RANDOM COEFFICIENT REGRESSIONS:
PARAMETRIC GOODNESS OF FIT TESTS

Pedro F. Delicado and Juan Romo

Abstract

Random coefficient regression models have been applied in different fields during recent years and they are a unifying frame for many statistical models. Recently, Beran and Hall (1992) opened the question of the nonparametric study of the distribution of the coefficients. Nonparametric goodness of fit tests were considered in Delicado and Romo (1994). In this paper we propose statistics for parametric goodness of fit tests and we obtain their asymptotic distributions. Moreover, we construct bootstrap approximations to these distributions, proving their validity. Finally, a simulation study illustrates our results.

Key words:
Goodness of fit, linear regression, random coefficient, parametric empirical processes.

1Research partially supported by DGICYT PB93-0232 (Spain).
Random coefficient regression models have been applied in different fields during recent years; their general form is

\[ Y_i = A_i + X_i B_i, \quad i \geq 1, \]  

(1.1)

where \( Y_i \) and \( A_i \) are \( p \times 1 \) random variables, \( B_i \) is a \( q \times 1 \) random variable and \( X_i \) is a \( p \times q \) random matrix. The triples \( \{(A_i, B_i, X_i) : i \geq 1\} \) are independent and identically distributed and, for each \( i \), \( (A_i, B_i) \) is independent of \( X_i \). The distribution of \( (A_i, B_i, X_i) \) is not known and we can observe the \( n \) pairs \( (Y_i, X_i) \), \( 1 \leq i \leq n \). These models include well known situations as random effects in ANOVA (see, e.g., Scheffé (1959)), deconvolution models (Fan (1991), van Es (1991)), location-scale mixture models or some heteroscedastic linear models. Their applications can be found in several fields (biology, econometrics, image compression) and Raj and Ullah (1981), Chow (1983), Nicholls and Quinn (1982), and Nicholls and Pagan (1985) survey this work. All this literature is focused on moments estimation, essentially mean and variance.

Beran and Hall (1992) began a nonparametric approach by considering the estimation of the joint distribution \( F_{AB} \) of the random parameter \((A, B)\). Beran (1991) introduced a minimum distance estimate and constructed prediction intervals for \( Y \). Beran and Millar (1991) construct a \( n^{1/2} \)-consistent minimum distance estimate of the coefficient distribution. Delicado and Romo (1994) present goodness of fit tests, obtain their asymptotic distributions under the null hypothesis and propose bootstrap approximations, proving their asymptotic validity.

In this paper, we study whether the distribution \( F_{AB} \) belongs to a parametric family \( \{F_{\theta} \mid \theta \in \Theta \subseteq \mathbb{R}^N\} \). The article is organized as follows. First, Section 1.1 contains the framework and the preliminary results needed for the rest of the paper. Section 2 gives the asymptotic distributions of the test statistics. In Section 3 we provide a bootstrap resampling strategy to approach these distributions and we prove their validity. Finally, a simulation experiment is carried out in Section 4 to check the performance of these tests.

1.1 Preliminaries

In model (1.1), the joint distribution \( F_{YX} = \mathcal{P}(F_{AB}, F_X) \) of \((Y_i, X_i)\) depends on both the distribution \( F_{AB} \) and the distribution \( F_X \) of \( X_i \). Let \( P_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{(Y_i, X_i)} \) and \( F_{X,n} = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i} \) be the empirical distributions associated to the observations \((Y_i, X_i)\) and \( X_i \), respectively. For the parametric family \( \{F_{\theta} \mid \theta \in \Theta \subseteq \mathbb{R}^N\} \), we want to study the test

\[
\begin{align*}
H_0 &: F_{AB} \in \{F_{\theta} \mid \theta \in \Theta \subseteq \mathbb{R}^N\} \\
H_1 &: F_{AB} \notin \{F_{\theta} \mid \theta \in \Theta \subseteq \mathbb{R}^N\}.
\end{align*}
\]  

(1.2)
We will assume identifiability in model (1.1), i.e., $\mathcal{P}(F_{AB}, F_X) = \mathcal{P}(\hat{F}_{AB}, F_X)$ implies $F_{AB} = \hat{F}_{AB}$; sufficient conditions for identifiability were given by Beran and Millar (1991). To construct the corresponding statistics, we will use the empirical processes

$$D_n = \sqrt{n} \left( P_n - \mathcal{P}(F_{AB}, F_X) \right),$$
$$J_n = \sqrt{n} \left( P_n - \mathcal{P}(F_{AB,n}, F_X) \right),$$

and

$$\hat{J}_n = \sqrt{n} \left( P_n - \mathcal{P}(F_{\hat{B}}, F_{X,n}) \right),$$

where $\hat{\theta}_n$ is an estimator of the true value $\theta_0$ in the parametric family $\{F_\theta | \theta \in \Theta \subseteq \mathbb{R}^n\}$. Consider the class $\mathcal{I} = \{I_{st} = (-\infty, s] \times (-\infty, t] : s \in \mathbb{R}^p, t \in \mathbb{R}^q \}$ of $(p + pq)$-dimensional semiintervals. Given $s \in \mathbb{R}^p, t \in \mathbb{R}^q$,

$$J_n(s, t) = J_n(I_{st}) = \sqrt{n} \left( \int_{\mathbb{R}^{p+q}} I_{(-\infty, I_{st})}(y, x) dP_n(y, x) - \int_{\mathbb{R}^{p+q}} F_{AB}(A + xB \leq s) I_{(-\infty, 0]}(x) dF_{X,n}(x) \right) = \sqrt{n} \int_{\mathbb{R}^{p+q}} (I_{(-\infty, 0]}(y) - F_{AB}(A + xB \leq s)) I_{(-\infty, 0]}(x) dP_n(y, x) = \sqrt{n} \int_{\mathbb{R}^{p+q}} f_{st}(y, x) dP_n(y, x) = \sqrt{n} P_n(f_{st}),$$

where $f_{st}(y, x) = (I_{(-\infty, 0]}(y) - F_{AB}(A + xB \leq s)) I_{(-\infty, 0]}(x)$. Observe that $\mathcal{P}(F_{AB}, F_X)(f_{st}) = 0$. Thus, it turns out that, for each $s, t$, $J_n(s, t) = D_n(f_{st})$. Consider $\mathcal{F} = \{f_{st} \mid s \in \mathbb{R}^p, t \in \mathbb{R}^q\}$. We refer to Giné and Zinn (1986) for the definitions of the classes $\mathcal{F}^1, \mathcal{F}^2, \mathcal{FF},$ and for the notions of weak convergence in $l^\infty(\mathcal{F})$, Vapnik-Cervonenkis and Donsker classes of functions.

An envelope for a class $\mathcal{F}$ of functions $f : \mathcal{X} \rightarrow \mathbb{R}$ is a measurable function $F$ such that $|f(x)| \leq F(x)$ for every $x \in \mathcal{X}$ and $f \in \mathcal{F}$. If $A \subseteq \mathcal{X}$ is a finite set and $\varepsilon > 0$, let

$$D(\varepsilon, A, F, \mathcal{F}) = \min\{k \mid \text{there exist } f_1, \ldots, f_k \in \mathcal{F} \text{ such that } \sup_{f \in \mathcal{F}} \min_{1 \leq i \leq k} \sum_{x \in A} (f(x) - f_i(x))^2 \leq \varepsilon^2 \sum_{x \in A} F(x)^2\},$$

and define $D(\varepsilon, F, \mathcal{F}) = \sup\{D(\varepsilon, A, F, \mathcal{F}) \mid A \subseteq \mathcal{X}, A \text{ finite}\}$, where $F$ is an envelope for $\mathcal{F}$. The corresponding entropy is $H(\varepsilon, F, \mathcal{F}) = \log D(\varepsilon, F, \mathcal{F})$.

We will use the following hypotheses in some of our results:

(a) $P_{F_{AB}}(A + xB = s) = 0$ for all $x \in \text{Supp}(X)$ and all $s \in \mathbb{R}^p$, and

(b) the distribution of $(A, B)$ is discrete.

The next theorem gives the asymptotic behavior of $D_n$ in the case of one-dimensional dependent variable $Y$. 

2
Theorem 1.1. (Delicado and Romo, 1994). If either (a) or (b) hold then
\[ D_n \to_w Z_{P(F_{AB}, F_X)} \] (1.3)
in \( L^\infty(\mathcal{F}) \), where \( Z_{P(F_{AB}, F_X)} \) is the \( P(F_{AB}, F_X) \)-brownian bridge with covariance structure given by
\[
\text{Cov}(f_{st}, f_{wu}) = \int_{(s \leq u)} (P_{F_{AB}}(A + xB \leq s \wedge u) - P_{F_{AB}}(A + xB \leq s)P_{F_{AB}}(A + xB \leq u)) \, dF_X(x).
\]
If the dimension of \( Y \) is larger than one, the following result provides conditions guaranteeing the convergence of the empirical processes.

Theorem 1.2. (Delicado and Romo, 1994). Assume that \( \text{Supp}(X) \) is compact and that \( Y_x = A + xB \) is absolutely continuous for all \( x \in \text{Supp}(X) \). Suppose also that the function \( h(x, s) = P_{F_{AB}}(A + xB \leq s) \) has uniformly bounded partial derivatives:
\[
\left| \frac{\partial h}{\partial x}(x, s) \right| \leq M_1, \quad x \in \mathbb{R}^p, \quad s \in \mathbb{R}^p,
\]
\[
\left| \frac{\partial h}{\partial s}(x, s) \right| \leq M_2, \quad x \in \mathbb{R}^p, \quad s \in \mathbb{R}^p.
\]

Then
(i) The family of probability measures \( \{ P_{Y_x}, x \in \text{Supp}(X) \} \) is tight: for all \( \varepsilon > 0 \) there exists a compact \( C(\varepsilon) \) such that \( P_{Y_x}(C(\varepsilon)) \geq 1 - \varepsilon, \quad x \in \text{Supp}(X). \) \( C(\varepsilon) \) can be chosen to be of the form
\[
C(\varepsilon) = [I(\varepsilon), u(\varepsilon)]_p = \{ s \in \mathbb{R}^p \mid I(\varepsilon) \leq s \leq u(\varepsilon) \}.
\]
(ii) \( D(\varepsilon, F, \mathcal{F}_p) \leq p^r \left( \frac{1}{2M_2} \right)^{-r} \text{Vol}([I(\varepsilon), u(\varepsilon)]_p). \)
(iii) If
\[
\int_0^1 \log \left( \text{Vol}([I(\varepsilon/2), u(\varepsilon/2) + 1_p \varepsilon/2M_2]) \right) \, d\varepsilon < \infty \tag{1.4}
\]
then \( \int_0^1 H(\varepsilon, F, \mathcal{F}_p) \, d\varepsilon < \infty \) and also \( \int_0^1 H(\varepsilon, F, \mathcal{F}) \, d\varepsilon < \infty \), where \( F = 1. \)

Finally, we recall three results on weak convergence. The first two relate weak and uniform convergence for a measure \( \mu \) and a sequence \( \{ \mu_n : n \in \mathbb{N} \} \) of measures defined on the Borel \( \sigma \)-algebra \( \mathcal{B} \) in a metric space \( \mathcal{X} \).

Theorem 1.3. (Ranga Rao, 1962). \( \{ \mu_n : n \in \mathbb{N} \} \) converges weakly to \( \mu \) if, and only if, for any uniformly bounded and equicontinuous class of functions \( \mathcal{F} \) defined on \( \mathcal{X} \),
\[
\lim_{n \to \infty} \sup_{f \in \mathcal{F}} \int fd\mu_n - \int fd\mu = 0.
\]
THEOREM 1.4. (Billingsley and Topsøe, 1967). Let $\mathcal{F}$ be a class of continuous functions on $X$. Then

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}} \left| \int f d\mu_n - \int f d\mu \right| = 0$$

if, and only if,

(i) $\omega_F(X) < \infty$, and

(ii) $\lim_{\varepsilon \to 0} \sup_{f \in \mathcal{F}} P(\varepsilon) > \varepsilon = 0$, for all $\varepsilon > 0$,

where $\omega_F(B) = \sup\{\|f(x) - f(y)\|, f \in \mathcal{F}, x, y \in B\}$ for $B \in \mathcal{B}$, $\omega_f(B) = \omega_{\{f\}}(B)$ and $S(x, \varepsilon)$ is the ball with center $x$ and radius $\varepsilon$.

The next result is due to Rubin and it can be seen in Billingsley (1968), Theorem 5.5.

THEOREM 1.5. Let $\{h_n : n \in \mathbb{N}\}$ and $h$ be measurable functions from $X$ to the metric space $X'$. Let $E = \{x \in X | \exists \{x_n\} \to x$, and $h_n(x_n) \to h(x)\}$. Assume that $E$ is a measurable subset of $X$ with $\mu(E) = 0$. Then $\mu_n \to \mu$ implies that $\mu_n h_n \to \mu h$.

2 Asymptotic distribution of the test statistics

To obtain the asymptotic distribution of the statistics used to test if $F_{AB}$ belongs to the parametric family $\{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^N\}$, we will assume that the model is continuous with respect to the parameter, i.e., $F_{\theta_n} \to F_{\theta_0}$ if $\theta_n \to \theta_0$, $\theta_0 \in \Theta$.

Let $\hat{\theta}_n$ be an estimate of $\theta_0$ based on the sample $(Y_i, X_i), i = 1, \ldots, n$, from model (1.1) with $p \geq 1$. Our goodness of fit statistics will be of either Kolmogorov-Smirnov or Cramér-von Misses type based on the empirical process

$$\hat{F}_n = \sqrt{n} \left( F_n - F_{\hat{\theta}_n, X_n} \right).$$

Let us recall the definitions of regular families and estimators (see, e.g., Shorack and Wellner (1986), p. 229).

DEFINITION 2.1. Let $\Delta = (\Delta_1, \ldots, \Delta_N)$ with $\Delta_j(\theta, x) = \hat{\partial}F_\theta/\hat{\partial} \theta_j$, $j = 1, \ldots, N$. We say that $\{F_\theta : \theta \in \Theta \subseteq \mathbb{R}^N\}$ in $\mathbb{R}^d$ is regular in $\theta$ if

$$F_{\theta}(z) = F_\theta(z) + \sum_{j=1}^N (\theta_j - \theta_j) \Delta_j(\theta, z) + R(\theta, \theta', z)$$

where $\|R(\theta, \theta', \cdot)\|_\infty = o(\|\theta - \theta'\|_2)$ and $\Delta_j, j = 1, \ldots, N$ are uniformly bounded in $z$.

Essentially, $\{F_\theta\}$ is regular if $F_\theta(z)$ is uniformly differentiable with respect to $\theta$. If $F_\theta(z)$ has bounded partial derivatives with respect to $\theta$ and the function $(\partial/\partial \theta)F(\cdot)$ is uniformly continuous (or even $\{(\partial/\partial \theta)F(z), z \in \mathbb{R}^p\}$ is equicontinuous) then $\{F_\theta\}$ is regular (see Pollard (1984), p. 119).
DEFINITION 2.2. Let \( \{Z_i : i \in \mathbb{N} \} \) be independent random variables with distribution \( F_\theta \).

The sequence of estimators \( \{\hat{\theta}_n : n \in \mathbb{N} \} \) is regular if, for all \( j \),
\[
\sqrt{n}(\hat{\theta}_{nj} - \theta_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(\theta, Z_i) + o_P(1),
\]
where \( E h_j(\theta, Z_i) = 0 \) and \( V(h_j(\theta, Z_i)) = \sigma_j^2 \).

Efficient estimators and M-estimators, for instance, are regular ones.

For \( x \in \mathbb{R}^p \), let \( F(x, \theta, s) = F(A + xB \leq s) \) be the distribution function of \( Y_x = A + xB \) when \( (A, B) \sim F_\theta \), with partial derivatives \( \Delta_j(x, \theta, s) = \partial F(x, \theta, s)/\partial \theta_j \). Let us denote by \( \theta_0 \) the true value of \( \theta \). Define the class \( \mathcal{H} = \{ h_n : \mathbb{R}^{p+1} \to \mathbb{R} \ | \ h_n(y, x) = \sum_{j=1}^{N} L_j(s, t, \theta_0) h_j(\theta_0, y, x), s \in \mathbb{R}^p, t \in \mathbb{R}^q \} \), where \( L_j(s, t, \theta) = E[\Delta_j(X, \theta, s)I(\theta_0, t)(X)] \) and \( h_j \) are the functions in the definition of regular estimators for the parameter \( \theta \).

THEOREM 2.1. Suppose \( \{ (Y_i, X_i), i \geq 1 \} \) are i.i.d. variables from model (1.1) with \( F_{AB} = F_{\theta_0} \). Let \( \{ F(x, \theta, s) \} \) be regular in \( \theta \) for all \( x \) and let \( \hat{\theta}_n = \theta(Y_1, X_1, Y_2, X_2, \ldots, Y_n, X_n) \) be a regular sequence of estimators of \( \theta \). Assume that \( \Delta_j(X_i, \theta, s) \) has finite expectation uniformly bounded in \( s \) and that
\[
\frac{1}{n} \sum_{i=1}^{n} \Delta_j(X_i, \theta_0, s) I(\theta_0, t)(X_i) - L_j(s, t, \theta_0) = o_P(1), \quad j = 1, \ldots, N,
\]
uniformly in \( s \) and \( t \). Suppose that hypotheses in either Theorem 1.1 or 1.2 hold for either \( p = 1 \) or \( p > 1 \), respectively. Then \( F_{\hat{\theta}_n} \) is a Donsker class for \( P(F_{\theta_0}, F_X) \) and
\[
D_n \to \mathcal{Z}_{P(F_{\theta_0}, F_X)} \text{ in } L^\infty(\mathcal{F} + \mathcal{H}),
\]
where \( \mathcal{Z}_{P(F_{\theta_0}, F_X)} \) is the brownian bridge with the corresponding covariance structure determined by \( P(F_{\theta_0}, F_X) \). Moreover, \( D_n(s, t + h_n) = \hat{J}_n(s, t) + o_P(1), \) for \( s \in \mathbb{R}^p, t \in \mathbb{R}^q \) uniformly in \( s \) and \( t \).

PROOF: The regularity of \( \{ F(x, \theta, s) \} \) and the linearity of \( P \) imply that
\[
P(F_{\hat{\theta}_n}, F_X, s, t) = \frac{1}{n} \sum_{i=1}^{n} P(F_{\hat{\theta}_n}, \delta_{X_i})(s, t) = \frac{1}{n} \sum_{i=1}^{n} F(X_i, \hat{\theta}_n, s) I_{(\theta_0, t]}(X_i) =
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} F(X_i, \theta_0, s) I_{(\theta_0, t]}(X_i) +
\]
\[
+ \sum_{j=1}^{N} (\hat{\theta}_{nj} - \theta_{nj}) \left( \frac{1}{n} \sum_{i=1}^{n} \Delta_j(X_i, \theta_0, s) I_{(\theta_0, t]}(X_i) \right) +
\]
\[
+ \frac{1}{n} \sum_{i=1}^{n} R(X_i, \theta_0, \hat{\theta}_n, s) I_{(\theta_0, t]}(X_i) =
\]

\[ \mathcal{P}(F_{\theta_0}, F_{X,n})(s, t) + \sum_{j=1}^{N} (\hat{\theta}_{n_j} - \theta_{0j}) E \left[ \Delta_j(X, \theta_0, s) I_{(\infty, \theta_j]}(X) \right] + \frac{1}{n} \sum_{i=1}^{n} R(X_i, \theta_0, \hat{\theta}_n, s) I_{(\infty, \theta_j]}(X_i), \]

where the last equality follows because \( E \Delta_j \) is finite. Thus,

\[
\hat{J}_n(s, t) = \hat{J}_n(I_{st}) = \sqrt{n} \left( P_n - \mathcal{P}(F_{\theta_0}, F_{X,n}) \right)(s, t) = \\
= \sqrt{n} \left( P_n - \mathcal{P}(F_{\theta_0}, F_{X,n}) \right)(s, t) - \sum_{j=1}^{N} \sqrt{n} R(X_i, \theta_0, \hat{\theta}_n, s) I_{(\infty, \theta_j]}(X_i), \]

(2.5)

The regularity of \( \{\hat{\theta}_n\} \) allows us to study the convergence of the terms in (2.5): firstly, the third term is \( o_p(1) \) uniformly in \( s \) and \( t \) because \( \sqrt{n}(\hat{\theta}_{n_j} - \theta_{0j}) = o_p(1) \); second,

\[
\sup_{s,t} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{n} R(X_i, \theta_0, \hat{\theta}_n, s) I_{(\infty, \theta_j]}(X_i) \right| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} o_p(1) = o_p(1),
\]

since \( \sqrt{n}(\hat{\theta}_n - \theta_0) = o_p(1) \); and finally,

\[
\sqrt{n}(\hat{\theta}_{n_j} - \theta_{0j}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(\theta_0, Y_i, X_i) + o_p(1) = \\
= \sqrt{n} \left( P_n - \mathcal{P}(F_{\theta_0}, F_X) \right)(h_j(\theta_0, \cdot, \cdot)) + o_p(1) = D_n(h_j(\theta_0, \cdot, \cdot)) + o_p(1),
\]

because \( \mathcal{P}(F_{\theta_0}, F_X)(h_j(\theta_0, \cdot, \cdot)) = 0 \).

By the uniform boundedness of \( E \Delta_j \), \( \|L_j(s, t, \theta_0)\|_\infty = \sup_{s,t} |L_j(s, t, \theta_0)| < \infty \) and so

\[
\hat{J}_n(s, t) = J_n(s, t) + \sum_{j=1}^{N} L_j(s, t, \theta_0) D_n(h_j(\theta_0, \cdot, \cdot)) + \sum_{j=1}^{N} L_j(s, t, \theta_0) o_p(1) + o_p(1) = \\
= D_n(f_{st} + h_{st}) + o_p(1)
\]

where \( h_{st} = \sum_{j=1}^{N} L_j(s, t, \theta_0) h_j(\theta_0, \cdot, \cdot) \in \mathcal{H} = \{ h : \mathbb{R}^{p+q} \rightarrow \mathbb{R} | h(y, x) = \sum_{j=1}^{N} L_j(s, t, \theta_0) h_j(\theta_0, y, x), s \in \mathbb{R}^p, t \in \mathbb{R}^q \} \subset \mathcal{H}' = \{ h : \mathbb{R}^{p+q} \rightarrow \mathbb{R} | h(y, x) = \sum_{j=1}^{N} \alpha_j h_j(\theta_0, y, x), \alpha_j \in \mathbb{R} \}. \)

\( \mathcal{H} \) is a Vapnik-Červonenkis class of functions because the dimension of \( \mathcal{H}' \) is finite (see, e.g., Pollard (1984), p. 30). Delicado and Romo (1994) obtain a bound for the entropy of the class \( \mathcal{F} \), so there exist positive constants \( A \) and \( w \) (see, e.g., Pollard (1984), p. 40, and Chapter II, Lemma 36) such that

\[
D(\varepsilon, F + H, \mathcal{F} + \mathcal{H}) < A \varepsilon^{-w},
\]

6
where \( F = 1 \) and \( H(x, y) = \sum_{j=1}^{N} \| L_j(s, t, \theta_0) \|_{\omega} \) are envelopes for \( \mathcal{F} \) and \( \mathcal{H} \), respectively.

Obviously, the function \( H \) has finite second moment. Moreover, since the functions \( h_j \) are measurable, the class \( \mathcal{H} \) is permissible (see Pollard (1984), p. 196). Proposition 2.1 in Delicado and Romo (1994) gives that \( \mathcal{F} \) is permissible under the conditions established in Section 1. So, the central limit theorem in Pollard (1982) leads to

\[
D_n \to_{w} Z_{\mathcal{P}(F_{\theta_0}, F_X)}
\]

in \( l^\infty(\mathcal{F} + \mathcal{H}) \).

The following corollary gives the asymptotic behavior of both the Kolmogorov-Smirnov \( K_n \) and Cramér-von Mises \( M_n \) test statistics.

**Corollary 2.1.** For the statistics

\[
\hat{K}_n = \sup_{s,t} |\hat{F}_n(s,t)| \quad \text{and} \quad \hat{M}_n = \left( \int_{\mathbb{R}^{+\times}} (\hat{J}_n(s,t))^2 \, dQ(s,t) \right)^{1/2},
\]

it holds that

\[
\hat{K}_n \to_{w} \| Z_{\mathcal{P}(F_{\theta_0}, F_X)} \|_{\mathcal{F} + \mathcal{H}}, \quad \text{and} \quad \hat{M}_n \to_{w} \| Z_{\mathcal{P}(F_{\theta_0}, F_X)} \|_{2Q}.
\]

### 3 Bootstrap approximations

It turns out that the asymptotic distributions obtained in Section 2 are not easy to handle. Resampling techniques—and, in particular, the bootstrap—provide a way to overcome this problem.

A straightforward bootstrap scheme based on resampling the pairs \((Y_i, X_i)\) cannot be implemented here because the functions \( f_{st} + h_{st} \) are not known: \( h_{st} \) depends on the true value \( \theta_0 \) of the parameter \( \theta \). However, since \( H_0 \) gives a parametric specification of the distribution of \((A, B)\), we propose a resampling strategy based on parametric bootstrap.

The algorithm for the bootstrap hypothesis test of \( H_0 \) using, e.g., \( K_n \) is the following:

1. Obtain the value \( \hat{K}_n = \sup_{s,t} |\hat{J}_n(s,t)| \) and estimate \( \theta \) by a regular statistic \( \hat{\theta}_n \) from the sample \((Y_i, X_i), i = 1, \ldots, n\).
2. Generate \((Y_i^*, X_i^*), i = 1, \ldots, n\) using parametric bootstrap:
   1. Obtain a sample \((A_i^*, B_i^*), i = 1, \ldots, n\) from \( F_{\hat{\theta}_n}\).
   2. Obtain a sample \(X_i^*, i = 1, \ldots, n\) from \( F_{X,\hat{\theta}_n}\).
2.3. Calculate the pseudo-values \( Y_i^* = A_i^* + X_i^* B_i^* , i = 1, \ldots , n \).

3. Compute the bootstrap empirical process for \( \hat{J}_n \): 
\[
\hat{J}_n^* = \sqrt{n} \left( \hat{P}_n^* - \mathcal{P}(F_{\hat{\theta}_n}, \hat{F}_{X,n}) \right)
\]
where \( \hat{P}_n^* \) and \( \hat{F}_{X,n}^* \) are, respectively, the empirical distributions obtained from \( (Y_i^*, X_i^*) \) and \( X_i^* , i = 1, \ldots , n \), and \( \hat{\theta}_n^* \) is the bootstrap replication of the estimate obtained from \( (Y_i^*, X_i^*) \).

4. Calculate \( \bar{K}_n^* \) from \( \hat{J}_n^* \).

5. Iterate steps 2, 3 and 4 \( B \) times to have \( \bar{K}_{n,b}^* , b = 1, \ldots , B \).

6. Reject \( H_0 \) if \( \bar{K}_{n,b}^* \) is larger than the \( \alpha \)-th upper quantile of the distribution of \( \bar{K}_{n,b}^* , b = 1, \ldots , B \).

Our next Proposition gives the relationship between the bootstrap versions of \( J_n \) and \( D_n \):
\[
\hat{J}_n^* = \sqrt{n} \left( \hat{P}_n^* - \mathcal{P}(F_{\hat{\theta}_n}, \hat{F}_{X,n}) \right)
\] and 
\[
\hat{D}_n^* = \sqrt{n} \left( \hat{P}_n^* - \mathcal{P}(F_{\hat{\theta}_n}, \hat{F}_{X,n}) \right).
\]
This is the key step that will allow us to establish below the validity of our bootstrap statistics.

**Proposition 3.1.** Suppose that the conditions in Theorem 2.1 hold and that the functions \( L_j(\theta, s, t) \) are derivable with respect to \( \theta \) around the true value \( \theta_0 \). Assume that the functions \( h_j(\theta, y, x), j = 1, \ldots , N \) are derivable with respect to \( \theta \) in a neighborhood of \( \theta_0 \) and the derivative at \( \theta = \theta_0 \) is continuous and bounded as a function of \( (y, x) \) except, possibly, in a null measure set. Then, for all \( s \in \mathbb{R}^a \), \( t \in \mathbb{R}^q \),
\[
\hat{J}_n^*(s, t) = \hat{D}_n^*(f_{st} + h_{st}) + o_P(1),
\]
uniformly in \( s \) and \( t \).

**Proof:** We have that
\[
\hat{D}_n^*(f_{st} + h_{st}) = \hat{D}_n^*(f_{st}) + \hat{D}_n^*(h_{st}).
\]
In the proof of Theorem 2.2 in Delicado and Romo (1994) it is shown that
\[
\hat{D}_n^*(f_{st}) = \sqrt{n} \left( \hat{P}_n^* - \mathcal{P}(F_{\hat{\theta}_n}, \hat{F}_{X,n}) \right)(s, t).
\]
Let us consider now the second term:
\[
\hat{D}_n^*(h_{st}) = \sqrt{n} \left( \hat{P}_n^* - \mathcal{P}(F_{\hat{\theta}_n}, \hat{F}_{X,n}) \right)(h_{st})
\]
\[
= \sqrt{n} \left( \hat{P}_n^* - \mathcal{P}(F_{\hat{\theta}_n}, \hat{F}_{X,n}) \right) \left( \sum_{j=1}^{N} l_j(s, t, \theta_0) h_j(\theta_0, Y, X) \right).
\]
The hypothesis on the derivatives of \( h_j \) allows us to approximate this last expression by using a one term Taylor expansion around \( \theta_0 \) to obtain
\[
\sum_{j=1}^{N} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_j(\hat{\theta}_n, Y_i^*, X_i^*) \right) L_j(s, t, \theta_0) - \sum_{j=1}^{N} L_j(s, t, \theta_0) \sqrt{n} \mathcal{P}(F_{\hat{\theta}_n}, F_{X,n})(h_j(\hat{\theta}_n, Y, X)) = 8
\]
\[ - \sum_{j=1}^{N} L_{j}(s, t, \theta_0) \sqrt{n(\hat{\theta}_n - \theta_0)^{T} \left( \hat{P}_n - \mathcal{P}(F_{s,t}, F_{X,n}) \right) H_{j}(\theta_0, Y, X)} + o_p(1) \]  

where \( H_{j} \) is the vector of partial derivatives of \( h_{j} \) with respect to \( \theta \) at \( \theta_0 \). By the definition of regular estimator, the second term is zero (the distribution of \( h_{j} \) has zero mean and variance equals to \( \sigma_{0j}^{2}(\theta_0, F_{X}) \)) if \((Y, X) \sim \mathcal{P}(F_{s,t}, F_{X})\).

Since \( \hat{\theta}_n \longrightarrow \theta_0 \), the continuity of the distributions with respect to \( \theta \) gives that for any distance \( d \) metrizing weak convergence, \( d(F_{\hat{\theta}_n}, F_{\theta_0}) \longrightarrow p 0 \). Proposition 2.1 in Beran and Millar (1991) and the fact that \( d(F_{X,n}, F_{X}) \longrightarrow a.s. 0 \) imply that \( d(\mathcal{P}(F_{\hat{\theta}_n}, F_{X,n}), \mathcal{P}(F_{\theta_0}, F_{X})) \) converges to zero in probability. By Theorem 2 in Beran, Le Cam, and Millar (1987) it follows that \( d(\hat{P}_n, \mathcal{P}(F_{\theta_0}, F_{X})) \longrightarrow p 0 \) and the hypothesis on the derivatives of \( h_{j} \) implies that \( H_{j} \) is continuous and bounded a.e. if \((Y, X) \sim \mathcal{P}(F_{s,t}, F_{X})\); so, \( \left( \hat{P}_n - \mathcal{P}(F_{\hat{\theta}_n}, F_{X,n}) \right) H_{j}(\theta_0, Y, X) \) converges to zero in probability. This proves that the second factor in (3.6) is \( o_p(1) \). Since the first factor is \( O_p(1) \), the whole term is \( o_P(1) \). Noting that \( \hat{\theta}_n = T_n(Y_1, Y_2, Y_3, \ldots, Y_n, X_n) \) conditioned on the sample is a regular estimator of \( \theta_0 \), we obtain that

\[ \hat{D}_n^*(h_{st}) = \sum_{j=1}^{N} \sqrt{n(\hat{\theta}_{nj} - \theta_{nj})} L_{j}(s, t, \theta_0) + o_P(1). \]

Using now the condition on \( L_{j} \),

\[ \hat{D}_n^*(h_{st}) = \sum_{j=1}^{N} \sqrt{n(\hat{\theta}_{nj} - \theta_{nj})} L_{j}(s, t, \theta_0) - \sum_{j=1}^{N} \sqrt{n(\hat{\theta}_{nj} - \theta_{nj})}(\hat{\theta}_n - \theta_0)^{T} \frac{\partial L_{j}}{\partial \theta}(s, t, \theta_0) + o_p(1) = \]

\[ = \sum_{j=1}^{N} \sqrt{n(\hat{\theta}_{nj} - \theta_{nj})} L_{j}(s, t, \theta_0) + o_p(1) = \]

\[ = \sum_{j=1}^{N} \sqrt{n(\hat{\theta}_{nj} - \theta_{nj})} E[\Delta_{j}(\hat{X}_{n}, \hat{\theta}_n, s) I_{(-\infty, t]}(X)] + o_p(1) = \]

\[ = \sum_{j=1}^{N} \sqrt{n(\hat{\theta}_{nj} - \theta_{nj})} \frac{1}{n} \sum_{i=1}^{n} \Delta_{j}(X_{i}, \hat{\theta}_n, s) I_{(-\infty, t]}(X_{i}) + o_p(1). \]

By the regularity of the distribution family and reasoning as in the first part of the proof of Theorem 2.1, we get \( \hat{D}_n^*(f_{st} + h_{st}) = \sqrt{n} \left( \hat{P}_n - \mathcal{P}(F_{s,t}, F_{X}) \right) \) \( (s, t) = \hat{J}_n^*(s, t) + o_P(1) \). \( \square \)

The following two theorems provide the asymptotic validity of the bootstrap version \( \hat{D}_n^* \) under two different sets of assumptions. Consider the following hypotheses:

(i) \( h_{j}(\theta_0, y, x), j = 1, \ldots, N \) are bounded in \((y, x)\) and continuous a.e. \( \mathcal{P}(F_{\theta_0}, F_{X}) \).

(ii) Let \( \mathcal{G}(a, b) = \{ g : R^0 \longrightarrow R | g(x) = f(a + xb, x), f \in \mathcal{F} \} \). As in Proposition 2.2 in Delicado and Romo (1994), it can be shown that \( \mathcal{G}(a, b) \) (and also the classes \( \mathcal{G}(a, b)h_{j} \))
\( j = 1, \ldots, N \) is a Donsker class for \( F_X \). So, the functions \( \psi_n^j(a, b) = \| F_{X,n} - F_X \|_{\mathcal{G}(a, b)} \) tend to zero in probability for all \( (a, b) \). Let \( \Omega^i = \{ (a, b) \mid \exists (a_n, b_n) \to (a, b) \text{ and } \psi_n^j(a_n, b_n) \neq 0 \} \). Assume that \( P_{F_0}(\Omega^i) = 0 \).

(ii) With the notation in (ii), \( \psi_n^j \) tends to zero uniformly over compact sets, \( j = 1, \ldots, N \).

(iii) \( h_j, j = 1, \ldots, N \) are uniformly continuous.

(iii') \( h_j, j = 1, \ldots, N \), are uniformly continuous except in a set whose closure has null probability under \( P(F_{\infty}, F_X) \): there exists \( \rho(h) \) such that \( P(F_{\infty}, F_X)(f_7; \rho) = 0 \) and for all \( \epsilon > 0 \) there exists \( \delta > 0 \) so that if \( (y_0, x_0) \notin \Gamma_{h_j} \) and \( (y', x') \in S((y_0, x_0), \delta) \) then \( |h_j(y, x) - h_j(y', x')| < \epsilon \).

(iv) For all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that \( \|(a, b) - (a', b')\| < \delta \) implies \( P_{F_X}(\{ x \mid a + xb \leq s, a' + xb' \leq s \}) < \epsilon \) for all \( s \in \mathbb{R}^p \).

(v) \( \mathcal{A} = \{ \alpha_s(x) = F(x, \theta_0, s), s \in \mathbb{R}^p \} \) is uniformly equicontinuous: for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that if \( \| x - x' \| < \delta \) then \( |\alpha_s(x) - \alpha_s(x')| < \epsilon \), for all \( s \in \mathbb{R}^p \).

(vi) \( P(F_{\infty}, F_X) \) is absolutely continuous with density function \( f_{P(F_{\infty}, F_X)} \) and there exists a largest mode where the density achieve its maximum.

**Theorem 3.1.** Suppose that the conditions in Theorem 2.1 and Proposition 3.1 hold. If (i), either (ii) or (ii)', (iii) and (iv) are true then

\[ \hat{D}_n^* \to_w Z_{P(F_{\infty}, F_X)} \text{ in probability.} \]

**Proof:** The proof relies on Corollary 2.7 in Giné and Zinn (1991). From that result, it follows that if \( \{ R_n \} \) are random measures and \( \| R_n - R_0 \|_\mathcal{G} \to 0 \) almost surely, then \( d(R_{n(h)}^{R_k}, Z_{R_0}) \to 0 \) almost surely for any distance \( d \) metrizing weak convergence, where \( \nu_{R_0}^R \) is the empirical process based on the probability measure \( R \) (see Giné and Zinn, 1991).

Now, if \( \| R_n - R_0 \|_\mathcal{G} \to 0 \) then any subsequence of \( R_n \) contains a further subsequence such that \( R_{n(h)} \to R_0 \) almost surely and thus, \( d(R_{n(h)}^{R_k}, Z_{R_0}) \to 0 \) almost surely; this implies that \( d(\nu_{R_0}^R, Z_{R_0}) \to_w 0 \) and we will say that \( \nu_{R_0}^R \to_w Z_{R_0} \) in probability.

Take \( \hat{D}_n^* = \nu_{R_0}^R \), with \( R_n = P(F_{\infty}, F_X) \), and \( R_0 = P(F_{\infty}, F_X) \). We have to show that \( \| R_n - R_0 \|_\mathcal{G} \to 0 \), where \( \mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4 \), and \( \mathcal{G}_1 = \mathcal{F} + \mathcal{H}, \mathcal{G}_2 = (\mathcal{F} + \mathcal{H})^2, \mathcal{G}_3 = (\mathcal{F} + \mathcal{H})^3, \mathcal{G}_4 = ((\mathcal{F} + \mathcal{H}))^2 \).

Note that

\[
\begin{align*}
\| R_n - R_0 \|_{\mathcal{G}_1} &\leq \| R_n - R_0 \|_{\mathcal{F}} + \| R_n - R_0 \|_{\mathcal{H}}, \\
\| R_n - R_0 \|_{\mathcal{G}_2} &\leq \| R_n - R_0 \|_{\mathcal{F}^2} + \| R_n - R_0 \|_{\mathcal{F} \mathcal{H}} + 2\| R_n - R_0 \|_{\mathcal{F} \mathcal{H}}, \\
\| R_n - R_0 \|_{\mathcal{G}_3} &\leq 2\| R_n - R_0 \|_{\mathcal{F}^2} + 2\| R_n - R_0 \|_{\mathcal{F} \mathcal{H}}, \\
\| R_n - R_0 \|_{\mathcal{G}_4} &\leq 3\| R_n - R_0 \|_{\mathcal{F}^3} + 3\| R_n - R_0 \|_{\mathcal{F} \mathcal{H}} + 4\| R_n - R_0 \|_{\mathcal{F} \mathcal{H}}.
\end{align*}
\]

We need to show that \( \| R_n - R_0 \|_C \to 0 \), where \( C \) is any of the classes \( \mathcal{F}, \mathcal{H}, \mathcal{F}^2, \mathcal{F} \mathcal{H}, \mathcal{F} \mathcal{H} \).
(a) \( \| R_n - R_0 \|_\mathcal{F} \longrightarrow p 0 \)

It holds that \( R_0(f_{st}) = 0 \) (see proof of Theorem 2.2 in Delicado and Romo (1994)).
Moreover,
\[
R_n(f_{st}) = \int \left[ \int f_n(y, x) d\mathcal{P}(F_{h_n}, \delta_x) \right] dF_{X,n} = \int \left[ P_{F_{h_n}}(A + xB \leq s) - P_{F_{h_n}}(A + xB \leq s) \right] I_{(\infty, s)}(x) dF_{X,n}.
\]
The regularity of the distribution of \( Y_x \) implies that the quantity between brackets is
\[
\sum_{j=1}^{N} (\hat{\theta}_{nj} - \theta_{0j}) \Delta_j(x, \theta_0, s) + R(x, \theta_0, \hat{\theta}_n, s),
\]
and so
\[
R_n(f_{st}) = \sum_{j=1}^{N} (\hat{\theta}_{nj} - \theta_{0j}) \left\{ \frac{1}{n} \sum_{i=1}^{n} \Delta_j(X_i, \theta_0, s) I_{(\infty, s)}(X_i) \right\} + o_p(s)(1) = \sum_{j=1}^{N} (\hat{\theta}_{nj} - \theta_{0j}) L_j(s, 1, \theta_0) + o_p(s)(1).
\]
The notation \( o_p(1) \) indicates that the term depends on \( s \); the hypothesis of regularity for \( Y_x \) ensures that the supremum in \( s \) of these quantities is \( o_p(1) \).
The hypothesis on \( L_j \) in Theorem 2.1 guarantees that the supremum in \( (s, t) \) of \( o_p(s)(1) \) is \( c_p(1) \). This, together with the uniform boundedness of \( L_j \) and the fact that \( \hat{\theta}_{nj} - \theta_{0j} = o_p(1), j = 1, \ldots, N, \) gives that
\[
\| R_n - R_0 \|_\mathcal{F} = \| R_n \|_\mathcal{F} \longrightarrow 0 \text{ in probability.}
\]

(b) \( \| R_n - R_0 \|_{\mathcal{H}} \longrightarrow p 0 \)

Let \( M = \max_{1 \leq j \leq N} \sup_{s, t} |L_j(s, t, \theta_0)| \), which is finite due to the hypothesis on \( E \Delta_j \).
Thus,
\[
\| R_n - R_0 \|_{\mathcal{H}} \leq N M \max_{1 \leq j \leq N} |(R_n - R_0)(h_j)|.
\]
Since \( \hat{\theta}_n \longrightarrow p \theta_0 \), we have that \( d(\hat{F}_{h_n}, F_{\theta}) \longrightarrow p 0 \). The continuity of functional \( \mathcal{P} \) (see Beran and Millar (1991), Proposition 2.1) implies that \( R_n \longrightarrow w R_0 \) in probability; hypothesis \( (i) \) gives that \( |(R_n - R_0)(h_j)| \longrightarrow p 0 \).

(c) \( \| R_n - R_0 \|_{\mathcal{H}} \longrightarrow p 0 \)

11
The argument in (b) applies to any finite-dimensional class of functions of the form

\[ C = \{ \sum_{j=1}^{N} \alpha_j h_j | \sum_{j=1}^{N} |\alpha_j| \leq K \} \]

when the functions \( h_j \) satisfy (i); in particular, this holds for the class \( \mathcal{H} \) and so, \( \| R_n - R_0 \|_{\mathcal{H}} \longrightarrow p \ 0 \).

(d) \( \| R_n - R_0 \|_{\mathcal{F}} \longrightarrow p \ 0 \)

From Theorem 2.2 in Delicado and Romo (1994), it follows that \( R_0(f_{st}f_{uv}) = F_X(r_{stuv}) \), where \( r_{stuv}(x) = (F(x, \theta_0, u \wedge s) - F(x, \theta_0, u))F(x, \theta_0, s)I_{(-\infty, \theta_0 \wedge 0]}(x) \), and that the class \( \mathcal{R} \) of these functions is Donsker for \( F_X \). Now, for \( R_n \) we have that

\[ R_n(f_{st}f_{uv}) = \int \left\{ \int f_{st}(y, x)f_{uv}(y, x)d\mathcal{P}(F_{dn}, \delta_x) \right\} dF_X \]

and the expression inside the first integral is

\[ E_{F_{dn}} \left[ I_{(-\infty, \theta_0 \wedge 0]}(A + xB) - P_{F_{dn}}(A + xB \leq u)I_{(-\infty, \theta_0 \wedge 0]}(A + xB) - P_{F_{dn}}(A + xB \leq s)I_{(-\infty, \theta_0 \wedge 0]}(A + xB) + P_{F_{dn}}(A + xB \leq u)P_{F_{dn}}(A + xB \leq s) \right] \]

\[ = \left[ F(x, \hat{\theta}_n, u \wedge s) - F(x, \hat{\theta}_n, u)F(x, \hat{\theta}_n, s) - F(x, \hat{\theta}_n, u)F(x, \theta_0, s) + F(x, \theta_0, s)F(x, \theta_0, u) \right] I_{(-\infty, \theta_0 \wedge 0]}(x) =

\[ = r_{stuv} + \sum_{j=1}^{N} (\hat{\theta}_{nj} - \theta_{0j}) \Delta_j(x, \theta_0, u \wedge s) + R(x, \theta_0, \hat{\theta}_n, u \wedge s) - \sum_{j=1}^{N} (\hat{\theta}_{nj} - \theta_{0j}) \Delta_j(x, \theta_0, s)F(x, \theta_0, u) + R(x, \theta_0, \hat{\theta}_n, s)F(x, \theta_0, u) \]

Thus,

\[ (R_n - R_0)(f_{st}f_{uv}) = (F_{X,n} - F_X)(r_{stuv}) + \]

\[ + \sum_{j=1}^{N} (\hat{\theta}_{nj} - \theta_{0j}) \left[ \frac{1}{n} \sum_{i=1}^{n} \Delta_j(X_i, \theta_0, u \wedge s)I_{(-\infty, \theta_0 \wedge 0]}(X_i) \right] - \frac{1}{n} \sum_{i=1}^{n} \Delta_j(X_i, \theta_0, u)F(X_i, \theta_0, s)I_{(-\infty, \theta_0 \wedge 0]}(X_i) - \frac{1}{n} \sum_{i=1}^{n} \Delta_j(X_i, \theta_0, s)F(X_i, \theta_0, u)I_{(-\infty, \theta_0 \wedge 0]}(X_i) + o_p(1). \]
So,

\[ \| R_n - R_0 \|_{\mathcal{F}} \leq \| F_{X,n} - F_X \|_{\mathcal{F}} + \sum_{j=1}^{N} |\hat{\theta}_n - \theta_j| \sup_{s,t,u,v} |L_j(u \wedge s, t \wedge v, \theta_0)| + \\
+ |L_j(u, t \wedge v, \theta_0)| + |L_j(s, t \wedge v, \theta_0)| + \phi^{(\text{stau})}(1) \] + \phi(1) = \\
\| F_{X,n} - F_X \|_{\mathcal{F}} + \phi(1),

(the supremum over \((s,t,u,v)\) of \(\phi^{(\text{stau})}(1)\) is \(\phi(1)\)), and this tends to zero in probability because \(\mathcal{R}\) is Donsker for \(F_X\).

(e) \( \| R_n - R_0 \|_{\mathcal{F}_N} \rightarrow_p 0 \)

From the definition of \(M\),

\[ \| R_n - R_0 \|_{\mathcal{F}_N} \leq N M \max_{1 \leq j \leq N} \| R_n - R_0 \|_{\mathcal{F}_j}, \]

where the class \(\mathcal{F}_j\) is the one obtained by multiplying the elements of \(\mathcal{F}\) by the function \(h_j\).

We will show that \( \| R_n - R_0 \|_{\mathcal{F}_j} \rightarrow_p 0 \), where \(h\) is any of the functions \(h_j, 1 \leq j \leq N\), for \(\theta\) equals to the true value \(\theta_0: h(y,x) = h_j(\theta_0, y, x)\). Let \(S_n = \mathcal{P}(F_{\theta_0}, F_X)\). Then

\[ \| R_n - R_0 \|_{\mathcal{F}_j} \leq \| R_n - S_n \|_{\mathcal{F}_j} + \| S_n - R_0 \|_{\mathcal{F}_j}, \]

and we will check that both of these elements tend to zero in probability.

We have that

\[ (R_n - S_n)(f) = \int \int f(x)h(a+xb,x)d(F_{\theta_0})(a,b). \]

The functions \(\psi_n(a,b) = \| F_{X,n} - F_X \|_{\mathcal{F}}(a,b)\) tend to zero in probability for all \((a,b)\) and \(|\psi_n(a,b)| \leq 2\|F\|_{\mathcal{F}}\|h\|\), where \(F\) is an envelope for \(\mathcal{F}\). Either hypothesis (ii) or (ii)' allow us to apply Theorem 1.5 to obtain

\[ \| R_n - S_n \|_{\mathcal{F}_j} \leq \int \psi_n(a,b)dF_{\theta_0} \rightarrow_p 0 \]

since \(\hat{\theta}_n \rightarrow_p \theta_0\) and so \(F_{\theta_0} \rightarrow_w F_{\theta_0}\), in probability.

Now, for \( \| S_n - R_0 \|_{\mathcal{F}_j} \), we have that

\[ (S_n - R_0)(f) = \int \int f(x)h(a+xb,x)d(F_{\theta_0})(a,b). \]

Let \(\lambda_{st}(a,b) = \int f(x)h(a+xb,x)d(F_X)(x)\). The class \(\mathcal{D} = \{\lambda_{st}\}\) is uniformly bounded. Let us show that is also equicontinuous and then apply Theorem 1.3. For a
function $\lambda_{st}$,

$$|\lambda_{st}(a,b) - \lambda_{st}(a',b')| \leq \int |f_{st}(a + xb, x)h(a + xb, x) - f_{st}(a' + xb', x)h(a' + xb', x)|dF_X(x) \leq$$

$$\leq \int_{C_1} |h(a + xb, x) - h(a' + xb', x)|dF_X(x) + \int_{C_2} |h(a + xb, x)|dF_X(x) + \int_{C_3} |h(a' + xb', x)|dF_X(x),$$

where $C_1(s, a, b, a', b') = \{ x \mid a + xb \leq s, a' + xb' \leq s \}$ and $C_2(s, a, b, a', b') = \{ x \mid a + xb \leq s, a' + xb' \leq s \}$ and $C_3(s, a, b, a', b') = C_2(s, a', b', a, b)$.

Given $\varepsilon > 0$, from (iii) we can find $\tau$ such that if $\|(a, b) - (a', b')\| < \tau$ then

$$|\lambda_{st}(a,b) - \lambda_{st}(a',b')| \leq \varepsilon + \|h\|_{\infty} P_{F_X}(C_2(s, a, b, a', b') \cup C_3(s, a, b, a', b'))$$

and hypothesis (iv) gives the equicontinuity of the family. Finally, Theorem 1.3 implies that

$$\lim_{n \to \infty} \sup_{s,t} \left| \int \lambda_{st}dF_{\Theta_n} - \int \lambda_{st}dF_{\Theta_0} \right| = 0 \text{ in probability.}$$

The next result gives the same conclusion under a different set of hypotheses.

**Theorem 3.2.** Suppose that conditions in Theorem 2.1 and Proposition 3.1 hold. If (i), either (iii) or (iii)', (e) and (vi) are true then

$$\hat{D}_n \rightarrow_{\text{w}} Z_{P(F_{\Theta_0}, F_X)} \text{ in probability.}$$

**Proof:** The proof is the same except for claim (e). So, it is enough to show part (e) under the present hypotheses. We use Theorem 1 in Billingsley and Topsoe (1967) (see Theorem 1.4) and we have to prove that $FH$ is a $P(F_{\Theta_0}, F_X)$-uniform class, where $h(\cdot, \cdot)$ is any of $h_j(\Theta_0, \cdot), j = 1, \ldots, N$. Hypothesis (i) in Theorem 1.4 holds because $FH$ is uniformly bounded; let us check (ii) in that theorem.

Let $(y_0, x_0) \in \mathbb{R}^{p+q}$. If $S((y_0, x_0), \delta)$ is the open ball with center $(y_0, x_0)$ and radius $\delta$, define $A^\delta = \bigcup_{\Theta \in A} S(\Theta, \delta)$. We have to establish that for all $\varepsilon > 0,

$$\lim_{\delta \to 0} \sup_{y \in x} P_{F(\Theta_0, F_X)} \left\{ \left( y_0, x_0 \right) : \sup_{(y', x') \in S((y_0, x_0), \delta)} |f_{st}(y, x)h(y, x) - f_{st}(y', x')h(y', x')| > \varepsilon \right\} = 0.$$}

For $(y, x), (y', x') \in S((y_0, x_0), \delta),

$$f_{st}(y, x)h(y, x) - f_{st}(y', x')h(y', x') =$$

$$= \left[ I_{(-\infty, \Theta_0]}(y) - F(x, \Theta_0, s) \right] I_{(-\infty, \Theta_0]}(x)h(y, x) - \left[ I_{(-\infty, \Theta_0]}(y') - F(x', \Theta_0, s) \right] I_{(-\infty, \Theta_0]}(x')h(y', x') = 14$$
\[
\begin{align*}
F(x', \theta_0, s) h(y', x') - F(x, \theta_0, s) h(y, x) \\
(1 - F(x, \theta_0, s)) h(y,x) - (1 - F(x', \theta_0, s)) h(y', x') \\
(1 - F(x, \theta_0, s)) h(y,x) + F(x', \theta_0, s) h(y', x') \\
-F(x, \theta_0, s) h(y,x) - (1 - F(x', \theta_0, s)) h(y', x') \\
-F(x, \theta_0, s) h(y,x) \\
F(x', \theta_0, s) h(y', x') \\
(1 - F(x, \theta_0, s)) h(y,x) \\
(1 - F(x', \theta_0, s)) h(y', x') \\
\end{align*}
\]

As (iii) implies (iii)', we will assume that (iii)' is true. So, the class \( \mathcal{A} \) is uniformly equicontinuous except a set \( \Gamma_h \) with null probability, where \( \mathcal{A} \) was defined in (v).

Let \( \Gamma(s,t, \delta) = \{(y,x): (s,t) \} \cap \{(y,x) \neq (s,t)\} \} \cup \Gamma_h \). Given \( (s,t) \), let \( (y_0,x_0) \notin \Gamma(s,t, \delta) \). For \( \delta \leq \delta_\varepsilon \),
\[
|f_{st}(y,x) h(y,x) - f_{st}(y',x') h(y',x')| \leq \varepsilon, \quad (y,x), (y',x') \in S((y_0,x_0), \delta).
\]
So, for all \( \varepsilon > 0 \), there exists \( \delta_\varepsilon \) such that for all \( (s,t) \), \( \delta \leq \delta_\varepsilon \) implies
\[
\sup_{(y,x),(y',x') \in S((y_0,x_0), \delta)} |f_{st}(y,x) h(y,x) - f_{st}(y',x') h(y',x')| > \varepsilon \leq P_{\mathcal{P}(F_\theta, F_X)}(\Gamma(s,t, \delta)).
\]
Now, let us see that (vi) gives that
\[
\lim_{\delta \to 0} \sup_{s,t} P_{\mathcal{P}(F_\theta, F_X)}(\Gamma(s,t, \delta)) = 0.
\]
Indeed, given \( \eta > 0 \), there exists a closed hypercube \( K_\eta \subset \mathbb{R}^{m+m} \) with probability larger than \( (1 - \eta) \) and such that
\[
P_{\mathcal{P}(F_\theta, F_X)}(\Gamma(s,t, \delta)) \leq P_{\mathcal{P}(F_\theta, F_X)}(\Gamma_\delta) + \eta + f_{\mathcal{P}(F_\theta, F_X)}(y_m,x_m) V(\delta),
\]
where \( (y_m, x_m) \) is the largest mode of the density of the distribution \( \mathcal{P}(F_\theta, F_X) \), and
\[
V(\delta) = \text{Vol}( \{ (y,x) \leq (s,t) \} \cap \{ (y,x) \neq (s,t) \} ) \leq (p+p) A_\delta
\]
where \( A_\delta \) is the surface of one of the faces of \( K_\eta \). So,
\[
\lim_{\delta \to 0} \sup_{s,t} P_{\mathcal{P}(F_\theta, F_X)}(\Gamma(s,t, \delta)) \leq P_{\mathcal{P}(F_\theta, F_X)}(\Gamma_\delta) + \eta = \eta,
\]
for any \( \eta > 0 \), and the result follows. \( \square \)

Condition (vi) follows, for instance, from the hypothesis in Theorem 1.2 (see Delicado and Romo (1994)). The next corollary provides the asymptotic validity of the statistics \( \hat{K}_n^* \) and \( \hat{M}_n^* \).

15
COROLLARY 3.1. For the statistics $$\hat{K}_n^* = \sup_{s,t} |\hat{J}_n^*(s, t)|$$ and $$\hat{M}_n^* = (\int_{\mathbb{R}^p+} (\hat{J}_n^*(s, t))^2 dQ(s, t))^\frac{1}{2}$$, it holds that

$$\hat{K}_n^* \rightarrow^\omega \|Z_{\mathcal{P}([F_{0},F_X])}\|_{\mathcal{F}+\mathcal{N}} \text{ a.s., and}$$

$$\hat{M}_n^* \rightarrow^\omega \|Z_{\mathcal{P}([F_{0},F_X])}\|_{\mathcal{F}+\mathcal{N},\mathcal{Q}} \text{ a.s.}$$

The strategy outlined in this paper can be also used to test one of the most relevant hypothesis in random coefficient regression models: the constancy of coefficients. This would entail to test that the distribution of $$B$$ is degenerated with parameters being the corresponding constant value, and assuming a parametric specification for the distribution of $$A$$.

4 A simulation study

We have conducted a Monte-Carlo experiment to study the size and power of these tests in practice. The data have been generated with the following algorithm. First, simulate independent $$(A_i, e_i), i = 1, \ldots, n$$ with $$A_i \sim F_A, e_i \sim F_e, A_i$$ and $$e_i$$ independent, and construct $$B_i = b_0 + \rho A_i + e_i, i = 1, \ldots, n$$. Then, take independent $$X_i, i = 1, \ldots, n$$ with distribution $$F_X$$ and, finally, calculate the observations $$Y_i = A_i + X_i B_i, i = 1, \ldots, n$$. $$b_0$$ is always equal to 1.

We label normal a model generated using variable $$A$$ with distribution $$N(0, 1)$$ and $$e$$ normally distributed such that $$E(e) = 0$$ and the standard deviation of $$B$$ is a specified value $$\sigma_B$$. The collection of simulations labelled Cauchy is constructed from $$A$$ with Cauchy distribution with zero median and interquantile semi range $$s_A$$ equal to one and $$B$$ is obtained from a Cauchy variable $$e$$ independent from $$A$$ such that the interquantile semi range of $$B$$ is a fixed value $$s_B$$. In our simulation study, we have considered each of this two situations with two sample sizes and two distributions for $$X (N(0, 1)$$ and $$N(2, 1))$$. 

[Table 1 about here]

Table 1 contains all the situations we have studied. They differ in the distribution of $$(A, B)$$, in the null and alternative hypotheses or in the parameters to be estimated.

In situations 1 to 4, the coefficients $$A$$ and $$B$$ are independent under $$H_0 (H_0: \rho = 0)$$; in 5 and 6, $$B$$ is degenerated under $$H_0$$, i.e., we are testing constancy of $$B$$. Third column in Table 1 specifies the unknown parameter in each situation; when they act as known values, we have taken $$\rho = \mu_A = m_A = 0, \mu_B = m_B = \sigma_A = \sigma_B = s_A = s_B = 1$$. In cases 1 to 4, $$\rho$$ has taken values in the set $$S^\rho_A = \{-0.9, -0.6, -0.4, -0.2, 0.2, 0.4, 0.6, 0.9\}$$ under $$H_A$$ and in cases 5 and 6, the scale parameter of $$B$$ belongs to the set $$S^\rho_B = \{0.2, 0.4, 0.6, 0.8, 1, 1.3, 1.6, 2\}$$.
We have used sample sizes \( n = 50 \) and \( n = 100 \) and we have simulated 500 samples for each combination of distributions and parameters and the number of bootstrap replications in each case was \( B = 500 \).

We present with some detail the estimation procedure corresponding to case 4. The model is

\[
Y_i = A_i + X_i B_i = m_A + X_i m_B + \varepsilon_i,
\]

where \( A, B \) have independent Cauchy distribution and \( \varepsilon_i = (A_i - m_A) + X_i (B_i - m_B), i = 1, \ldots, n \), are also Cauchy. The median of \( \varepsilon_i \) is 0 and its scale parameter \( s_i = s_A + |X_i| s_B \) is unknown. We propose here an estimator of the parameters of \( A \) and \( B \) in the spirit of Hildreth and Houck (1968), but based on minimum absolute deviations regression.

First, we obtain initial minimum absolute deviations estimations of \( m_A \) and \( m_B \) from the pairs \((Y_i, X_i)\) and calculate the estimated residuals \( \hat{\varepsilon}_i = Y_i - \hat{m}_A - X_i \hat{m}_B \). Note that

\[
|\hat{\varepsilon}_i| = \text{med}(|\varepsilon_i|) + (|\hat{\varepsilon}_i| - \text{med}(|\varepsilon_i|)) \approx \text{med}(|\varepsilon_i|) + \hat{\varepsilon}_i = s_i + v_i = s_A + |X_i| s_B + v_i,
\]

where \( v_i \) is a random variable with zero median.

Second, estimate the scale parameters of \( A \) and \( B \) by using minimum absolute deviations from the regression

\[
|\hat{\varepsilon}_i| = s_A + |X_i| s_B + v_i, \quad i = 1, \ldots, n.
\]

Let \( \hat{s}_A \) and \( \hat{s}_B \) be these estimates. Now, estimate the scale of \( \varepsilon_i \) by using \( \hat{s}_i = \hat{s}_A + |X_i| \hat{s}_B \).

Third, calculate the generalized minimum absolute deviations estimates of \( m_A \) and \( m_B \) as the minimum absolute deviations of the regression

\[
\frac{Y_i}{\hat{s}_i} = \frac{1}{\hat{s}_i} m_A + \frac{X_i}{\hat{s}_i} m_B + \frac{\varepsilon_i}{\hat{s}_i}.
\]

Table 2 presents the empirical sizes of the described simulation experiment. The sizes obtained in situations 1, 3 and 4 are very satisfactory: only 2 out of 72 are significantly different from the theoretical sizes at 95%. In situation 2, the estimated sizes are more often significantly different from the theoretical ones, specially for \( \alpha = 0.05 \) and when \( X \) is centered at 2. Testing for constant coefficients (5 and 6) gives empirical sizes which are very different from the expected ones: when \( A \) has a normal distribution, they are much lower than the nominal sizes and the opposite happens for a Cauchy coefficient \( A \).
Finally, we described the results on the power of the tests against models with different values of one parameter. Figures 1 and 2 contain two types of graphs: some of them represent the empirical power functions and the remainder ones are multiple box-plots; each of these represents estimates of the unknown parameters corresponding to a value of the parameter under either the null or the alternative. We give some results for $n = 100$ (the results for $n = 50$ are qualitatively similar but with lower power and larger dispersion for the parameter estimations).

Figure 1 corresponds to situation 3 in Table 1. With $X \sim N(0,1)$, we see the same behavior for $K_n$ for both positive and negative values for $\rho$ and lack of identifiability for $M_n$ with negative values of $\rho$. Also, lack of identifiability appears for $X \sim N(2,1)$ and $\rho$ positive (see Figure 1.(b)). For all values of $\rho$, we get unbiased estimations of $m_B$ with dispersion increasing with the absolute values of $\rho$, except when $\rho$ is positive and $X \sim N(2,1)$ because this case cannot be distinguished from $\rho = 0$.

The graphs in Figure 2 give the results for situation 4 in Table 1. Now, we are estimating four parameters and so the empirical powers are lower than the ones obtained in Figure 1, when we were estimating just one parameter; this is especially clear for negative $\rho$ and $X \sim N(2,1)$. When $X$ is centered at zero, the power functions for both $K_n$ and $M_n$ are similar; this didn’t happen for tests different from this one only in the number of estimated parameters. The multiple box-plots show that all the estimations, especially for scale parameters are very sensitive to the value of $\rho$ used to generate the data; let us recall that the procedure is designed to act correctly under the null hypothesis ($\rho = 0$).

Finally, Figure 3 contains the power functions for the different specified situations. Part (a) corresponds to normal $(A, B)$ with $X \sim N(2,1)$ and this is better than the result we obtained when $X \sim N(0,1)$. Moreover, comparing with other results available from our simulations, we got slightly higher powers for estimated parameters in this case. In part (b), we estimate the parameters $\mu_A, \mu_B, \sigma_A$ and $\sigma_B$ and $X \sim N(0,1)$; the power is larger when we estimate four parameter than when none or just one parameter is estimated. The results in part (c) are slightly better than the ones we obtained for $X \sim N(0,1)$ and the same phenomenon happens with the Cauchy situation; in part (d) with respect to the case with $X \sim N(2,1)$. To end, we may also report that the power is low when the four parameters are estimated in the Cauchy case.
References


<table>
<thead>
<tr>
<th>Situation</th>
<th>((A, B)) model</th>
<th>Unknown parameters</th>
<th>Specified parameters under (H_0)</th>
<th>Specified parameters under (H_A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normal</td>
<td>(\mu_B)</td>
<td>(\rho = 0)</td>
<td>(\rho \in S_A^1)</td>
</tr>
<tr>
<td>2</td>
<td>Normal</td>
<td>(\mu_A, \mu_B, \sigma_A, \sigma_B)</td>
<td>(\rho = 0)</td>
<td>(\rho \in S_A^1)</td>
</tr>
<tr>
<td>3</td>
<td>Cauchy</td>
<td>(m_B)</td>
<td>(\rho = 0)</td>
<td>(\rho \in S_A^1)</td>
</tr>
<tr>
<td>4</td>
<td>Cauchy</td>
<td>(m_A, m_B, s_A, s_B)</td>
<td>(\rho = 0)</td>
<td>(\rho \in S_A^1)</td>
</tr>
<tr>
<td>5</td>
<td>Normal</td>
<td>(\mu_A, \mu_B, \sigma_A)</td>
<td>(\sigma_B = 0)</td>
<td>(\sigma_B \in S_A^2)</td>
</tr>
<tr>
<td>6</td>
<td>Cauchy</td>
<td>(m_A, m_B, s_A)</td>
<td>(s_B = 0)</td>
<td>(s_B \in S_A^2)</td>
</tr>
</tbody>
</table>

Table 1: Simulation design for estimated parameters.
<table>
<thead>
<tr>
<th>Situation</th>
<th>$X \sim N(0,1)$</th>
<th>$X \sim N(2,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 50$</td>
<td>$n = 100$</td>
</tr>
<tr>
<td></td>
<td>$M_n$</td>
<td>$K_n$</td>
</tr>
<tr>
<td>1</td>
<td>$.002$</td>
<td>$.006$</td>
</tr>
<tr>
<td></td>
<td>$.046$</td>
<td>$.048$</td>
</tr>
<tr>
<td></td>
<td>$.104$</td>
<td>$.118$</td>
</tr>
<tr>
<td></td>
<td>$.036$</td>
<td>$.022$</td>
</tr>
<tr>
<td>2</td>
<td>$.068$</td>
<td>$.060$</td>
</tr>
<tr>
<td></td>
<td>$.114$</td>
<td>$.114$</td>
</tr>
<tr>
<td></td>
<td>$.010$</td>
<td>$.014$</td>
</tr>
<tr>
<td>3</td>
<td>$.062$</td>
<td>$.044$</td>
</tr>
<tr>
<td></td>
<td>$.114$</td>
<td>$.102$</td>
</tr>
<tr>
<td></td>
<td>$.006$</td>
<td>$.016$</td>
</tr>
<tr>
<td>4</td>
<td>$.052$</td>
<td>$.066$</td>
</tr>
<tr>
<td></td>
<td>$.092$</td>
<td>$.096$</td>
</tr>
<tr>
<td>5</td>
<td>$.000$</td>
<td>$.000$</td>
</tr>
<tr>
<td></td>
<td>$.000$</td>
<td>$.004$</td>
</tr>
<tr>
<td></td>
<td>$.004$</td>
<td>$.010$</td>
</tr>
<tr>
<td>6</td>
<td>$.042$</td>
<td>$.020$</td>
</tr>
<tr>
<td></td>
<td>$.072$</td>
<td>$.056$</td>
</tr>
<tr>
<td></td>
<td>$.120$</td>
<td>$.110$</td>
</tr>
</tbody>
</table>

Table 2: Empirical sizes. The three values appearing in each cell are the empirical sizes corresponding to the theoretical ones of .01, .05 and .1, from top to the bottom.
Figure 1: Estimation 3. Power function for $H_0: \rho = 0$ against different values of $\rho$, with estimated $m_B$. $(A, B)$ are Cauchy and $\alpha = 0.05$. $X \sim N(0,1)$ in the left hand side and $X \sim N(2,1)$ in the right one.
Figure 2: Estimation 4. Power function for $H_0 : \rho = 0$ against different values of $\rho$, with estimated $m_A, m_B, s_A, s_B$. $(A, B)$ are Cauchy, $\alpha = 0.05$ and $X \sim N(0, 1)$.
Figure 3: Power functions for:

(a) $H_0: \rho = 0$ against different values of $\rho$, with estimated $\mu_B$. $(A, B)$ are normal, $\alpha = 0.05$ and $X \sim N(2, 1)$.

(b) same, with estimated $\mu_A, \mu_B, \sigma_A, \sigma_B$ and $X \sim N(0, 1)$.

(c) $H_0: \sigma_B = 0$ against different values of $\sigma_B$, with estimated $\mu_A, \mu_B, \sigma_A$. $(A, B)$ are normal, $\alpha = 0.05$ and $X \sim N(2, 1)$.

(d) $H_0: s_B = \theta$ against different values of $s_B$, with estimated $m_A, m_B, s_A$. $(A, B)$ are Cauchy, $\alpha = 0.05$ and $X \sim N(2, 1)$. 