The Annals of Statistics
1984, Vol. 12, No. 2, 601–611

REGRESSION MODELS WITH INFINITELY MANY PARAMETERS: CONSISTENCY OF BOUNDED LINEAR FUNCTIONALS[1]

BY KER-CHAU LI

Purdue University

Consider a linear model with infinitely many parameters given by $y = \sum_{i=1}^{\infty} x_i\theta_i + \varepsilon$ where $\mathbf{x} = (x_1, x_2, \cdots)'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)'$ are infinite dimensional vectors such that $\sum_{i=1}^{\infty} x_i^2 < \infty$ and $\sum_{i=1}^{\infty} \theta_i^2 < \infty$. Suppose independent observations $y_1, y_2, \cdots, y_n$ are observed at levels $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$. Under suitable conditions about the error distribution, the set of all bounded linear functionals $T(\boldsymbol{\theta})$ for which there exists an estimate $\hat{T}_n$ such that $\hat{T}_n \to T(\boldsymbol{\theta})$ in probability will be characterized. An application will be extended to the nonparametric regression problem where the response curve f is smooth on the interval [0, 1] in the sense that f has an $(m-1)$th derivative that is absolutely continuous and $\int_0^1 f^{(m)}(t)^2\,dt < \infty$.

1. Introduction. Consider a linear model given by

$$y = \sum_{i=1}^{\infty} x_i\theta_i + \varepsilon = \langle \mathbf{x}, \boldsymbol{\theta} \rangle + \varepsilon, \tag{1.1}$$

where the unknown $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)'$ and the known $\mathbf{x} = (x_1, x_2, \cdots)'$ are infinite dimensional vectors in the Hilbert space $\ell^2 = \{(a_1, a_2, \cdots)' \mid \sum_{i=1}^{\infty} a_i^2 < \infty\}$; $\langle \cdot, \cdot \rangle$ denotes the inner product of ℓ^2; ε is the random error satisfying certain conditions to be specified later. Let $\Theta \subset \ell^2$ be the parameter space. (1.1) extends the scope of the usual linear model where Θ is typically a finite dimensional space.

Suppose we are interested in estimating $T(\boldsymbol{\theta}) = \sum_{i=1}^{\infty} c_i\theta_i$, for some c_i, $i = 1, 2, \cdots$, with $\sum_{i=1}^{\infty} c_i^2 < \infty$. In standard terminology, $T(\cdot)$ is called a bounded linear functional on ℓ^2. Assume that independent observations $y_1, y_2, \cdots, y_n$ are observed at levels $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$. Let $\hat{T}_n$ be an estimate for $T(\boldsymbol{\theta})$. From the asymptotic viewpoint, a minimum requirement for reasonable $\hat{T}_n$ to possess seems to be $\hat{T}_n \to \hat{T}(\boldsymbol{\theta})$ in probability. However, $\hat{T}_n$ may fail to be consistent either because it is a poor estimate or because the underlying structure of the problem (particularly, the behavior of $\mathbf{x}_i$ sequence) does not admit any consistent estimates at all. Thus before checking the consistency of any proposed estimate, one should first examine the intrinsic structure of the given asymptotic setup. For this purpose, it is desired to obtain a necessary and sufficient condition on the $\mathbf{x}_i$ sequence to guarantee the existence of a consistent $\hat{T}_n$ for $T(\boldsymbol{\theta})$. In Section 2, such conditions will be derived (Theorem 2.1). Under the given model (1.1), we call $T(\cdot)$ a consistently-estimable bounded linear (c.e.b.l. hereafter) functional

Received April 1982; revised March 1983.
[1] Research supported by NSF Grant No. MCS-82-00631.
AMS 1980 *subject classification.* Primary 62G05; secondary 62J05.
Key words and phrases. Asymptotic consistency, consistent region of degree k, Fisher's information, Hilbert space, limiting point of degree k, nonparametric regression.

for Θ if there exists some consistent estimate $\hat{T}_n$ for $T(\theta)$, $\theta \in \Theta$. Our result shows that $T(\cdot)$ is c.e.b.l. if and only if $T(\theta) = 0$ for any $\theta \in \Theta$ such that $\sum_{i=1}^{\infty} \langle \mathbf{x}_i, \theta \rangle^2 < \infty$. Further discussion on Theorem 2.1 will be provided in Section 3.

Section 4 is devoted to applying these results to solve a problem in nonparametric regression. Suppose that the response curve f is $m - 1$ time continuously differentiable on [0, 1] with $f^{(m-1)}$ absolutely continuous and $\int_0^1 f^{(m)}(t)^2 \, dt < \infty$. Independent observations $y_i = f(t_i) + \varepsilon_i$, $i = 1, 2, \cdots, n$, are made and random errors ε_i's are i.i.d. with the common distribution possessing a finite Fisher's information and a finite variance. The goal is to characterize the set of all c.e.b.l. functionals. In particular, we would like to obtain conditions to determine whether the kth derivative of f at a point t is c.e.b.l. or not. The limiting behavior of the sequence of design points $\{t_1, t_2, \cdots\}$ is crucial here. We call a point t^* *a limiting point of degree k* for the sequence $\{t_i\}$ if there exists a subsequence $\{t_i'\}$ of $\{t_i\}$ such that $t_i' \to t^*$ as $i \to \infty$ and $\sum_{i=1}^{\infty} (t_i' - t^*)^{2k} = \infty$. When $k = 0$, this is exactly the usual definition of limiting points. Our result shows that for $0 \le k \le m - 1$, $f^{(k)}(t^*)$ is c.e.b.l. if and only if t^* is a limiting point of degree k. This brings up a connection with a result on polynomial regression obtained by Wu (1980).

When Θ is finite-dimensional (i.e., under the usual linear model), consistency for least squares estimates has been studied by Drygas (1976) and Wu (1980); the latter is more relevant to our work here. Wu showed that (2.4) in Theorem 2.1 was necessary and sufficient for $\hat{T}_n \to T(\theta)$, where $\hat{T}_n$ is the least squares estimate for $\mathbf{T}(\theta)$. Note that for the usual linear model, using the least squares estimate has been a common practice. But in the case that Θ is infinite dimensional, *it is typical that the least squares estimate no longer works*: it can fail to be consistent easily (for examples, see Section 4). Thus it is interesting to observe that for the finite dimensional Θ, if the least squares estimate is inconsistent, then no other estimates can be consistent. For the nonlinear regression models (with finitely many parameters), a necessary condition for a given setup to admit a consistent estimate was obtained by Wu (1981).

2. Main results. In this section, the notation $\langle \cdot, \cdot \rangle$ will be used to denote either the inner product in ℓ^2 or the inner product in R^n without ambiguity.

Consider the linear regression model with infinitely many parameters given by (1.1). Since the results we shall derive here may also be applicable to the case where the parameter space Θ is not the entire ℓ^2 (e.g., $\Theta = \{\theta \mid \langle \theta, \theta \rangle \le \delta^2, \theta \in \ell^2\}$ for some known real number δ) and to the case where the usual linear model with finitely many parameters is considered, a suitable condition on the parameter space will be given as follows. Let Θ^* be the closed linear space generated by Θ. For $\delta > 0$, let $B(\delta) = \{\theta \mid \langle \theta, \theta \rangle \le \delta, \theta \in \Theta^*\}$. Assume that

$$(2.1) \qquad \Theta \text{ contains } B(\delta^*) \text{ for some } \delta^* > 0.$$

The probability distribution of the random error ε is assumed to satisfy the following two conditions:

$$(2.2) \quad E\varepsilon = 0 \text{ and } 0 < \operatorname{Var} \varepsilon = \sigma^2 < \infty \quad (\sigma \text{ may or may not be known}),$$

and

(2.3) ε has a finite Fisher's information (i.e., ε has a density f which is positive (a.e.), absolutely continuous and $\int_{-\infty}^{\infty}(f'(x)^2/f(x))\,dx < \infty$).

We now present the main theorem of this paper. The convention that the "inf" of an empty set is $+\infty$ will be adopted. $\mathbf{0}$ denotes $(0, 0, \cdots)'$.

THEOREM 2.1. *Assume that* (2.1) ~ (2.3) *hold. For the regression model* (1.1), *the following statements are equivalent*:

(i) $T(\cdot)$ *is a* c.e.b.l. *functional for* $\boldsymbol{\theta} \in \Theta$.

(ii) (*Pairwise consistency*). *For any* $\boldsymbol{\theta}^* \in \Theta$, $T(\cdot)$ *is a* c.e.b.l. *functional when the parameter space is restricted to* $\{\mathbf{0}, \boldsymbol{\theta}^*\}$.

(iii) *For any* $\boldsymbol{\theta} \in \Theta$ *such that* $T(\boldsymbol{\theta}) \neq 0$,

$$\sum_{i=1}^{\infty} \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle^2 = \infty. \tag{2.4}$$

(iv) *For any* $\delta > 0$,

$$\liminf_{n\to\infty}\{\sum_{i=1}^{n} \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle^2 \mid \boldsymbol{\theta} \in B(\delta),\ T(\boldsymbol{\theta}) = 1\} = \infty. \tag{2.5}$$

(v) *There exists a sequence of estimators* $\{\hat{T}_n\}$, *where* $\hat{T}_n$ *is based on the first* n *observations, such that* $E(\hat{T}_n - T(\boldsymbol{\theta}))^2 \to 0$, *as* $n \to \infty$, *for any* $\boldsymbol{\theta} \in \Theta$.

PROOF.

"(i) $\Rightarrow$ (ii)" holds obviously.

"(ii) $\Rightarrow$ (iii)". Suppose there exists a $\boldsymbol{\theta}^* \in \Theta$ such that $T(\boldsymbol{\theta}^*) \neq 0$ but (2.4) does not hold. Let $h_i = \langle \mathbf{x}_i, \boldsymbol{\theta}^* \rangle$. Then $\sum_{i=1}^{\infty} h_i^2 < \infty$. Let $\mathbf{P}_n$ and $\mathbf{Q}_n$ be the probability measure of $(y_1, \cdots, y_n)'$ when $\boldsymbol{\theta} = \mathbf{0}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ respectively. Consider the case where the parameter space is restricted to $\{\boldsymbol{\theta}, \boldsymbol{\theta}^*\}$. It is clear that (ii) implies that $\mathbf{P}_n$ and $\mathbf{Q}_n$ are asymptotically mutually-singular. However, this is contradictory to Theorem 1 of Shepp (1965) (for a different proof, see the Appendix of Li, 1982). Note that a similar argument was used in Wu (1981).

"(iii) $\Rightarrow$ (iv)". Since the "inf" of an empty set is $+\infty$, we may consider those δ such that $\{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in B(\delta),\ T(\boldsymbol{\theta}) = 1\} \neq \phi$ only. For such a δ and any $c > 0$, define

$$A_n = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in B(\delta),\ T(\boldsymbol{\theta}) = 1, \quad \text{and} \quad \sum_{i=1}^{n} \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle^2 \le c\}.$$

Since $A_n \subset A_{n-1}$, it suffices to show that $A_n = \phi$ for some n. Observe that due to (2.1), (2.4) holds for any $\boldsymbol{\theta} \in B(\delta^*)$ with $T(\boldsymbol{\theta}) \neq 0$. This in turn implies that (2.4) is satisfied by any $\boldsymbol{\theta} \in B(\delta)$ with $T(\boldsymbol{\theta}) \neq 0$. Now we have

$$\cap_{n=1}^{\infty} A_n = \phi; \tag{2.6}$$

otherwise, for any $\boldsymbol{\theta} \in \cap_{n=1}^{\infty} A_n$, we get $\sum_{i=1}^{n} \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle^2 \le c < \infty$ for any n, a contradiction to (2.4). Now, consider the weak topology on the space Θ^* (since Θ^* is a Hilbert space, weak topology and weak* topology are identical). By Alaoglu's Theorem (c.f. Royden 1972, page 202), $B(\delta)$ is weakly compact. Since

$T(\cdot)$ and $\langle \mathbf{x}_i, \cdot \rangle$ are weakly continuous, it is clear that A_n is weakly closed. Moreover, since $A_n \subset B(\delta)$, A_n is also weakly compact. From compactness, (2.6) implies that there exists some N such that $\bigcap_{n=1}^{N} A_n = \phi$.

Since $A_n \subset A_{n-1}$, we get $A_N = \phi$. Thus the proof is complete.

"(iv) $\Rightarrow$ (v)". Let ν be the element in Θ^* such that $\langle \nu, \theta \rangle = T(\theta)$ for any $\theta \in \Theta^*$ (such a ν exists because of Riesz representation theorem). Without loss of generality, we may assume that $\langle \nu, \nu \rangle = 1$ and $\mathbf{x}_i \in \Theta^*$, $i = 1, 2, \cdots$. Before constructing the consistent estimators, we shall make a suitable orthogonal transformation on the vector of observations $(y_1, \cdots y_n)'$ and choose a convenient complete orthonormal system on Θ^* so that the design matrix takes the form

$$\left(\begin{array}{c|cccccc|c} x_{10} & x_{11} & 0 & \cdot & \cdot & 0 & & 0 \cdots \\ x_{20} & 0 & x_{22} & 0 & \cdot & \cdot & & 0 \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \cdots \\ x_{n0} & 0 & \cdot & \cdot & \cdot & x_{nn} & & 0 \cdots \end{array}\right) \tag{2.7}$$

and the first coordinate of the new parameter is what we want to estimate. To carry out this idea, let $\mathbf{V}_n$ be the vector space generated by $\{\nu, \mathbf{x}_1, \cdots, \mathbf{x}_n\}$ and let $\mathbf{U}_n = \{\mathbf{u} \mid \langle \mathbf{u}, \nu \rangle = 0 \text{ and } \mathbf{u} \in \mathbf{V}_n\}$. Consider the linear transformation L from $\mathbf{U}_n$ to R^n defined by mapping $\mathbf{u}$ to $(\langle \mathbf{x}_1, \mathbf{u} \rangle, \cdots \langle \mathbf{x}_n, \mathbf{u} \rangle)'$. It is a well-known fact in linear algebra that there exists an orthonormal basis $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ in $\mathbf{U}_n$ and an orthonormal basis $\{\mathbf{g}_1, \cdots, \mathbf{g}_n\}$ in R^n such that $L(\mathbf{e}_i) = m_i \mathbf{g}_i$ for some nonnegative number m_i, $i = 1, \cdots, n$ (m_i may be taken as the square roots of the eigenvalues of $L'L$). We extend the orthonormal basis $\{\nu, \mathbf{e}_1, \cdots, \mathbf{e}_n\}$ in $\mathbf{V}_n$ to a complete orthonormal system in Θ^* by adding an arbitrary complete orthonormal system in the orthogonal complement of $\mathbf{V}_n$ in Θ^* to the set $\{\nu, \mathbf{e}_1, \cdots, \mathbf{e}_n\}$. Write $\mathbf{y}^{(n)} = (y_1, \cdots, y_n)'$, $\varepsilon^{(n)} = (\varepsilon_1, \cdots, \varepsilon_n)'$, $\mathbf{A}^{(n)} = (\langle \mathbf{x}_1, \nu \rangle, \cdots, \langle \mathbf{x}_n, \nu \rangle)'$; let $Z_i = \langle \mathbf{y}^{(n)}, \mathbf{g}_i \rangle$ and $\varepsilon_i' = \langle \varepsilon^{(n)}, \mathbf{g}_i \rangle$. Now, it is clear that

$$Z_i = \langle A^{(n)}, \mathbf{g}_i \rangle \langle \nu, \theta \rangle + m_i \langle \mathbf{e}_i, \theta \rangle + \varepsilon_i', \quad i = 1, \cdots, n, \tag{2.8}$$

and the random errors ε_i' are uncorrelated with the common variance σ^2. Thus for $(Z_1, \cdots, Z_n)'$ and the complete orthonormal system $\{\nu, \mathbf{e}_1, \cdots\}$, the design matrix is of the form (2.7) with $x_{i0} = \langle \mathbf{A}^{(n)}, \mathbf{g}_i \rangle$ and $x_{ii} = m_i$. Note that to be precise we should have used the notations $x_{i0}^{(n)}$ and $x_{ii}^{(n)}$ instead of x_{i0} and x_{ii} (and $\mathbf{e}_i^{(n)}$ instead of $\mathbf{e}_i$) because the transformation L depends on n. However we omit the superscript (n) to avoid complexity of notation.

Now, construct the estimator $\hat{T}_n$ by setting

$$\hat{T}_n = \frac{\sum_{i=1}^n x_{i0} b_i^{-2} Z_i}{\sum_{i=1}^n x_{i0}^2 b_i^{-2}}, \tag{2.9}$$

where $b_i = \max\{x_{ii}, 1\}$.

To establish (v), it suffices to show that $E\hat{T}_n - T(\theta) \to 0$ and $\operatorname{Var} \hat{T}_n \to 0$. Since Z_i are uncorrelated,

$$\operatorname{Var} \hat{T}_n = \sigma^2 \left(\textstyle\sum_{i=1}^n x_{i0}^2 b_i^{-4}\right) / \left(\textstyle\sum_{i=1}^n x_{i0}^2 b_i^{-2}\right)^2 \leq \sigma^2 / \textstyle\sum_{i=1}^n x_{i0}^2 b_i^{-2}.$$

We now proceed to show that

$$(2.10)\qquad \textstyle\sum_{i=1}^{n} x_{i0}^2 b_i^{-2} \to \infty \quad \text{as} \quad n \to \infty$$

in order to get Var $\hat{T}_n \to 0$.

Let $I_n = \{i \mid i \le n, x_{ii} < 1\}$. Write

$$\textstyle\sum_{i=1}^{n} x_{i0}^2 b_i^{-2} = \sum_{i \in I_n} x_{i0}^2 + \sum_{i \notin I_n} x_{i0}^2 x_{ii}^{-2}.$$

Suppose (2.10) does not hold. Then there exists some positive number M such that

$$(2.11)\qquad \textstyle\sum_{i \in I_n} x_{i0}^2 \le M,$$

$$(2.12)\qquad \textstyle\sum_{i \notin I_n} x_{i0}^2 x_{ii}^{-2} \le M,$$

for n in an infinite subset of positive integers. Let $\omega^{(n)} = \nu - \sum_{i \notin I_n} x_{i0} x_{ii}^{-1} \mathbf{e}_i$. Because of (2.12) it is clear that $\omega^{(n)} \in \{\theta \mid T(\theta) = 1 \text{ and } \theta \in B(1 + M)\}$. From the definition of L and $\mathbf{A}^{(n)}$, we have

$$\begin{aligned}\inf\{\textstyle\sum_{i=1}^{n} \langle \mathbf{x}_i, \theta \rangle^2 \mid \theta \in B(1 + M), T(\theta) = 1\} &\le \textstyle\sum_{i=1}^{n} \langle \mathbf{x}_i, \omega^{(n)} \rangle^2 \\ &= \| \mathrm{L}(\omega^{(n)} - \nu) + \mathbf{A}^{(n)} \|^2,\end{aligned}$$

where $\| \cdot \|$ is the Euclidean norm in R^n. Since $\{\mathbf{g}_i\}$ is an orthonormal basis, the last expression equals

$$\textstyle\sum_{i=1}^{n} \langle \mathbf{g}_i, L(\omega^{(n)} - \nu) + \mathbf{A}^{(n)} \rangle^2,$$

which, since $L(\mathbf{e}_j) = x_{jj} \mathbf{g}_j$, in turn equals

$$\begin{aligned}&\textstyle\sum_{i=1}^{n} \langle \mathbf{g}_i, -\sum_{j \notin I_n} (x_{j0}/x_{jj}) x_{jj} \mathbf{g}_j + \mathbf{A}^{(n)} \rangle^2 \\ &\qquad = \textstyle\sum_{i \in I_n} \langle \mathbf{g}_i, \mathbf{A}^{(n)} \rangle^2 + \sum_{i \notin I_n} (\langle \mathbf{g}_i, \mathbf{A}^{(n)} \rangle - x_{i0})^2 \\ &\qquad = \textstyle\sum_{i \in I_n} x_{i0}^2.\end{aligned}$$

The last equality follows from the definition of x_{i0}. Now by (2.11), this is contradictory to (2.5) for $\delta = M + 1$. Hence (2.10) holds and Var $\hat{T}_n \to 0$. It remains to show that $E\hat{T}_n - T(\theta) \to 0$ for any $\theta \in \Theta$. For this purpose, it suffices to verify that

$$(2.13)\qquad \frac{\sum_{i=1}^{n} x_{i0} b_i^{-2} x_{ii} \langle \mathbf{e}_i, \theta \rangle}{\sum_{i=1}^{n} x_{i0}^2 b_i^{-2}} \to 0, \quad \text{as} \quad n \to \infty.$$

Now, by the Cauchy-Schwartz inequality, the numerator of (2.13) does not exceed

$$(\textstyle\sum_{i=1}^{n} x_{i0}^2 b_i^{-2})^{1/2} \cdot (\sum_{i=1}^{n} x_i^2 b_i^{-2} \langle e_i, \theta \rangle^2)^{1/2} \le (\sum_{i=1}^{n} \mathrm{x}_{i0}^2 \mathrm{b}_i^{-2})^{1/2} \| \theta \|.$$

Therefore by (2.10), (2.13) holds and consequently $\hat{T}_n$ is consistent. The proof for "(iv) $\Rightarrow$ (v)" is now complete. Finally, "(v) $\Rightarrow$ (i)" holds obviously. □

3. Discussion. Several important features about Theorem 2.1 are now in order. First, "(ii) $\Rightarrow$ (i)" means "pairwise consistency implies consistency", which certainly may not be true in other contexts. In fact, the following example demonstrates what may happen without the structure given by (1.1).

EXAMPLE 1. Suppose $\Theta = \{(\theta_1, \cdots, \theta_n, \cdots) \mid \sum_{i=1}^{\infty} \theta_i^2 = \infty\} \cup \{(0, 0, \cdots)\}$. The observations y_i satisfy the model $y_i = \theta_i + \varepsilon_i$, $i = 1, 2, \cdots$, where ε_i are i.i.d. normal with mean 0 and variance σ^2. Suppose we want to estimate θ_1. Without much difficulty, it can be verified that when the parameter space is restricted to two points $\{(0, 0, \cdots), (\theta_1^*, \cdots, \theta_n^*, \cdots)\}$, then consistent estimates exist. But, it is also clear that when the parameter space is the whole Θ, θ_1 is not consistently estimable.

The next important feature about Theorem 2.1 concerns the statement (iii). By the equivalence between (iii) and (i), the consistency problem (which is stochastic in nature and therefore is relatively complicated) can now be reduced to verifying the deterministic equation (2.4).

A useful consequence is given in the following:

COROLLARY 3.1. *The set of all* c.e.b.l. *functionals is a closed linear space.*

PROOF. Let $\Theta' = \{\boldsymbol{\theta}: \boldsymbol{\theta} \in \Theta \text{ and } \sum_{i=1}^{\infty} \langle \mathbf{x}_i, \boldsymbol{\theta} \rangle^2 < \infty\}$. Then T is a c.e.b.l. functional if $T(\boldsymbol{\theta}) = 0$ for $\boldsymbol{\theta} \in \Theta'$. Hence this corollary follows from the fact that the orthogonal complement of any subset in a Hilbert space is closed. □

However, unlike the finite dimensional case, the set Θ' may not be a closed space; it may only be a dense subset of the orthogonal complement of the space of all c.e.b.l. functionals. Thus caution should always be taken when one wants to characterize the set of all c.e.b.l. functionals; see, for instance, Example 2 of Section 4.

We now discuss the case where (2.2) is violated and the observations may be dependent or correlated and may have unequal variances. First, we observe that the independence assumption about observations is needed only when verifying "(ii) ⇒ (iii)". Thus, if the observations are dependent but uncorrelated with common error distribution and satisfy (2.2), then it still holds that (iii) ⇒ (iv) ⇒ (v) ⇒ (i) ⇒ (ii). Thus (iii) is sufficient for T to be a consistent estimator in most situations. Next, suppose the covariances of the observations are known up to a constant. Denote the covariance matrix of the first n observations by V_n. Let A_n be the lower triangular matrix such that $A_n V_n A'_n = I_{nxn}$. Now, transform the original data $(y_1, \cdots, y_n)'$ to $(z_1, \cdots, z_n)' = A_n(y_1, \cdots, y_n)'$. For the new data, the observations are now homoscedastic and uncorrelated. Let $(a_{i1}, \cdots, a_{ii}, 0, \cdots)$ be the ith row of A_n. The regression model for z_i becomes

$$z_i = \langle \textstyle\sum_{j=1}^{i} a_{ij} \mathbf{x}_j, \boldsymbol{\theta} \rangle + \varepsilon_i'.$$

(Note that since A_{n-1} is the left-upper submatrix of A_n, z_i should be independent of n.) Thus writing $\mathbf{x}_i^* = \sum_{j=1}^{i} a_{ij} \mathbf{x}_j$, we can establish "(iii) ⇒ (iv) ⇒ (v) ⇒ (i) ⇒ (iii)" after substituting $\mathbf{x}_i$ by $\mathbf{x}_i^*$ in the Theorem 2.1. Moreover, if $\{\varepsilon_i\}$ is Gaussian, then the ε_i', $i = 1, 2 \cdots$ are independent; hence "(ii) ⇒ (iii)" holds and the analogue of Theorem 2.1 is now established. However, to what extent the condition about the existence of the second moment of the error distribution can be released is not clear to the present author yet.

A comment about the consistent estimators constructed by the method used in the proof of "(iv) $\Rightarrow$ (v)" is given below. By examining the proof carefully, it is not hard to see that we may replace b_n in (2.9) by $b_n = \max\{x_{ii}, \lambda\}$, or by

$$b_n = (\lambda^2 + x_{ii}^2)^{1/2}, \tag{3.1}$$

where λ is any fixed positive number. The role of λ here is similar to the role of the ridge constant in the ridge regression or the role of a smoothing parameter in any ill-posed problem; it controls the tradeoff between the variance and the bias. Therefore, one might expect that an adaptive choice of λ should be more useful in practice. This should be investigated in the future. Also, for other commonly-used estimation procedures such as the smoothing spline method (which essentially uses (2.9) with b_n similar to (3.1)) in the nonparametric regression setting of Section 4, their consistency property should also be examined under the general framework discussed here.

An anonymous referee kindly provided the author with a proof of "(iv) $\Rightarrow$ (v)" that depends only on Banach structures of the parameter space Θ. When reduced to our Hilbertian Θ, his estimator is somewhat similar to (2.9) with (3.1).

The equivalence between (v) and (i) is also interesting. Without the specific setup, particularly the conditions (2.2) and (2.3), (i) generally does not imply (v).

Our last remark concerns the estimability of the linear combinations of parameters as defined in Scheffé (1958); i.e., a linear combination of parameters is estimable if there exists an unbiased estimator. In the finite dimensional case, it is true that if $T(\cdot)$ is consistently estimable then $T(\cdot)$ is estimable. But this is not necessarily the case in the infinite dimensional situation. $T(\cdot)$, when represented as an element in the Hilbert space concerned, can be outside of the linear space generated by $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ for any n but still retains the property of consistent estimability. This will become clearer when we consider the nonparametric regression setting of the next section.

4. Nonparametric regression.

In this section, the following nonparametric regression problem will be considered. For $m \geq 1$ (m is considered as a fixed integer hereafter), let $W_2^m[0, 1] = \{f^{(m-1)}$ is absolutely continuous on $[0, 1]$ and $\int_0^1 f^{(m)}(t)^2\, dt < \infty\}$. $W_2^m[0, 1]$ is a separable Hilbert space when equipped with the inner product $\langle f, g\rangle \sum_{i=0}^m \int_0^1 f^{(i)}(t) g^{(i)}(t)\, dt$. (It should be clear from the context whether $\langle \cdot, \cdot\rangle$ is the inner product of $W_2^m[0, 1]$ or the inner product of ℓ^2). Suppose a sequence of points in $[0, 1]$, $\{t_1, t_2, \cdots\}$ is given. We observe y_i, $i = 1, 2, \cdots$, which follow the nonparametric regression model:

$$y_i = f(t_i) + \varepsilon_i, \tag{4.1}$$

where $f \in W_2^m[0, 1]$ and ε_i are i.i.d. with a common distribution satisfying the conditions (2.2) and (2.3). Our goal is then to characterize the set of all c.e.b.l. functions on $W_2^m[0, 1]$.

Let $\{f_1, f_2, \cdots\}$ be a complete orthonormal system of $W_2^m[0, 1]$. Any f in $W_2^m[0, 1]$ can be represented as $\sum_{j=1}^{\infty} \langle f, f_j \rangle f_j$. In particular, for any $t \in [0, 1]$, the bounded linear functional $D_t^{(0)}$, defined by $\langle D_t^{(0)}, f \rangle = f(t)$, can be written as $\sum_{j=1}^{\infty} \langle D_t^{(0)}, f_j \rangle f_j$. Take $\theta_j = \langle f, f_j \rangle$ and $x_j = \langle D_t^{(0)}, f_j \rangle = f_j(t)$. Denote $\boldsymbol{\theta} = (\theta_1, \theta_2 \cdots)'$ and $\mathbf{x} = (x_1, x_2, \cdots)'$. When any observation y is made at the point t, we can rewrite (4.1) as

$$y = f(t) + \varepsilon = \langle D_t^{(0)}, f \rangle + \varepsilon = \langle \mathbf{x}, \boldsymbol{\theta} \rangle + \varepsilon.$$

Hence our setup is indeed a special case of (1.1) with $\Theta = \ell^2$. We may apply Theorem 2.1 to derive the desired results as follows.

LEMMA 4.1. *$T(\cdot)$ is a* c.e.b.l. *functional if and only if*

$$\sum_{i=1}^{\infty} f(t_i)^2 = \infty, \quad \textit{for any } f \textit{ in } \quad W_2^m[0, 1] \quad \textit{such that} \quad T(f) \neq 0. \tag{4.2}$$

PROOF. Represent f and $D_{t_i}^{(0)}$ as $\boldsymbol{\theta}$ and $\mathbf{x}_i$. It follows that $\langle \mathbf{x}_i, \boldsymbol{\theta} \rangle = f(t_i)$. Therefore Theorem 2.1 applies. (4.2) follows from (2.4). □

Now let us consider an important class of bounded linear functionals on $W_2^m[0, 1]$, namely the differential functional $D_t^{(k)}$, which maps any f in $W_2^m[0, 1]$ to its kth derivative at the point t, $f^{(k)}(t)$. Note that $D_t^{(k)}$ is a bounded linear functional only when $0 \le k \le m - 1$. To characterize the set of all $D_t^{(k)}$ which are consistently estimable we shall, equivalently, determine the *consistency region of degree k*, defined to be the set

$$C_k = \{t \mid t \in [0, 1] \text{ and } D_t^{(k)} \text{ is consistently estimable}\}. \tag{4.3}$$

Recall the definition of the limiting points of degree k from Section 1. Observe that according to our definition, a limiting point of degree k is also a limiting point of degree less than k. Other useful properties about limiting points are described in the following lemma. The topology considered here is restricted to [0, 1]; e.g., (½] is an open set, etc. We omit the proof.

LEMMA 4.2. *A point t^* is a limiting point of degree k if and only if*

$$\sum_{t_i \in \mathbf{N}}^{\infty} (t_i - t^*)^{2k} = \infty, \quad \textit{for any open neighborhood } \mathbf{N} \textit{ of } t^*. \tag{4.4}$$

In particular, the set of all limiting points of degree k is closed, and the set of all limiting points which is of degree 0 *but is not of degree k is discrete.*

The consistency region of degree k is now characterized below.

THEOREM 4.1. *For any integer k such that $0 \le k \le m - 1$, the consistency region of degree k, C_k, consists of all limiting points of degree k.*

The following properties of C_k follow from Lemma 4.2 and Theorem 4.1.

COROLLARY 4.1. *C_k is compact, $C_k \supset C_{k+1}$, and $C_0 - C_{m=1}$ is discrete* (*and is therefore countable.*)

PROOF OF THEOREM 4.1. First, we show that for any limiting point t of degree k, $D_t^{(k)}$ is a c.e.b.l. functional. By Lemma 4.1, it suffices to show that for any f such that $f^{(k)}(t^*) \neq 0$, $\sum_{i=1}^{\infty} f(t_i)^2 = \infty$. Let γ be the smallest nonnegative integer such that $f^{(\gamma)}(t^*) \neq 0$. Because of the continuity of $f^{(\gamma)}$, it is easy to see that $|f(t)| \geq \frac{1}{2}|f^{(\gamma)}(t^*)(t - t^*)|^{\gamma}$ for any t in an open neighborhood $\mathbf{N}$ of t^*. Since $\gamma \leq k$ and t^* is a limiting point of degree k, it follows that t^* is also a limiting point of degree γ. Apply Lemma 4.2 (taking k to be γ) and we conclude that

$$\sum_{i=1}^{\infty} f(t_i)^2 \geq \sum_{t_i \in \mathbf{N}} f(t_i)^2 \geq \tfrac{1}{2} f^{(\gamma)}(t^*)^2 \sum_{t_i \in \mathbf{N}} (t_i - t^*)^{2\gamma} = \infty.$$

(4.2) is now established and $D_{t^*}^{(k)}$ is therefore a c.e.b.l. functional.

Next, for any t^* which is not a limiting point of degree k, we shall demonstrate that there exists f in $W_2^m[0, 1]$ such that $f^{(k)}(t^*) \neq 0$ and $\sum_{i=1}^{\infty} f(t_i)^2 < \infty$; this then implies that $D_{t^*}^{(k)}$ is not a c.e.b.l. functional due to Lemma 4.1. By Lemma 4.2, let $\mathbf{N}$ be an open neighborhood of t^* such that $\sum_{t_i \in \mathbf{N}}^{\infty} (t_i - t^*)^{2k} < \infty$. Construct a function f in $W_2^m[0, 1]$ such that $f(t) = 0$ for any $t \notin \mathbf{N}$, $f^{(\gamma)}(t^*) = 0$ for any $\gamma < k$, and $f^{(k)}(t^*) \neq 0$. It follows that $\sum_{i=1}^{\infty} f(t_i)^2 = \sum_{t_i \in \mathbf{N}} f(t_i)^2 \leq \sum_{t_i \in \mathbf{N}} M (t_i - t^*)^{2k} < \infty$, where $M \geq \sup\{f^{(k)}(t)^2 \mid t \in [0, 1]\}$. Therefore $D_{t^*}^{(k)}$ is not a c.e.b.l. functional, and the proof is complete. □

By Theorem 4.1, we may easily see whether $D_t^{(k)}$ is a c.e.b.l. functional or not. We call $D_t^{(k)}$ a c.e.b.l. *functional of differential type* if $t \in C_k$. An application of Corollary 3.1 then shows that any element in the closed linear space generated by all c.e.b.l. functionals of differential type is also a c.e.b.l. functional. Naturally, we would like to know if there are any other c.e.b.l. functionals or not. The following example gives some clues to the answer. It also demonstrates that the space of all θ such that (2.4) does not hold may not be closed as was already pointed out in Section 3.

EXAMPLE 2. Suppose $t_i \to t^*$ as $i \to \infty$ and $\sum_{i=1}^{\infty} (t_i - t^*)^{2m} = \infty$. By Theorem 4.1, $D_{t^*}^{(k)}$, $k = 0, 1, \cdots, m - 1$, are the only c.e.b.l. functionals of differential type. We now show that the linear space generated by $D_{t^*}^{(k)}$, $k = 0, 1, \cdots, m - 1$, is exactly the set of all c.e.b.l. functionals.

Consider an f in $W_2^m[0, 1]$ which is orthogonal to $D_{t^*}^{(k)}$, $k = 0, 1, \cdots, m - 1$; i.e., $f^{(k)}(t^*) = 0$, for $0 \leq k \leq m - 1$. To show that $\langle f, \cdot \rangle$ is not a c.e.b.l. functional we have to find a g such that $\langle f, g \rangle \neq 0$ and $\sum_{i=1}^{\infty} g(t_i)^2 < \infty$ (by Lemma 4.1). By some trivial argument it can be shown that for any $\varepsilon > 0$, there exists a g which equals 0 in a small open interval $\mathbf{I}$ containing t^* such that $\| f - g \| < \varepsilon$. Take ε small enough to insure that $\langle f, g \rangle \neq 0$. It follows that $\sum_{i=1}^{\infty} g(t_i)^2 = \sum_{i=1}^{n} g(t_i)^2 < \infty$, where n is an integer such that $t_i \in \mathbf{I}$ for $i > n$. Thus the desired result is established.

Note that in this example, the set of all g such that $\sum_{i=1}^{\infty} g(t_i)^2 < \infty$ is not closed. To see this, consider the function $g_0(t) = (t - t^*)^m$. Obviously, $g_0^{(0)}(t^*) = \cdots = g_0^{(m-1)}(t^*) = 0$ and $\sum_{i=1}^{\infty} g_0(t_i)^2 = \infty$. However, if the set of all g such that $\sum_{i=1}^{\infty} g(t_i)^2 < \infty$ were closed, then this set would equal the orthogonal complement of the space of all c.e.b.l. functionals (i.e., $\{g \mid g^{(0)}(t^*) = \cdots = g^{(m-1)}(t^*) = 0\} = \{g \mid \sum_{i=1}^{\infty} g(t_i)^2 < \infty\}$). Thus a contradiction is obtained because g_0 can't be in both sets.

The argument used in the Example 2 can be generalized to show that the set of all c.e.b.l. functionals equals the closed space generated by all c.e.b.l. functionals of differential type when $C_0 - C_{m-1}$ is a finite set (but the cardinality of C_{m-1} may be infinite). Consider f in $W_2^m[0, 1]$ such that $f^{(0)}(t) = \cdots = f^{(m-1)}(t) = 0$ for all $t \in C_{m-1}$ and $f^{(k)}(t) = 0$ for all $t \in C_k$, $k = 0, 1, \cdots, m-2$. Since C_{m-1} is compact and $C_0 - C_{m-1}$ is a finite set, without much difficulty we can construct a function g in $W_2^m[0, 1]$ such that

(4.5) $g(t) = 0$ for t in a union of finitely many open intervals covering C_{m-1}; $g^{(k)}(t) = 0$ for $t \in C_k$, $k = 0, 1, \cdots, m-2$; and $\| f - g \| < \varepsilon$ for $\varepsilon > 0$.

It follows that for ε small enough, we have $\langle f, g \rangle \neq 0$ and $\sum_{i=1}^{\infty} g(t_i)^2 < \infty$. Thus f is not a c.e.b.l. functional and the desired result is obtained.

However, for the case that $C_0 - C_{m-1}$ is not a finite set, it is still not clear to the present author whether a g satisfying (4.5) exists or not. Thus it remains unknown whether the set of all c.e.b.l. functionals equals the closed linear space generated by all c.e.b.l. functionals of differential type or not.

Finally, we draw the connection between our results here and those obtained by Wu (1980) in the consideration of the polynomial regression model. Suppose the model is $y_j = f(t_j) + \varepsilon_j = \sum_{i=0}^{m-1} \theta_i t_j^i + \varepsilon_j$, where ε_j are independent random errors satisfying conditions (2.2) and (2.3) and $t_j \in [0, 1]$. Then the following statements hold:

(i) The consistency region of degree k ($0 \le k \le m-1$) contains (but may not equal) all the limiting points of degree k.
(ii) The set of all c.e.b.l. functionals equals the linear space generated by the c.e.b.l. functionals of differential type obtained by (i).
(iii) From (ii), it is easy to identify the consistency region (of degree k): it equals [0, 1] if the dimension of the set of all c.e.b.l. functionals is m; otherwise it equals the set of all limiting points of degree k.

These results can be derived from Theorem 4 of Wu (1980) (in particular, (4.4) and (4.6) can be simplified; see Li, 1982). At the revision of this paper, the author learned that Wang and Wu (1983) essentially derived these results by a different method which also works for the extended Tchybycheff system. It is interesting to observe that their Proposition 2 can be derived from our Theorem 4.1 because their space of response functions has only finite dimensions and is contained in our richer space $W_2^m[0, 1]$. For the same reason, the above results (i) ~ (iii) for polynomial regression follow immediately from our results for nonparametric regression (Note that we have established (ii) for finite $C_0 - C_{m-1}$; but when $C_0 - C_{m-1}$ is infinite, due to the finiteness of the dimensions of Θ, (ii) is also obvious). Wang and Wu (1984) have further results extending their earlier paper to the multivariate case.

Acknowledgements. I am grateful to the associate editor and a referee for providing valuable suggestions on improving the presentation of this work. In particular, they brought to my attention the paper of L. Shepp (1965). Another anonymous referee kindly provided me with an alternative proof of (iii) $\Rightarrow$ (iv) of Theorem 2.1 (see Section 3).

REFERENCES

DRYGAS, H. (1976). Weak and strong consistency of the least squares estimates in regression models. *Z. Wahrsch. verw. Gebiete.* **34** 119–127.

LI, K. C. (1982). Regression models with infinitely many parameters: Consistency of bounded linear functionals. Mimeograph Series #82-10. Department of Statistics, Purdue University.

ROYDEN, H. L. (1972). *Real Analysis.* Macmillan, New York.

SCHEFFÉ H. (1958). *The Analysis of Variance.* Wiley, New York.

SHEPP, L. A. (1965). Distinguishing a sequence of random variables from a translate of itself. *Ann. Math. Statist.* **36** 1107–1112.

WANG, S. G. and WU, C. F. (1983). Further results on the consistent directions of least squares estimators. *Ann. Statist.* **11** 1257–1263.

WANG, S. G. and WU C. F. (1984). Consistent directions of the least squares estimators in linear models. *Proc. Pacific Area Statist. Conf.*, to appear.

WU, C. F. (1980). Characterizing the consistent directions of least squares estimates. *Ann. Statist.* **8** 789–801.

WU, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907