Ann. Inst. Statist. Math.
Vol. 45, No. 4, 639–654 (1993)

SEMIPARAMETRIC RANDOM COEFFICIENT REGRESSION MODELS*

RUDOLF BERAN

Department of Statistics, University of California, Berkeley, Berkeley, CA 94720, U.S.A.

(Received March 23, 1992; revised January 11, 1993)

Abstract. Linear regression models with random coefficients express the idea that each individual sampled may have a different linear response function. Technically speaking, random coefficient regression encompasses a rich variety of submodels. These include deconvolution or affine-mixture models as well as certain classical linear regression models that have heteroscedastic errors, or errors-in-variables, or random effects. This paper studies minimum distance estimates for the coefficient distributions in a general, semiparametric, random coefficient regression model. The analysis yields goodness-of-fit tests for the semiparametric model, prediction regions for future responses, and confidence regions for the distribution of the random coefficients.

Key words and phrases: Minimum distance, empirical characteristic function, errors-in-variables, deconvolution, random effects, statistical inference.

1. Introduction

In a random coefficient linear regression model, the response Y_i of individual i is related to the covariates X_i through the linear equation

$$Y_i = A_i + X_i B_i, \qquad 1 \leq i \leq n, \tag{1.1}$$

where the coefficients (A_i, B_i) are random elements. For the purposes of this paper, Y_i is a $p \times 1$ random vector of responses, X_i is a $p \times q$ random matrix of covariates, and the coefficients (A_i, B_i) are $p \times 1$ and $q \times 1$ random vectors respectively. The triples $\{(A_i, B_i, X_i)\}$ are assumed i.i.d. and the coefficients (A_i, B_i) are independent of the covariates X_i for every i. The distributions F_{AB} of (A_i, B_i) and F_X of X_i are unknown, though possibly restricted in ways that will be described later. Observed is the sample S_n consisting of the n pairs $\{(Y_i, X_i) : 1 \leq i \leq n\}$. The initial goals are to estimate F_{AB} and to test the goodness-of-fit of model (1.1).

The regression model just described expresses three ideas about the data. First is the supposition that the i-th response Y_i depends linearly on the i-th

* This research was supported in part by NSF Grant DMS 9001710.

set of covariates X_i. Second is the notion that the coefficients (A_i, B_i) of the linear response function can vary with i. Third is the belief that S_n behaves like a simple random sample from a large population. The first two ideas are embodied by equation (1.1). The third idea is expressed by the i.i.d. assumption on $\{(Y_i, X_i, A_i, B_i)\}$.

Model (1.1) is equivalent mathematically to a multivariate linear regression model with random regressors and structured heteroscedastic errors. To see this, write $A_i = a + \tilde{A}_i$ and $B_i = b + \tilde{B}_i$ where $a = EA_i$ and $b = EB_i$. The classical homoscedastic linear regression model then corresponds to the special case in which the distribution of B_i is supported in one point. In the econometric literature, models like (1.1) have been used to analyze panel data (cf. Hsiao (1986)) and to investigate heteroscedasticity (cf. Hildreth and Houck (1968), Goldfeld and Quandt (1972), Chapter 3, and Amemiya (1977)). Surveys of work on random coefficient regression models, on their autoregressive analogs, and on models combining both features are given by Raj and Ullah (1981), Chow (1983), Nicholls and Pagan (1985), and Newbold (1988).

In the statistical literature, various submodels of (1.1) are well-established under several labels. When the distribution of X_i is supported on one known point, then (1.1) includes the random effects models of ANOVA (cf. Scheffé (1959), Chapter 7) and the models studied in nonparametric deconvolution (cf. Fan (1991), van Es (1991)). Errors-in-variables linear regression (cf. Spiegelman (1979)) corresponds to a submodel in which the distribution of B_i is supported in one unknown point. When the $\{X_i : 1 \leq i \leq n\}$ are not observed but have a known distribution, then (1.1) becomes an affine-mixture model.

Much of this literature on model (1.1) seeks to estimate the first two moments of F_{AB}, often under the additional assumption that A_i and B_i are independent or uncorrelated. Beran and Hall (1992) first treated nonparametric estimation of F_{AB}. Under the assumption that A_i and B_i are independent scalar random variables, their paper gives consistent nonparametric estimators for the marginal coefficient distribution F_A and F_B. Because the construction requires estimating many moments of (A_i, B_i), it is difficult to extend their method to the multivariate model (1.1). The asymptotic distributions of their estimators are unknown.

Beran and Millar (1991) introduced and studied a class of nonparametric minimum distance estimators for F_{AB}. Their estimation idea is to fit the distribution of (Y_i, X_i) under model (1.1) to the empirical distribution of the sample S_n. The criterion of fit is the distance between these two distributions, measured in a metric for weak convergence. Under conditions that ensure the identifiability of F_{AB}, such nonparametric minimum distance estimators are consistent for F_{AB}.

Particularly tractable are the metrics for weak convergence generated by $L_2(Q)$ norms on characteristic functions, Q being any probability measure with full support on R^{pq+p}. For such Hilbertian distances and for F_{AB} restricted to a certain parametric family of distributions (see Example 2 below), Beran and Millar (1991) established $n^{1/2}$-consistency of the minimum distance estimator for F_{AB}. This result is the starting point for the semiparametric model and asymptotic distribution theory that are developed in the present paper.

Consider the semiparametric version of model (1.1) in which the unknown distribution F_X of X_i is unrestricted but the unknown distribution of (A_i, B_i) belongs to a parametric family $\mathcal{F}_{AB} = \{F_{AB}(\theta) : \theta \in \Theta\}$. Here Θ is an open subset in R^k. Three examples illustrate the flexibility of such models:

Example 1. $F_{AB}(\theta)$ is a discrete distribution supported on r distinct sites $\{s_i : 1 \leq i \leq r\}$ in R^{p+q}. These sites are ordered by their first coordinates, with ties broken by second coordinate ordering, and so forth. The probability supported on each site s_i is $1/r$. Here $\theta = (s_1, s_2, \ldots, s_r)$ and the dimension of θ is $k = (p+q)r$.

Example 2. The support of $F_{AB}(\theta)$ is discrete as in Example 1. The probability supported on site s_i is now p_i, where $p_i > 0$ and $\sum_{i=1}^{r-1} p_i < 1$. Here $\theta = (s_1, s_2, \ldots, s_r, p_1, \ldots, p_{r-1})$ and the dimension of θ is $k = (p+q+1)r - 1$. This model was analyzed in Beran and Millar (1991).

Example 3. $\mathcal{F}_{AB}$ is a canonical exponential family model indexed by θ. Here Θ is the interior of the natural parameter space.

The models of Examples 1 and 2 are rich enough to approximate a wide range of distributions for (A_i, B_i). These two models do not induce classically regular semiparametric models (1.1), in the sense of Begun *et al.* (1983), because the support of $F_{AB}(\theta)$ depends on θ. However, the model of Example 3 is classically regular in this sense. On the other hand, all three examples are regular from the viewpoint of the minimum distance procedures studied in this paper.

Section 2 gives general conditions under which the minimum $L_2(Q)$-distance estimator $\hat{\theta}_n$ of θ is asymptotically normal. Under these same conditions, the minimized distance between the empirical characteristic function of the sample S_n and the modelled characteristic function of (Y_i, X_i) also has a limiting distribution. Both limiting distributions, especially the latter, are complicated analytically. For purposes of statistical inference, it is natural to consider bootstrap estimates of these and related sampling distributions. The theory in Section 2 implies the consistency of such semiparametric bootstrap distributions and of their quantiles.

However, calculation of the minimum distance estimator $\hat{\theta}_n$ requires a minimization over a k-dimensional parameter space, where k is typically not small (cf. Examples 1 and 2 above and the computational discussion in Beran and Millar (1991)). Thus, directly bootstrapping $\hat{\theta}_n$ several hundred times or more may not be practical. Subsection 3.1 suggests replacing the bootstrap recalculation θ_n^* of $\hat{\theta}_n$ with a linearized approximation motivated by the asymptotics. Bootstrap distributions based on this approximation retain the essential consistency property while being much easier to compute.

Subsection 3.2 uses the linearized bootstrap to construct a goodness-of-fit test for the semiparametric random coefficient regression model and to construct confidence regions for the coefficient distribution $F_{AB}(\theta)$. In another direction, Subsection 3.2 also develops prediction regions for Y_{n+1}, given $X_{n+1} = x$ and the learning sample S_n. Section 4 proves the main results of the paper.

2. Minimum distance fits

The main topics of this section are: the semiparametric random coefficient regression model; definition of the minimum distance estimator $\hat{\theta}_n$ and of the associated minimized distance; and the asymptotic distributions of both statistics. Triangular array asymptotics are used throughout. This approach establishes the local uniformity of the weak convergence studied and thereby ensures the consistency of the natural semiparametric bootstrap estimates.

2.1 *The model and procedures*

To the basic structure of model (1.1) described in the first paragraph of the Introduction, we add the following elements:

$F_{AB}(\theta)$ denotes the distribution of (A_i, B_i) which belongs to a family $\mathcal{F}_{AB} = \{F_{AB}(\theta) : \theta \in \Theta\}$ of distributions on R^{p+q};

The parameter space Θ is an open subset of R^k;

F_X denotes the distribution of X_i, which is restricted to a family $\mathcal{F}_X$ of distributions on R^{pq};

$P(\theta, F_X)$ denotes the distribution of (Y_i, X_i) under model (1.1);

$\hat{P}_n$ denotes the empirical distribution of the sample $S_n = \{(Y_i, X_i) : 1 \leq i \leq n\}$;

$\hat{F}_{X,n}$ denotes the empirical distribution of the $\{X_i; 1 \leq i \leq n\}$.

The semiparametric random coefficient regression model asserts that the observable pairs $\{(Y_i, X_i) : 1 \leq i \leq n\}$, which constitute the sample S_n, are i.i.d., with common distribution belonging to the family $\{P(\theta, F_X) : \theta \in \Theta,\ F_X \in \mathcal{F}_X\}$. Examples 1 to 3 in the Introduction illustrate the scope of this model.

To define the Beran and Millar (1991) minimum distance estimates for θ requires the characteristic functions of $\hat{P}_n$ and of $P(\theta, F_X)$. The first of these, the empirical chf $\hat{\varphi}_n$, is given by

$$\hat{\varphi}_n(t, u) = n^{-1} \sum_{j=1}^{n} \exp[i\langle t, Y_j\rangle + i\langle u, X_j\rangle], \tag{2.1}$$

where $t \in R^p$, $u \in R^{pq}$ and $\langle \cdot, \cdot \rangle$ denotes the inner product of appropriate dimension. Let $\varphi(\theta)$ denote the chf of $F_{AB}(\theta)$,

$$\varphi(t, v; \theta) = E \exp[i\langle t, A\rangle + i\langle v, B\rangle], \tag{2.2}$$

where $t \in R^p$, $v \in R^q$ and (A, B) has distribution $F_{AB}(\theta)$. The chf of $P(\theta, F_X)$ is then $\psi(\theta, F_X)$, defined by

$$\psi(t, u; \theta, F_X) = E\{\varphi(t, X_j' t; \theta) \exp[i\langle u, X_j\rangle]\}, \tag{2.3}$$

where $t \in R^p$, $u \in R^{pq}$, and X_j' denotes the transpose of X_j, which has distribution F_X.

Suppose P_1 and P_2 are any two distribution on R^{p+pq}, with chf's φ_1 and φ_2 respectively. Let Q be a probability measure on R^{p+pq} that has full support and let $\|\cdot\|$ denote the $L_2(Q)$ norm on chf's:

$$\|\varphi_i\| = \left[\int |\varphi_i|^2 dQ\right]^{1/2}. \tag{2.4}$$

Then, the distance between distributions P_1, P_2 defined by $\|\varphi_1 - \varphi_2\|$ metrizes weak convergence of probabilities on R^{p+pq}. Moment-matching considerations suggest a choice of Q that puts most of its mass near the origin. However, this aspect is not important for the asymptotic theory in this paper. We say that $\hat{\theta}_n$ is a *minimum distance estimator* of θ if

$$\|\hat{\varphi}_n - \psi(\hat{\theta}_n, \hat{F}_{x,n})\| = \inf_{\theta \in \Theta} \|\hat{\varphi}_n - \psi(\theta, \hat{F}_{X,n})\| + o(n^{-1/2}). \tag{2.5}$$

Let

$$T_n = n^{1/2} \inf_{\theta \in \Theta} \|\hat{\varphi}_n - \psi(\theta, \hat{F}_{X,n})\| \tag{2.6}$$

denote the rescaled *minimized distance* between the empirical distribution and the fitted semiparametric model. Relatively large values of T_n are evidence that this semiparametric model does not fit the sample. Later, in Subsections 2.2 and 3.2, we will see how to quantify the phrase "relatively large".

In definition (2.5), the distribution of F_X is estimated by the empirical distribution $\hat{F}_{X,n}$. One can instead estimate the pair (θ, F_X) by the minimum distance criterion. The additional computational complexity of this joint estimation is daunting, however.

2.2 *Asymptotic distributions*

The asymptotic distributions of $\hat{\theta}_n$ and of the minimized distance T_n will be derived under certain regularity assumptions on the semiparametric model $\{P(\theta, F_X) : \theta \in \Theta,\, F_X \in \mathcal{F}_X\}$. These assumptions are weak enough to include the three examples in the Introduction. Subsection 2.3 gives details on this point. In broad outline, the reasoning follows the minimum distance literature (cf. Pollard (1980) and references therein). Some additional argument is needed to handle the effects of estimating the infinite-dimensional nuisance parameter F_X by the empirical distribution $\hat{F}_{X,n}$.

Let $(\theta_0, F_{X,0})$ denote a fixed point in $\Theta \times \mathcal{F}_X$. We will require the following assumptions:

A1 (strong identifiability). If $\|\psi(\theta, F_X) - \psi(\theta_0, F_{X,0})\| \to 0$, then $\theta \to \theta_0$.

A2 (norm differentiability). If $F_X \Rightarrow F_{X,0}$ and $\theta \to \theta_0$, then there exists a $k \times 1$ vector function $\eta_0 = \eta(\theta_0, F_{X,0})$, whose components belong to $L_2(Q)$, such that

$$|\theta - \theta_0|^{-1} \|\psi(\theta, F_X) - \psi(\theta_0, F_X) - \langle \theta - \theta_0, \eta_0 \rangle\| \to 0. \tag{2.7}$$

A3 (nonsingularity). There exits a finite positive constant C such that

$$\|\langle t, \eta_0 \rangle\| \geq C|t| \tag{2.8}$$

for every $t \in R^k$.

Note that the hypothesis in A1 implies that $F_X \Rightarrow F_{X,0}$. Convenient sufficient conditions for A1 to A3 are the topic of Subsection 2.3.

To state the basic result in the asymptotic behavior of $\hat{\theta}_n$ and T_n requires a little more notation. For any sequence $\{\theta_n \in \Theta\}$ such that $\{n^{1/2}(\theta_n - \theta_0)\}$ is bounded and for any sequence $\{F_{X,n} \in \mathcal{F}_X\}$ such that $F_{X,n} \Rightarrow F_{X,0}$, let

$$W_n = n^{1/2}[\hat{\varphi}_n - \psi(\theta_n, F_{X,n})] - n^{1/2}[\psi(\theta_0, \hat{F}_{X,n}) - \psi(\theta_0, F_{X,n})]. \tag{2.9}$$

This is a complex-valued process whose realizations are elements of $L_2(Q)$. Let

$$\hat{t}_n = \operatorname{argmin}_{t \in R^k} \|W_n - \langle t, \eta_0 \rangle\|. \tag{2.10}$$

By easy algebra, the unique minimizing value in (2.10) is

$$\hat{t}_n = \left[\int \mathcal{R}(\bar{\eta}_0 \eta_0') dQ\right]^{-1} \int \mathcal{R}(\bar{\eta}_0 W_n) dQ, \tag{2.11}$$

where $\bar{\eta}_0$ denotes the coordinatewise complex conjugate of the vector function η_0 and $\mathcal{R}$ denotes real part. The matrix inverse exists whenever A3 holds.

PROPOSITION 2.1. *Suppose that assumptions* A1 *to* A3 *are satisfied, that* $\{n^{1/2}(\theta_n - \theta_0)\}$ *is bounded and that* $F_{X,n} \Rightarrow F_{X,0}$. *Then, under the sequence of models* $\{P(\theta_n, F_{X,n})\}$,

$$n^{1/2}(\hat{\theta}_n - \theta_n) = \hat{t}_n + o_p(1) \tag{2.12}$$

and

$$\begin{aligned} T_n &= \inf_{t \in R^k} \|W_n - \langle t, \eta_0 \rangle\| + o_p(1) \\ &= \|W_n - \langle \hat{t}_n, \eta_0 \rangle\| + o_p(1). \end{aligned} \tag{2.13}$$

The proofs of this and all other Propositions appear in Section 4. To derive the asymptotic distributions of $\{n^{1/2}(\hat{\theta}_n - \theta_n)\}$ and of $\{T_n\}$ from Proposition 2.1, we need to consider the weak convergence of the processes $\{W_n\}$. Let $\{Z(y,x) : y \in R^p, x \in R^{pq}\}$ be a gaussian process with mean zero and covariance structure

$$\operatorname{Cov}[Z(y,x), Z(y',x')] = G(y \wedge y', x \wedge x') - G_0(y,x) G_0(y',x') \tag{2.14}$$

where G_0 is the cdf of $P(\theta_0, F_{X,0})$ and $x \wedge x'$ denotes the vector of coordinatewise minima. Define the complex-valued process $W_0 = \{W_0(t,u) : t \in R^p, u \in R^{pq}\}$ by

$$\begin{aligned} W_0(t,u) = &\iint \exp[i\langle t, y \rangle + i\langle u, x \rangle] dZ(y,x) \\ &- \iint \varphi(t, x't; \theta_0) \exp[i\langle u, x \rangle] dZ(y,x). \end{aligned} \tag{2.15}$$

Under $\{P(\theta_n, F_{X,n})\}$, the processes $\{W_n\}$ converge weakly, as random elements of $L_2(Q)$, to the process W_0. This follows from the central limit theorem in $L_2(Q)$.

Let $H_n(\theta, F_X)$ denote $\mathcal{L}[n^{1/2}(\hat{\theta}_n - \theta) \mid P(\theta, F_X)]$ and let $J_n(\theta, F_X)$ denote $\mathcal{L}[T_n \mid P(\theta, F_X)]$. Correspondingly, let

$$H(\theta_0, F_{X,0}) = \mathcal{L}\left\{\left[\int \mathcal{R}(\bar{\eta}_0\eta_0')dQ\right]^{-1} \int \mathcal{R}(\bar{\eta}_0 W_0)dQ\right\} \tag{2.16}$$

and

$$J(\theta_0, F_{X,0}) = \mathcal{L}\left\{\inf_{t\in R^k} \|W_0 - \langle t, \eta_0\rangle\|\right\}. \tag{2.17}$$

Note that $H(\theta_0, F_{X,0})$ is a multivariate normal distribution with mean zero. On the other hand, $J(\theta_0, F_{X,0})$ can be represented as the distribution of the square root of a weighted sum of independent chi-squared random variables, the weights being functions of $(\theta_0, F_{X,0})$.

PROPOSITION 2.2. *Suppose that the assumptions for Proposition* 2.1 *hold. Then, under the sequence of models* $\{P(\theta_n, F_{X,n})\}$,

$$\begin{aligned} H_n(\theta_n, F_{X,n}) &\Rightarrow H(\theta_0, F_{X,0}), \\ J_n(\theta_n, F_{X,n}) &\Rightarrow J(\theta_0, F_{X,0}). \end{aligned} \tag{2.18}$$

Hence also

$$\begin{aligned} H_n(\hat{\theta}_n, \hat{F}_{X,n}) &\Rightarrow H(\theta_0, F_{X,0}), \\ J_n(\hat{\theta}_n, \hat{F}_{X,n}) &\Rightarrow J(\theta_0, F_{X,0}) \end{aligned}$$

in probability.

The random probability measures $H_n(\hat{\theta}_n, \hat{F}_{X,n})$ and $J_n(\hat{\theta}_n, \hat{F}_{X,n})$ are the natural semiparametric bootstrap estimates for the sampling distributions $H_n(\theta_n, F_{X,n})$ and $J_n(\theta_n, F_{X,n})$. By Proposition 2.2, these bootstrap estimates are consistent and so provide a basis for statistical inference concerning θ or the semiparametric random coefficient regression model itself. For instance, the goodness-of-fit test that rejects the assumed model whenever T_n exceeds the $(1 - \alpha)$-th quantile of $J_n(\hat{\theta}_n, \hat{F}_{X,n})$ has asymptotic probability α of rejecting when the model is correct. Unfortunately, the calculation of the desired quantile by straightforward Monte Carlo methods is not easy because of the minimization step. Nor is it convenient to compute the $(1 - \alpha)$-th quantile of the estimated asymptotic distribution $J(\hat{\theta}_n, \hat{F}_{X,n})$, which is an alternative asymptotically valid critical value for the test. Subsection 3.1 describes how these computational difficulties can be relieved through a linearized bootstrap algorithm.

2.3 *Checking assumptions*

Assumptions A1 to A3 on the model $\{P(\theta, F_X) : \theta \in \Theta, F_X \in \mathcal{F}_X\}$ are abstract in character. The following discussion provides more concrete sufficient conditions under which these assumptions hold.

Sufficient conditions for A1. Suppose that $\mathcal{F}_{AB}$ consists of probabilities supported on a fixed compact and that $\{x't : x \in \text{support } (F_{X,0})\}$ contains a non-empty open set in R^q for every $t \neq 0$ in R^p. Then the hypothesis in A1 implies

$$F_{AB}(\theta_n) \Rightarrow F_{AB}(\theta_0) \tag{2.19}$$

by Proposition 2.2 of Beran and Millar (1991). The desired conclusion is now equivalent to strong identifiability of the parametric model $\mathcal{F}_{AB}$.

Sufficient conditions for A2. Suppose that for every $(t, u) \in R^{p+pq}$ and for every (θ, F_X) in a neighborhood of $(\theta_0, F_{X,0})$, the chf $\psi(t, u; \theta, F_X)$ has partial derivatives $\{\eta_{\theta,F_X}(t, u) : 1 \leq j \leq k\}$ with respect to θ; and that these partial derivatives are continuous in (θ, F_X) over a neighborhood of $(\theta_0, F_{X,0})$. Suppose as well that the convergence $\theta \to \theta_0$, $F_X \Rightarrow F_{X,0}$ implies

$$\|\eta_{\theta,F_X,j}\| \to \|\eta_{\theta_0,F_{X,0},j}\| < \infty, \qquad 1 \leq j \leq k. \tag{2.20}$$

Then A2 holds with $\eta_0 = \{\eta_{\theta_0,F_{X,0},j} : 1 \leq j \leq k\}$. The proof rests upon the fundamental theorem of calculus and the Cauchy-Schwarz inequality.

Equivalent condition for A3. Since θ is finite dimensional, nonsingularity in the sense of A3 is equivalent to linear independence of the components of η_0 (cf. Pollard (1980)).

We illustrate the usefulness of these sufficient conditions by applying them to Example 1 of the Introduction. The analysis for Examples 2 and 3 is not significantly harder.

Example 1. (continued). Suppose for simplicity that $p = q = 1$, so that the sites in $\theta = (s_1, \ldots, s_r)$ have the form $s_j = (a_j, b_j)$, where a_j and b_j are real. Here, (2.3) reduces to

$$\psi(t, u; \theta, F_X) = r^{-1} \sum_{j=1}^{r} \exp[ita_j + ixtb_j] \exp(iux) dF_X(x). \tag{2.21}$$

Suppose that $\int t^2 dQ(t)$ is finite and $\mu(F_X) = \int |x| dF_X(x)$ is finite and weakly continuous over all $F_X \in \mathcal{F}_X$. The components of $\eta_{\theta,F_X}(t, u)$ are the $k = 2r$ elements

$$\begin{aligned} \frac{\partial\psi(t,u)}{\partial a_j} &= r^{-1} \int it \exp[ita_j + ixtb_j] \exp(iux) dF_X(x), \\ \frac{\partial\psi(t,u)}{\partial b_j} &= r^{-1} \int ixt \exp[ita_j + ixtb_j] \exp(iux) dF_X(x), \end{aligned} \tag{2.22}$$

as j ranges from 1 to r. The sufficient conditions for A2 hold by dominated convergence.

Since $F_{AB}(\theta)$ puts mass $1/r$ on each of the distinct, partially ordered sites $\{s_j : 1 \leq j \leq r\}$, the strong identifiability of the parametric model $\mathcal{F}_{AB}$ is evident. Hence, A1 holds if the support of $F_{X,0}$ contains a non-empty open set and the sites $\{s_i\}$ lie within a given compact. Finally, A3 holds because the partial derivatives (2.22) are linearly independent.

3. Statistical inference

Using the asymptotic distributions or bootstrap distributions of Proposition 2.2 is extremely computer-intensive. This section introduces a simpler linearized bootstrap algorithm for estimating sampling distributions of interest and then solves three inferential problems. These are: testing the fit of the semiparametric model $\{P(\theta, F_X) : \theta \in \Theta,\ F_X \in \mathcal{F}_X\}$; devising a prediction region for Y_{n+1}, given $X_{n+1} = x$ and the learning sample S_n; and constructing a confidence region for the distribution $F_{AB}(\theta)$ of the random coefficients.

3.1 *The linearized bootstrap*

Let $S_n^* = \{(Y_i^*, X_i^*) : 1 \le i \le n\}$ denote a bootstrap sample of size n drawn from the fitted random coefficient regression model. That is, the conditional distribution $\mathcal{L}[(Y_i^*, X_i^*) \mid S_n]$ is $P(\hat{\theta}_n, \hat{F}_{X,n})$ for every i and the pairs $\{(Y_i^*, X_i^*) : 1 \le i \le n\}$ are conditionally independent, given S_n. Write φ_n^* for the empirical chf of S_n^*, $F_{X,n}^*$ for the empirical distribution of the $\{X_i^* : 1 \le i \le n\}$, and θ_n^* for the minimum distance estimator of θ, recalculated from S_n^*. In other words, θ_n^* satisfies

$$\|\varphi_n^* - \psi(\theta_n^*, F_{X,n}^*)\| = \inf_{\theta\in\Theta} \|\varphi_n^* - \psi(\theta, F_{X,n}^*)\| + o(n^{-1/2}). \tag{3.1}$$

The two bootstrap distributions that were defined in Proposition 2.2 can be re-interpreted as conditional distributions:

$$\begin{aligned} H_n(\hat{\theta}_n, \hat{F}_{X,n}) &= \mathcal{L}[n^{1/2}(\theta_n^* - \hat{\theta}_n) \mid S_n], \\ J_n(\hat{\theta}_n, \hat{F}_{X,n}) &= \mathcal{L}\left[n^{1/2} \inf_{\theta\in\Theta} \|\varphi_n^* - \psi(\theta, F_{X,n}^*)\| \mid S_n\right]. \end{aligned} \tag{3.2}$$

It is these representations that justify the usual Monte Carlo approximations to the two bootstrap distributions. Yet, it is the difficulty of the minimization step that may render this direct Monte Carlo approach impractical.

The asymptotic approximations established in Proposition 2.1 suggest the following simplification. By analogy with (2.9), define the process

$$\tilde{W}_n^* = n^{1/2}[\varphi_n^* - \psi(\hat{\theta}_n, \hat{F}_{X,n})] - n^{1/2}[\psi(\hat{\theta}_n, F_{X,n}^*) - \psi(\hat{\theta}_n, \hat{F}_{X,n})]. \tag{3.3}$$

This is close to but not quite a bootstrapped version of W_n. Let $\eta_n = \eta_0(\hat{\theta}_n, \hat{F}_{X,n})$ and let

$$\tilde{\theta}_n^* = \hat{\theta}_n + n^{-1/2}\left[\int \mathcal{R}(\bar{\eta}_n \eta_n')dQ\right]^{-1} \int \mathcal{R}(\bar{\eta}_n \tilde{W}_n^*)dQ. \tag{3.4}$$

The statistic $\tilde{\theta}_n^*$ is a linear approximation to θ_n^* that is motivated by conclusion (2.12) of Proposition 2.1. The linearized bootstrap estimates for the sampling distributions $H_n(\theta_n, F_{X,n})$ and $J_n(\theta_n, F_{X,n})$ are defined to be

$$\begin{aligned} \tilde{H}_n(\hat{\theta}_n, \hat{F}_{X,n}) &= \mathcal{L}[n^{1/2}(\tilde{\theta}_n^* - \hat{\theta}_n) \mid S_n], \\ \tilde{J}_n(\hat{\theta}_n, \hat{F}_{X,n}) &= \mathcal{L}[n^{1/2}\|\varphi_n^* - \psi(\tilde{\theta}_n^*, F_{X,n}^*)\|] \end{aligned} \tag{3.5}$$

by analogy with (3.2). Both of these distributions admit computationally easier Monte Carlo approximations than do $H_n(\hat{\theta}_n, \hat{F}_{X,n})$ and $J_n(\hat{\theta}_n, \hat{F}_{X,n})$.

Proposition 3.1 below, which justifies the definitions (3.5), requires an additional regularity property on the model $\{P(\theta, F_X) : \theta \in \Theta, F_X \in \mathcal{F}_X\}$:

A4 (norm continuity of the derivative). If $F_X \Rightarrow F_{X,0}$ and $\theta \to \theta_0$, then the components $\{\eta_{\theta,F_X,j} : 1 \le j \le k\}$ of the derivative in A2 satisfy

$$\|\eta_{\theta,F_X,j} - \eta_{\theta_0,F_{X_0},j}\| \to 0, \qquad 1 \le j \le k. \tag{3.6}$$

The sufficient conditions for A2 that are given in Subsection 2.3 imply A4, by Vitali's theorem.

PROPOSITION 3.1. *Suppose that the assumptions for Proposition* 2.2 *and assumption* A4 *hold. Then, under the sequence of models* $\{P(\theta_n, F_{X,n})\}$,

$$\begin{aligned} \tilde{H}_n(\hat{\theta}_n, \hat{F}_{X,n}) &\Rightarrow H(\theta_0, F_{X,0}), \\ \tilde{J}_n(\hat{\theta}_n, \hat{F}_{X,n}) &\Rightarrow J(\theta_0, F_{X,0}) \end{aligned} \tag{3.7}$$

in probability. Moreover,

$$\tilde{\theta}_n^* - \theta_n^* = o_p(n^{-1/2}) \tag{3.8}$$

under the joint distribution of (S_n, S_n^*).

The linearization of θ_n^* and the subsequent analysis of $\tilde{\theta}_n^*$ resemble, in logical structure, LeCam's (1969) study of one-step maximum likelihood estimates. Related as well is the algorithm of Schucany and Wang (1991), who use essentially a one-step Newton-Raphson approximation in order to bootstrap estimators that are defined iteratively.

3.2 *Applications to inference*

We will now describe three statistical methods that illustrate the usefulness of Propositions 2.2 and 3.1.

A. *Testing model-fit.* Does the semiparametric model $\{P(\theta, F_X) : \theta \in \Theta, F_X \in \mathcal{F}_X\}$ fit the data? Let $c_n(\alpha)$ denote the $(1-\alpha)$-th quantile of the linearized bootstrap distribution $\tilde{J}_n(\hat{\theta}_n, \hat{F}_{X,n})$. Then the test which rejects whenever T_n exceeds $c_n(\alpha)$ has asymptotic rejection probability α under the null hypothesis that the model is correct. This conclusion follows from Propositions 2.2 and 3.1, by the reasoning for Theorem 2.1 and Example 1 in Beran (1986). The test statistic T_n can be replaced by $n^{1/2}\|\hat{\varphi}_n - \psi(\hat{\theta}_n, \hat{F}_{X,n})\|$, in view of (2.5) and (2.6). The test is consistent against all alternatives because T_n tends in probability to ∞ under each alternative, while $c_n(\alpha)$ does not.

B. *Confidence sets for* $F_{AB}(\theta)$. For the models of Examples 1 and 2 in the Introduction, the cdf of $F_{AB}(\theta)$ is not a differentiable function of θ. Natural confidence sets for $F_{AB}(\theta)$ can nevertheless be constructed in the following way: For Example 1, we devise simultaneous confidence boxes for the r sites on which the uniform probability distribution F_{AB} is supported. For Example 2, we add to these confidence boxes simultaneous confidence intervals for the site probabilities $\{p_i\}$.

Consider Example 1 with $p = q = 1$, so that each support site $s_i = (a_i, b_i)$ lies in R^2 and $\theta = (a_1, b_1, \ldots, a_r, b_r)'$ has dimension $k = 2r$. Let $U = \{\pm e_i : 1 \leq i \leq k\}$, where the $\{e_i\}$ are the standard orthonormal basis for R^k. Let $K_{n,u}(\cdot, \theta, F_X)$ denote the cdf of $\mathcal{L}[n^{1/2}u'(\hat{\theta}_n - \theta) \mid P(\theta, F_X)]$ and let $K_n(\cdot, \theta, F_X)$ denote the cdf of $\mathcal{L}\{\max_{u \in U} K_{n,u}[n^{1/2}u'(\hat{\theta}_n - \theta)] \mid P(\theta, F_X)\}$. The linearized bootstrap estimates for these two cdf's are

$$\tilde{K}_{n,u}(x, \hat{\theta}_n, \hat{F}_{X,n}) = \Pr[n^{1/2}u'(\tilde{\theta}_n^* - \hat{\theta}_n) \leq x \mid S_n] \tag{3.9}$$

and

$$\tilde{K}_n(x, \hat{\theta}_n, \hat{F}_{X,n}) = \Pr\left\{\max_{u \in U} \tilde{K}_{n,u}[n^{1/2}u'(\tilde{\theta}_n^* - \hat{\theta}_n)] \mid S_n\right\} \tag{3.10}$$

respectively.

Let

$$\tilde{d}_{n,u}(\alpha) = \tilde{K}_{n,u}^{-1}[\tilde{K}_n^{-1}(\alpha, \hat{\theta}_n, \hat{F}_{X,n}), \hat{\theta}_n, \hat{F}_{X,n}], \qquad u \in U. \tag{3.11}$$

These critical values can be approximated by a one-stage resampling algorithm (see Beran (1988), Section 2.2). A simultaneous confidence set for the components of θ is then

$$C_n = \{\theta \in \Theta : u'\theta \geq u'\hat{\theta}_n - n^{-1/2}\tilde{d}_{n,u}(\alpha)\ \forall u \in U\}, \tag{3.12}$$

which is the intersection of the confidence half-spaces

$$C_{n,u} = \{\theta \in \Theta : u'\theta \geq u'\hat{\theta}_n - u^{-1/2}\tilde{d}_{n,u}(\alpha)\}. \tag{3.13}$$

Geometrically, C_n consists of confidence boxes for the sites $\{(a_i, b_i)\}$ in R^2, centered at the estimated site values $\{(\hat{a}_{n,i}, \hat{b}_{n,i})\}$; here the pair $(\hat{a}_{n,i}, \hat{b}_{n,i})$ consists of components $2i - 1$ and $2i$ of $\hat{\theta}_n$. The sides of these confidence boxes are parallel to the two coordinate axes of R^2.

The point of construction (3.12) is twofold. First, under the assumptions for Proposition 3.1,

$$\lim_{n \to \infty} P^n_{\theta_n, F_{X,n}}[\theta_n \in C_n] = \alpha. \tag{3.14}$$

Second, the confidence set C_n is balanced in the sense that

$$\lim_{n \to \infty} P^n_{\theta_n, F_{X,n}}[\theta_n \in C_{n,u}] = K^{-1}(\alpha, \theta_0, F_{X,0}) \tag{3.15}$$

for every $u \in U$, the limit not depending on u. These two results follow from Theorem 4.1 in Beran (1988).

Geometrically, (3.15) means that the four halfspaces $C_{n,u}$ whose boundaries define the sides of the confidence box for site (a_i, b_i) each have the same asymptotic probability of containing (a_i, b_i); and that these asymptotic probabilities do not depend on i. In this way, the r confidence boxes defined by C_n give a clear picture of the relative precisions with which we can estimate the sites $\{(a_i, b_i)\}$. At the same time, the simultaneous coverage probability for θ is asymptotically α.

Confidence set C_n for θ induces a confidence set D_n for $F_{AB}(\theta)$ through

$$D_n = \{F_{AB}(\theta) \in \mathcal{F}_{AB} : \theta \in C_n\}. \tag{3.16}$$

The correspondence between C_n and D_n is one-to-one. The foregoing discussion extends in an obvious way to Example 2.

C. *Prediction sets for* Y_{n+1}. We turn to the question of predicting Y_{n+1} under the model $\{P(\theta, F_X) : \theta \in \Theta, F_X \in \mathcal{F}_X\}$, given $X_{n+1} = x$ and the learning sample S_n. Suppose $p = q = 1$. For every real x, let $G_x(\cdot, \theta)$ denote the cdf of $A_i + xB_i$ under the semiparametric model. Let

$$\begin{aligned} \hat{c}_{x,L}(\alpha) &= G_x^{-1}[(1-\alpha)/2, \hat{\theta}_n], \\ \hat{c}_{x,U}(\alpha) &= G_x^{-1}[(1+\alpha)/2, \hat{\theta}_n] \end{aligned} \tag{3.17}$$

and define the prediction interval for Y_{n+1}, given $X_{n+1} = x$, to be

$$D_{x,n} = \{y : \hat{c}_{x,L}(\alpha) \le y \le \hat{c}_{x,U}(\alpha)\}. \tag{3.18}$$

The conditional coverage probability of $D_{x,n}$ for Y_{n+1}, given $X_{n+1} = x$ and S_n, is

$$CP(D_{x,n} \mid x, S_n, \theta) = G_x[c_{x,U}(\alpha), \theta] - G_x[c_{x,L}(\alpha), \theta]. \tag{3.19}$$

The coverage probability of $D_{x,n}$ for Y_{n+1}, given $X_{n+1} = x$, is

$$\begin{aligned} CP(D_{x,n} \mid x, \theta) &= \Pr[Y_{n+1} \in D_{x,n} \mid X_{n+1} = x, \theta] \\ &= E[CP(D_{x,n} \mid x, S_n, \theta)) \end{aligned} \tag{3.20}$$

the expectation being taken with respect to the distribution of the learning sample S_n.

Suppose that the conditional cdf $G_x(t, \theta)$ is continuous in t for every $\theta \in \Theta$ and $\hat{\theta}_n$ converges in probability to θ_0. Proposition 2.1 gives sufficient conditions for the latter. Then, as n increases,

$$\begin{aligned} &CP(D_{x,n} \mid x, S_n, \theta_0) \xrightarrow{p} \alpha, \\ &CP(D_{x,n} \mid x, \theta_0) \to \alpha \end{aligned} \tag{3.21}$$

for every support point of F_{X_0}. Indeed, if $\{\theta_n\}$ is any sequence converging to θ_0, then $G_x(\cdot, \theta_n)$ converges weakly to $G_x(\cdot, \theta_0)$ because $F_{AB}(\theta_n) \Rightarrow F_{AB}(\theta_0)$ under assumption A2. The convergence is uniform in the first argument, by Polya's theorem. Thus

$$\sup_t |G_x(t, \hat{\theta}_n) - G_x(t, \theta_0)| \xrightarrow{p} 0, \tag{3.22}$$

which implies

$$
\begin{aligned}
G_x[\hat{c}_{x,L}(\alpha), \theta_0] &= (1-\alpha)/2 + o_p(1), \\
G_x[\hat{c}_{x,U}(\alpha), \theta_0] &= (1+\alpha)/2 + o_p(1)
\end{aligned} \tag{3.23}
$$

and therefore (3.21).

An important assumption in this argument is the continuity of $G_x(t, \theta)$ in t. This condition does not hold in Examples 1 and 2 of the Introduction but does hold in Example 3 if $F_{AB}(\theta)$ has a Lebesgue density for every $\theta \in \Theta$. In Examples 1 and 2, not all coverage probabilities α are possible, because $G_x(\cdot, \theta)$ is a discrete distribution. This need not be a problem when the number of support sites is moderately large.

It follows from the argument above that the prediction interval $D_{x,n}$ is probability-centered in the sense that both $\Pr[Y_{n+1} > \hat{c}_{x,u}(\alpha) \mid X_{n+1} = x, \theta_0]$ and $\Pr[Y_{n+1} < \hat{c}_{x,L}(\alpha) \mid X_{n+1} = x, \theta_0]$ converge to the common value $(1-\alpha)/2$ as sample size increases. The concept of a probability-centered prediction region can be extended to q-dimensional responses Y_{n+1} by a projection-pursuit approach (Beran (1991), Example 2).

4. Proofs

The arguments for Propositions 2.1, 2.2 and 3.1 rest upon the following purely analytical result. In the statement below, φ denotes the chf of $\mathcal{L}[(Y_i, X_i)]$, which is *not* required at this point to how a distribution within the semiparametric model $\{P(\theta, F_X) : \theta \in \Theta, F_X \in \mathcal{F}_X\}$.

PROPOSITION 4.1. *Suppose that assumptions* A1 *to* A3 *are satisfied and that* $\|\varphi - \psi(\theta_0, F_{X,0})\| \to 0$. *Then*

$$
\begin{aligned}
\inf_{\theta \in \Theta} \|\varphi - \psi(\theta, F_X)\| = &\inf_{t \in R^k} \|\varphi - \psi(\theta_0, F_X) - \langle t, \eta_0 \rangle\| \\
&+ o(\|\varphi - \psi(\theta_0, F_X)\|).
\end{aligned} \tag{4.1}
$$

PROOF. The hypothesis on φ implies that $F_X \Rightarrow F_{X,0}$ and therefore $\|\varphi - \psi(\theta_0, F_X)\| \to 0$. Let N be any neighborhood of θ_0. By the triangle inequality,

$$
\begin{aligned}
\inf_{\theta \notin N} \|\varphi - \psi(\theta, F_X)\| - \|\varphi - \psi(\theta_0, F_X)\| \geq &\inf_{\theta \notin N} \|\psi(\theta, F_X) - \psi(\theta_0, F_X)\| \\
&- 2\|\varphi - \psi(\theta_0, F_X)\|.
\end{aligned} \tag{4.2}
$$

By A1 and the convergence above, the right side of (4.2) is ultimately positive. Thus

$$
\inf_{\theta \in \Theta} \|\varphi - \psi(\theta, F_X)\| = \inf_{\theta \in N} \|\varphi - \psi(\theta, F_X)\| \tag{4.3}
$$

for all sufficiently small values of $d_0 = \|\varphi - \psi(\theta_0, F_{X,0})\|$.

Let

$$r_\theta = \psi(\theta, F_X) - \psi(\theta_0, F_X) - \langle \theta - \theta_0, \eta_0 \rangle. \tag{4.4}$$

By A2, there exists a neighborhood N_0 of θ_0 such that $\|r_\theta\| \leq 2^{-1}C|\theta - \theta_0|$ if $\theta \in N_0$ and d_0 is sufficiently small. Here C is the positive, finite constant of assumption A3. By the triangle inequality and A3,

$$\begin{aligned} &\|\varphi - \psi(\theta, F_X)\| - \|\varphi - \psi(\theta_0, F_X)\| \\ &\geq \|\langle \theta - \theta_0, \eta_0 \rangle\| - \|r_\theta\| - 2\|\varphi - \psi(\theta_0, F_X)\| \\ &\geq 2^{-1}C|\theta - \theta_0| - 2\|\varphi - \psi(\theta_0, F_X)\| \end{aligned} \tag{4.5}$$

provided $\theta \in N_0$ and d_0 is sufficiently small.

Let $d = \|\varphi - \psi(\theta_0, F_X)\|$. The right side of (4.5) is strictly positive if $|\theta - \theta_0| > 4d/C$. Thus

$$\inf_{\theta \in N_0} \|\varphi - \psi(\theta, F_X)\| \geq \inf_{|\theta - \theta_0| \leq 4d/C} \|\varphi - \psi(\theta, F_X)\| \tag{4.6}$$

for all sufficiently small d_0. In view of (4.3), (4.6), and the trivial reverse inequality,

$$\inf_{\theta \in \Theta} \|\varphi - \psi(\theta, F_X)\| = \inf_{|\theta - \theta_0| \leq 4d/C} \|\varphi - \psi(\theta, F_X)\| \tag{4.7}$$

provided d_0 is sufficiently small.

Write $\theta = \theta_0 + t$. From A2,

$$\inf_{|t| \leq 4d/C} \|\varphi - \psi(\theta_0 + t, F_X)\| = \inf_{|t| \leq 4d/C} \|\varphi - \psi(\theta_0, F_X) - \langle t, \eta_0 \rangle\| + o(d) \tag{4.8}$$

as $d_0 \to 0$. This approximation and (4.7) yield

$$\inf_{\theta \in \Theta} \|\varphi - \psi(\theta, F_X)\| = \inf_{|t| \leq 4d/C} \|\varphi - \psi(\theta_0, F_X) - \langle t, \eta_0 \rangle\| + o(d) \tag{4.9}$$

as $d_0 \to 0$.

It remains only to show that the infimum on the right side of (4.9) can be replaced by the infimum over all $t \in R^k$. Indeed, if $|t| > 4d/C$, then by A3

$$\begin{aligned} \|\varphi - \psi(\theta_0, F_X) - \langle t, \eta_0 \rangle\| &\geq \|\langle t, \eta_0 \rangle\| - \|\varphi - \psi(\theta_0, F_X)\| \\ &\geq C|t| - d > 3d \\ &\geq \inf_{t \in R^k} \|\varphi - \psi(\theta_0, F_X) - \langle t, \eta_0 \rangle\|. \end{aligned} \tag{4.10}$$

This completes the proof of Proposition 4.1.

PROOF OF PROPOSITION 2.1. Consider the empirical process

$$V_n = n^{1/2}[\hat{\varphi}_n - \psi(\theta_0, \hat{F}_{X,n})]. \tag{4.11}$$

From the definition (2.9) of W_n and A2,

$$\|V_n - W_n - n^{1/2}\langle(\theta_n - \theta_0), \eta_0\rangle\| = o(1). \tag{4.12}$$

Since $W_n \Rightarrow W_0$, the random variables $\{\|V_n\|\}$ are tight. Thus, by Proposition 4.1 and (4.12),

$$\begin{aligned} T_n &= n^{1/2} \inf_{\theta\in\Theta} \|\hat{\varphi}_n - \psi(\theta, \hat{F}_{X,n})\| \\ &= \inf_t \|V_n - \langle t, \eta_0\rangle\| + o_p(1) \\ &= \inf_t \|W_n - \langle t, \eta_0\rangle\| + o_p(1) \end{aligned} \tag{4.13}$$

as asserted in (2.13).

Let $\hat{\theta}_n$ be any minimum distance estimator and let

$$U_n = n^{1/2}\|\hat{\varphi}_n - \psi(\hat{\theta}_n, \hat{F}_{X,n})\|. \tag{4.14}$$

From (4.7) in the proof of Proposition 4.1, $\hat{\theta}_n = \theta_0 + n^{-1/2}\hat{s}_n$, where ultimately $|\hat{s}_n| \le 4\|V_n\|/C$. Equivalently, $\hat{\theta}_n = \theta_n + n^{-1/2}\tilde{s}_n$, where $\tilde{s}_n = \hat{s}_n - n^{1/2}(\theta_n - \theta_0)$. Assume without loss of generality that $W_n \to W_0$ (Skorokhod versions) and that $n^{1/2}(\theta_n - \theta_0) \to h$ (compactness). Then also $V_n \to W_0 + \langle h, \eta_0\rangle$ and, by (4.13),

$$U_n \to \|W_0 - \langle t_0, \eta_0\rangle\| \tag{4.15}$$

where t_0 is the unique value of $t \in R^k$ that minimizes $\|W_0 - \langle t, \eta_0\rangle\|$.

Now suppose that $\tilde{s}_n - \hat{t}_n \not\to 0$, for $\hat{t}_n$ defined in (2.11). Since $\hat{t}_n \to t_0$ and $|\tilde{s}_n|$ is bounded asymptotically, assume without loss of generality (compactness) that $\tilde{s}_n \to s_0 \ne t_0$. Then, from (4.14) and A2,

$$\begin{aligned} U_n &= \|V_n - \langle h + s_0, \eta_0\rangle\| + o(1) \\ &\to \|W_0 - \langle s_0, \eta_0\rangle\| \\ &> \|W_0 - \langle t_0, \eta_0\rangle\|. \end{aligned} \tag{4.16}$$

The contradiction between (4.15) and (4.16) establishes (2.12).

PROOF OF PROPOSITION 2.2. This follows from Proposition 2.1 and the weak convergence of the $\{W_n\}$ to W_0, as random elements of $L_2(Q)$.

PROOF OF PROPOSITION 3.1. Let

$$\tilde{W}_n = n^{1/2}[\hat{\varphi}_n - \psi(\theta_n, F_{X,n})] - n^{1/2}[\psi(\theta_n, \hat{F}_{X,n}) - \psi(\theta_n, F_{X,n})] \tag{4.17}$$

and let

$$\tilde{\theta}_n = \theta_n + n^{-1/2}\left[\int \mathcal{R}(\bar{\xi}_n\xi_n')dQ\right]^{-1} \int \mathcal{R}(\bar{\xi}_n\tilde{W}_n)dQ, \tag{4.18}$$

where $\xi_n = \eta_0(\theta_n, F_{X,n})$. Comparing (4.17) with (2.9) and using A2 shows that

$$\tilde{W}_n = W_n + o_p(1) \tag{4.19}$$

under $\{P(\theta_n, F_{X,n})\}$. With the help of A4 and (2.12), the asymptotic approximation

$$\tilde{\theta}_n = \hat{\theta}_n + o_p(n^{-1/2}) \tag{4.20}$$

also holds under $\{P(\theta_n, F_{X,n})\}$. The desired conclusions are implied by (4.19), (4.20) and Propositions 2.1 and 2.2.

References

Amemiya, T. (1977). A note on a heteroscedastic model, *J. Econometrics*, **6**, 365–370.

Begun, J. M., Hall, W. J., Huang, W.-M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models, *Ann. Statist.*, **11**, 432–452.

Beran, R. (1986). Simulated power functions, *Ann. Statist.*, **14**, 171–173.

Beran, R. (1988). Balanced simultaneous confidence sets, *J. Amer. Statist. Assoc.*, **83**, 679–686.

Beran, R. (1991). Designing bootstrap prediction bands, *Nonparametric Functional Estimation and Related Topics* (ed. G. Roussas), Nato ASI Series, 577–586, Kluwer, Dordrecht.

Beran, R. and Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions, *Ann. Statist.*, **20**, 1970–1984.

Beran, R. and Millar, P. W. (1991). Minimum distance estimation in random coefficient regression models (unpublished).

Chow, G. C. (1983). Random and changing coefficient models, *Handbook of Econometrics* **2** (eds. Z. Griliches and M. D. Intriligator), North-Holland, Amsterdam.

Fan, J. (1991). On the optimal rates for nonparametric deconvolution problems, *Ann. Statist.*, **19**, 1257–1272.

Goldfeld, S. M. and Quandt, R. E. (1972). *Nonlinear Methods in Econometrics*, North-Holland, Amsterdam.

Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients, *J. Amer. Statist. Assoc.*, **63**, 584–595.

Hsiao, C. (1986). *Analysis of Panel Data*, Econometric Society Monographs, Cambridge Univ. Press, Cambridge.

LeCam, L. M. (1969). *Théorie Asymptotique de la Décision Statistique*, Univ. de Montréal.

Newbold, P. (1988). Some recent developments in time series analysis—III, *Internat. Statist. Rev.*, **56**, 17–29.

Nicholls, D. F. and Pagan, A. R. (1985). Varying coefficient regression, *Handbook of Statistics* **5** (eds. E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 413–449, North-Holland, Amsterdam.

Pollard, D. (1980). The minimum distance method of testing, *Metrika*, **27**, 43–70.

Raj, B. and Ullah, A. (1981). *Econometrics, A Varying Coefficients Approach*, Croom-Helm, London.

Scheffé, H. (1959), *The Analysis of Variance*, Wiley, New York.

Schucany, W. R. and Wang, S. (1991). One-step bootstrapping for smooth iterative procedures, *J. Roy. Statist. Soc. Ser. B*, **53**, 587–596.

Spiegelman, C. (1979). On estimating the slope of a straight line when variables are subject to error, *Ann. Statist.*, **7**, 201–206.

van Es, A. J. (1991). Uniform deconvolution: nonparametric maximum likelihood and inverse estimation, *Nonparametric Functional Estimation and Related Topics* (ed. G. Roussas), NATO ASI Series, 191–198, Kluwer, Dordrecht.