal{X}^{(n)}, \mathcal{A}^{(n)}) = \prod_{i=1}^{n} (\mathcal{X}_i, \mathcal{A}_i), \quad (\mathcal{X}, \mathcal{A}) = \prod_{i=1}^{+\infty} (\mathcal{X}_i, \mathcal{A}_i).$$

$Q^{(n)}$ is the joint distribution of $(\mathbf{X}_1, \dots, \mathbf{X}_n)$, and $\{Q^{(n)},\ n = 1, 2, \dots\}$ are assumed to have a unique extension Q on $(\mathcal{X}, \mathcal{A})$. μ is a σ-finite measure on $(\mathcal{X}, \mathcal{A})$, $q = \mathrm{d}Q/\mathrm{d}\mu,\ Q << \mu$. Denote by $E_Q,\ E_q$ the expectation operator and by $E(.|.)$ the conditional expectation. Then, the finite dimensional projections of μ and q are respectively $\mu^{(n)} = E(\mu|\mathcal{A}^{(n)}),\ q^{(n)} = \mathrm{d}Q^{(n)}/\mathrm{d}\mu_n = E(q|\mathcal{A}^n)\ a.s.$.

Definition 2.2 *A sequence* $\{\phi_n\}, 0 \le \phi_n \le 1$, *of* $\mathcal{A}^{(n)}$*-measurable functions on* $\mathcal{X}$ *is consistent for testing between the sets* H_0 *and* H_1 *of probability measures on* $(\mathcal{X}, \mathcal{A})$ *if* $\lim_{n\longrightarrow\infty} E_P\phi_n = 0$ *for every* $P \in H_0$, *and* $\lim_{n\longrightarrow\infty} E_Q\phi_n = 1$ *for every* $Q \in H_1$.

Kakutani (1948) characterized equivalence and orthogonality of two infinite *direct* product measures in terms of Hellinger distances between their finite dimensional projections. Kraft (1955, p. 126, 128) provided a similar characterization for *any* two infinite dimensional distributions, and used it to derive conditions for the existence of consistent statistical procedures.

Proposition 2.1 (Kraft, 1955, Theorem 1, p. 128)
Let P *and* Q *be probability measures on* $(\mathcal{X}, \mathcal{A})$ *with finite dimensional projections* $p^{(n)}$ *and* $q^{(n)}$. *The following statements are equivalent.*

1. *P and Q are orthogonal.*
2. $lim_{n\longrightarrow\infty}\rho(p^{(n)}, q^{(n)}) = 0$.
3. *There exists a consistent test for* $H_0 = \{P\}$ *against* $H_1 = \{Q\}$.

Using Kakutani (1948) results, Shepp (1965) provided conditions for distinguishing a sequence of random variables from a translate of itself.

Proposition 2.2 (Shepp, 1965, Theorem 1, part (ii), p. 1108)
Let $\mathbf{W} = (W_1, \dots, W_n, \dots)$ *be an infinite vector of independent identically distributed (i.i.d.) random variables with common distribution* $F = F(\mathrm{d}w)$, *and* $\mathbf{a} = (a_1, \dots, a_n, \dots)$ *a numerical sequence. If* Q *and* $Q^{\mathbf{a}}$ *are the distributions of* $\mathbf{W}$ *and* $\mathbf{W} + \mathbf{a}$, *respectively, and* F *has finite Fisher information, then*

a) *Q and $Q^{\mathbf{a}}$ are orthogonal if* $\sum_{n=1}^{+\infty} a_n^2 = +\infty$, *and*
b) *Q and $Q^{\mathbf{a}}$ are equivalent if* $\sum_{n=1}^{+\infty} a_n^2 < +\infty$.

Li (1984) considered observations $Y_1, \ldots, Y_n$ at predetermined levels $\mathbf{x}_1, \ldots, \mathbf{x}_n$, from the linear model

$$y = \sum_{i=1}^{+\infty} x_i \theta_i + \epsilon = \langle \mathbf{x}, \theta \rangle + \epsilon = \theta(\mathbf{x}) + \epsilon; \tag{2}$$

the unknown $\theta = (\theta_1, \ldots, \theta_k, \ldots)$ and the *known* predictors $\mathbf{x} = (x_1, \ldots, x_k, \ldots)$ are both infinite dimensional vectors in $l^2 = \{(a_1, \ldots, a_k, \ldots) : \sum_{k=1}^{+\infty} a_k^2 < +\infty\}$; $\theta \in \Theta \subseteq l^2$, $\langle , \rangle$ denotes the inner products in l^2 and in R^k, $\theta(\mathbf{x})$ denotes $\langle \mathbf{x}, \theta \rangle$.

The error ϵ has:

i) Mean 0, variance σ^2 (that may be unknown), and
ii) Finite Fisher information;
iii) $\epsilon_1, \ldots, \epsilon_n, \ldots$ are *i.i.d.*.
Let Θ^* be the closed linear space generated by Θ. For $\delta > 0$, let

$$B(\delta) = \{\theta : \langle \theta, \theta \rangle \leq \delta, \ \theta \in \Theta^*\}$$

iv) Assume that Θ contains $B(\delta^*)$ for some $\delta^* > 0$.

The next theorem provides necessary and sufficient conditions on the sequence $\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots$ that will guarantee the existence of consistent estimates $\hat{T}_n$ of the bounded linear functional $T(\theta) = \sum_{i=1}^{+\infty} c_i \theta_i$; $(c_1, \ldots, c_n, \ldots) \in l^2$. $T(\theta)$ is then called consistently-estimable bounded linear (c.e.b.l.) functional.

Proposition 2.3 (Li, 1984, p. 603) *Under the assumptions* **i)–iv)**, *the following statements are equivalent for model (2):*

a) *T is a c.e.b.l. functional for $\theta \in \Theta$.*
b) *For any $\theta^* \in \Theta$, T is a c.e.b.l. functional when the parameter space is restricted to $\{0, \theta^*\}$.*
c) *For any $\theta \in \Theta$ such that $T(\theta) \neq 0$,*

$$\sum_{i=1}^{+\infty} \langle \mathbf{x}_i, \theta \rangle^2 = +\infty.$$

d) *For any $\delta > 0$,*

$$\liminf_{n \longrightarrow \infty} \left\{ \sum_{i=1}^{n} \langle \mathbf{x}_i, \theta \rangle^2 | \theta \in B(\delta), T(\theta) = 1 \right\} = +\infty.$$

e) *There exists a sequence of estimators $\{\hat{T}_n\}$ based on the first n observations, such that $\lim_{n \longrightarrow \infty} E[\hat{T}_n - T(\theta)]^2 = 0$, for any $\theta \in \Theta$.*

3 The results

The next proposition is an extension of Proposition 2.3. The necessary and sufficient conditions obtained for random sequences $\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots$ were intuitively expected.

Proposition 3.1 *Let* $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n), \ldots$ *be random vectors that follow model (2),* $Y_i \in \mathcal{Y}_i \subseteq R, \mathbf{X}_i \in \mathcal{X}_i,\ i = 1, \ldots, n, \ldots$.

Under the assumptions **i)–iv)***, and*

$$\mathbf{v})\ P[\mathbf{X}_i \in l^2] = 1,\ i = 1, \ldots, n, \ldots ,$$

the following statements are equivalent.

a) *T is a c.e.b.l. functional for $\theta^* \in \Theta$.*

b) *For any $\theta \in \Theta$, T is a c.e.b.l. functional when the parameter space is restricted to $\{0, \theta\}$.*

c) *For any $\theta \in \Theta$ such that $T(\theta) \neq 0$,*

$$\sum_{i=1}^{+\infty} \langle \mathbf{X}_i, \theta \rangle^2 = +\infty\ \ a.s.\,.$$

d) *For any $\delta > 0$,*

$$\liminf_{n \longrightarrow \infty} \left\{ \sum_{i=1}^{n} \langle \mathbf{X}_i, \theta \rangle^2 | \theta \in B(\delta), T(\theta) = 1 \right\} = +\infty\ \ a.s.\,.$$

Proof **a)** $\Longrightarrow$ **b)** It is clear.

b) $\Longrightarrow$ **c)** Let g_θ and $g_{\theta,\mathbf{X}}$ be respectively the joint densities of $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n), \ldots$ and $(Y_1, \ldots, Y_n, \ldots,)$ given $(\mathbf{X}_1, \ldots, \mathbf{X}_n, \ldots)$, and $g_\theta^{(n)}$, $g_{\theta,\mathbf{X}}^{(n)}$ the corresponding densities of the n-dimensional vectors.
By Proposition 2.1, the existence of a consistent test for the given hypotheses implies that

$$\lim_{n \longrightarrow \infty} H^2(g_\theta^{(n)}, g_0^{(n)}) = 2. \tag{3}$$

Then,

$$H^2(g_\theta^{(n)}, g_0^{(n)}) = E_{q^{(n)}} H^2(g_{\theta,\mathbf{X}}^{(n)}, g_{0,\mathbf{X}}^{(n)}) = E_q H^2(g_{\theta,\mathbf{X}}^{(n)}, g_{0,\mathbf{X}}^{(n)})$$

and (3) becomes

$$\lim_{n \longrightarrow \infty} E_q H^2(g_{\theta,\mathbf{X}}^{(n)}, g_{0,\mathbf{X}}^{(n)}) = 2. \tag{4}$$

From Proposition 2.1, for every $\mathbf{X}$-realization $\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots\}$

$$\lim_{n \longrightarrow \infty} H^2(g_{\theta,\mathbf{X}}^{(n)}, g_{0,\mathbf{X}}^{(n)}) = H^2(g_{\theta,\mathbf{X}}, g_{0,\mathbf{X}})$$

and since $H^2(f, g) \leq 2$, for all f, g, from the bounded convergence theorem (4) holds if and only if (denoted by *iff* in the sequel)

$$H^2(g_{\theta,\mathbf{X}}, g_{0,\mathbf{X}}) = 2 \; a.s.\,. \tag{5}$$

From assumption **(iii)**, with $\theta(\mathbf{x})$ denoting $\langle \mathbf{x}, \theta \rangle$, the conditional density of the Y's given the $\mathbf{X}$'s is

$$g_{\theta,\mathbf{X}} = \prod_{i=1}^{+\infty} p_{\theta(\mathbf{X}_i)}. \tag{6}$$

From Eq. (6), (**i**), (**ii**) and Proposition 2.2, Eq. (5) holds *iff*

$$\sum_{i=1}^{+\infty} \langle \mathbf{X}_i, \theta \rangle^2 = \sum_{i=1}^{+\infty} \theta^2(\mathbf{X}_i) = +\infty \;\; a.s..$$

c) $\Longrightarrow$ **d)** For each $\mathbf{X}$-realization that satisfies **c)**, use the equivalence of **c)** and **d)** in Proposition 2.3.
d) $\Longrightarrow$ **a)** For each $\mathbf{X}$-realization that satisfies **d)** and from the equivalence of **d)** and **e)** in Proposition 2.3, there is a least squares estimate $\hat{T}_n$ such that

$$\lim_{n \longrightarrow \infty} E(\hat{T}_n - T(\theta))^2 = 0, \text{ and thus,}$$

$$\lim_{n \longrightarrow \infty} P[|\hat{T}_n - T(\theta)| > \epsilon] = 0. \tag{7}$$

Using the bounded convergence theorem for the conditional probabilities $P[|\hat{T}_n - T(\theta)| > \epsilon | \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n, \ldots)$, Eq. (7) is obtained unconditionally.

Remark 3.1 In Proposition 3.1, the *i.i.d.* assumption **(iii)** was used to show that **b)** implies **c)**. If only **i)** holds, similar implications between **a)** and **d)** hold as those mentioned in Li (1984, p. 606, after Corollary 3.1), that is, **c)** $\Longrightarrow$ **d)** $\Longrightarrow$ **a)** $\Longrightarrow$ **b)**.

Consider now the nonparametric regression set-up

$$Y_i = f(X_i) + \epsilon_i, i = 1, \ldots, n; \tag{8}$$

the errors satisfy **i)–iii)**, the X_i's take values in [0, 1], $f \in W_2^m[0, 1]$,

$$W_2^m[0,1] = \{f : [0,1] \longrightarrow R,\ f^{(m-1)} \text{ is absolutely continuous on } [0,1],\ \int_0^1 f^{(m)}(t)^2 \mathrm{d}t < +\infty\};$$

$f^{(k)}$ is the k-th derivative of f, m is a positive integer.

$W_2^m[0, 1]$ is a Hilbert space equipped with the inner product

$$\langle\langle f, g\rangle\rangle = \sum_{k=0}^{m} \int_0^1 f^{(k)}(x) g^{(k)}(x) \mathrm{d}x.$$

Using the decomposition in Li (1984, p. 608), if $\{f_j;\ j \geq 1\}$ denotes a complete orthonormal system in $W_2^m[0, 1]$, model (8) can be written as the model (2)

$$Y_i = f(X_i) + \epsilon_i = \sum_{j=1}^{+\infty} w_{i,j}\theta_j + \epsilon_i =< \mathbf{W}_i, \theta > +\epsilon_i, \quad i = 1, \ldots, n; \qquad (9)$$

$\theta_j = \langle\langle f, f_j\rangle\rangle$ and $w_{i,j} = f_j(X_i)$. Assume also that (**iv**) holds for the obtained parameter space Θ, and that (**v**) holds for the $\mathbf{W_i}$'s. An application of Proposition 3.1 gives the next result.

Corollary 3.1 *In model (9), $T(.)$ is a c.e.b.l. functional if and only if*

$$\sum_{i=1}^{+\infty} f(X_i)^2 = +\infty \ a.s., \textit{ for any } f \in W_2^m[0,1] \textit{ such that } T(f) \neq 0.$$

The next proposition provides necessary conditions for the existence of consistent estimates in both parametric and nonparametric regression. Part **f)** extends Theorem 1 in (Wu, 1981, p. 503) for random predictors.

Proposition 3.2 *Let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ be random vectors that follow the regression model*

$$Y_i = f(\mathbf{X}_i, \theta) + \epsilon_i, \quad i = 1, \ldots, n, \qquad (10)$$

with θ (resp. f) the unknown parameter.

Assume that $Y_i \in \mathcal{Y}_i \subseteq R$, $\mathbf{X}_i \in \mathcal{X}_i$ and has marginal density q_i, the errors ϵ_i have mean 0 and variance $\sigma^2>0$ unknown, and are independent with densities p_i, $i = 1, \ldots, n$.

Then, the following statements are equivalent.

a) *There is a consistent test for the hypotheses $H_o = \{\theta\}$, $H_1 = \{\eta\}$ (resp. $H_o = \{f\}$, $H_1 = \{g\}$).*

b) $\sum_{i=1}^{+\infty} E_{q_i} H^2(p_{i,f(\mathbf{X}_i,\theta)}, p_{i,f(\mathbf{X}_i,\eta)}) = +\infty$ *(resp.* $\sum_{i=1}^{+\infty} E_{q_i} H^2(p_{i,f(\mathbf{X}_i,\theta)}, p_{i,g(\mathbf{X}_i,\theta)}) = +\infty$*).*

c) $H^2(\Pi_{i=1}^{+\infty} p_{i,f(\mathbf{X}_i,\theta)}, \Pi_{i=1}^{+\infty} p_{i,f(\mathbf{X}_i,\eta)}) = 2\ a.s.$ *(resp.* $H^2(\Pi_{i=1}^{+\infty} p_{i,f(\mathbf{X}_i,\theta)}, \Pi_{i=1}^{+\infty} p_{i,g(\mathbf{X}_i,\theta)}) = 2\ a.s.$*).*

If $\mathcal{X}_i = \mathcal{X}_1$, $q_i = q_1$, the errors are identically distributed $p_i = p$, $i = 1, \ldots, n, \ldots$, then **a)–c)** *are equivalent to*

d) $E_{q_1} H^2(p_{f(\mathbf{X}_1,\theta)}, p_{f(\mathbf{X}_1,\eta)}) > 0$ *(resp.* $E_{q_1} H^2(p_{f(\mathbf{X}_1,\theta)}, p_{g(\mathbf{X}_1,\theta)}) > 0$*).*

e) $\sum_{i=1}^{+\infty} H^2(p_{f(\mathbf{X}_i,\theta)}, p_{f(\mathbf{X}_i,\eta)}) = +\infty\ a.s.$ *(resp.* $\sum_{i=1}^{+\infty} H^2(p_{f(\mathbf{X}_i,\theta)}, p_{g(\mathbf{X}_i,\theta)}) = +\infty\ a.s.$*).*

If the errors have in addition finite Fisher information, **a)–e)** *are equivalent to*

f) $\sum_{i=1}^{+\infty} [f(\mathbf{X}_i,\theta) - f(\mathbf{X}_i,\eta)]^2 = +\infty$ *(resp.* $\sum_{i=1}^{+\infty} [f(\mathbf{X}_i,\theta) - g(\mathbf{X}_i,\theta)]^2 = +\infty$*).*

Proof It is enough to provide the proof for $H_o = \{\theta\},\ H_1 = \{\eta\}$.

a) $\Longleftrightarrow$ **b)** Note that from Eq. (1)

$$H^2(g_\theta^n, g_\eta^n) = 2\left[1 - \prod_{i=1}^{n} \int_{\mathcal{X}_i}\int_{\mathcal{Y}_i} \sqrt{p_{f(\mathbf{x}_i,\theta)}(y_i),\, p_{f(\mathbf{x}_i,\eta)}(y_i)}\, q_i(\mathbf{x}_i)\mathrm{d}y_i \mathrm{d}\mathbf{x}_i\right]$$
$$= 2\left[1 - \prod_{i=1}^{n}[1 - 0.5E_{q_i} H^2(p_{f(\mathbf{X}_i,\theta)}, p_{f(\mathbf{X}_i,\eta)})]\right].$$

Then, **a)** holds by Proposition 2.1 *iff*

$$\lim_{n\to+\infty} H^2(g_\theta^n, g_\eta^n) = 2 \Longleftrightarrow \lim_{n\to+\infty} \prod_{i=1}^{n}[1 - 0.5E_{q_i} H^2(p_{i,f(\mathbf{X}_i,\theta)}, p_{i,f(\mathbf{X}_i,\eta)})] = 0,$$

which holds *iff*

$$\lim_{n\to+\infty} \sum_{i=1}^{n} E_{q_i} H^2(p_{i,f(\mathbf{X}_i,\theta)}, p_{i,f(\mathbf{X}_i,\eta)}) = +\infty. \tag{11}$$

a) $\Longleftrightarrow$ **c)** As in the proof **b)** $\Longrightarrow$ **c)** in Proposition 1,

$$\lim_{n\to+\infty} H^2(g_\theta^n, g_\eta^n) = 2 \Longleftrightarrow H^2(\prod_{i=1}^{+\infty} p_{f(\mathbf{x}_i,\theta)}, \prod_{i=1}^{+\infty} p_{f(\mathbf{x}_i,\eta)}) = 2\ a.s..$$

b) $\Longleftrightarrow$ **d)** If $q_i = q_1,\ p_i = p,\ i = 1, \ldots, n$, Eq. (11) holds *iff*

$$\lim_{n\to+\infty} n\ E_{q_1} H^2(p_{f(\mathbf{X}_1,\theta)}, p_{f(\mathbf{X}_1,\eta)}) = +\infty, \text{ or } iff$$

$$E_{q_1} H^2(p_{f(\mathbf{X}_1,\theta)}, p_{f(\mathbf{X}_1,\eta)}) > 0.$$

d) $\Longleftrightarrow$ **e)** If $E_{q_1} H^2(p_{f(\mathbf{X}_1,\theta)}, p_{f(\mathbf{X}_1,\eta)}) > 0$ then $\sum_{i=1}^{+\infty} H^2(p_{f(\mathbf{X}_i,\theta)}, p_{f(\mathbf{X}_i,\eta)}) = +\infty\ a.s.$. The converse also holds since $\sum_{i=1}^{+\infty} H^2(p_{f(\mathbf{X}_i,\theta)}, p_{f(\mathbf{X}_i,\eta)}) = +\infty\ a.s.$ and $E_{q_1} H^2(p_{f(\mathbf{X}_1,\theta)}, p_{f(\mathbf{X}_1,\eta)}) = 0$ lead to the contradiction $\sum_{i=1}^{+\infty} H^2(p_{f(\mathbf{X}_i,\theta)}, p_{f(\mathbf{X}_i,\eta)}) = 0\ a.s.$.

c) $\Longleftrightarrow$ **f)** Follows from Proposition 2.2, as in the proof in Proposition 3.1.

Acknowledgements Many thanks are due to the Associate Editor and the referees for useful suggestions that served to improve the presentation of this paper.

References

Bustos, O. H. (1982). General M-estimates for contaminated pth-order autoregressive processes: consistency and asymptotic normality. Robustness in autoregressive processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *59*, 491–504.

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics*, *34*, 447–456.

Drygas, H. (1976). Weak and strong consistency of the least squares estimates in regression models. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, *34*, 119–127.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, *40*, 633–643.

Jurečkova, J. (1989). Consistency of M-estimators in a linear model, generated by non-monotone and discontinuous ψ-functions. *Probability and Mathematical Statistics*, *10*, 1–10.

Kakutani, S. (1948). On equivalence of infinite product measures. *Annals of Mathematics*, *49*, 214–224.

Koul, H. L. (1996). Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Annals of Statistics*, *24*, 380–404.

Kraft, C. (1955). Some conditions for consistency and uniform consistency of statistical procedures. *University of California Publications in Statistics*, *2*, 125–142.

Lai, T. L., Robbins, H., & Wei, C. Z. (1978). Strong consistency of least-squares estimates in multiple regression. *Proceedings of the National Academy of Sciences of the United States of America*, *75*, 3034–3036.

Lai, T. L., & Robbins, H. (1977). Strong consistency of least-squares estimates in regression models. *Proceedings of the National Academy of Sciences of the United States of America*, *74*, 2667–2669.

Le Cam L., & Yang G. (1990). *Asymptotics in Statistics: Some Basic Concepts.* New York: Springer-Verlag.

Li, K. -C. (1984). Regression models with infinitely many parameters: consistency of bounded linear functionals. *Annals of Statistics*, *12*, 601–611.

Liese, F., & Vajda, I. (1994). Consistency of M-estimates in general regression models. *Journal of Multivariate Analysis*, *50*, 93–114.

Malinvaud, E. (1970). The consistency of nonlinear regressions. *Annals of Mathematical Statistics*, *41*, 956–969.

Massam, H. (1987). Consistent directions for least squares estimates. *The Canadian Journal of Statistics*, *15*, 87–90.

Pötscher, B. M., & Prucha, I. R. (1986). A class of partially adaptive one-step M-estimators for the nonlinear regression model with dependent observations. *Journal of Econometrics*, *32*, 219–251.

Richardson, G. D., & Bhattacharyya, B. B. (1986). Consistent estimators in nonlinear regression for a noncompact parameter space. *Annals of Statistics*, *14*, 1591–1596.

Shepp, L. A. (1965). Distinguishing a sequence of random variables from a translate of itself. *Annals of Mathematical Statistics*, *36*, 1107–1112.

Van de Geer, S., & Wegkamp, M. (1996). Consistency for the least squares estimator in nonparametric regression. *Annals of Statistics*, *24*, 2513–2523.

Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, *9*, 501–513.

Wu, C. F. (1980). Characterizing the consistent directions of least squares estimates. *Annals of Statistics*, *8*, 789–801.

Yatracos, Y. G. (1988). A lower bound on the error in nonparametric regression type problems. *Annals of Statistics*, *16*, 1180–1187.