Working Paper 94-52
Statistics and Econometrics Series 20
December 1994

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe, Madrid (Spain)
Fax (341) 624-9849

GOODNESS OF FIT TESTS IN RANDOM COEFFICIENT REGRESSION MODELS[1]

Pedro F. Delicado and Juan Romo

Abstract

Random coefficient regressions have been applied in a wide range of fields, from biology to economics, and constitute a common frame for several important statistical models. A nonparametric approach to inference in random coefficient models was initiated by Beran and Hall. In this paper we introduce and study goodness of fit tests for the coefficient distributions; their asymptotic behaviour under the null hypothesis is obtained. We also propose bootstrap resampling strategies to approach these distributions and prove their asymptotic validity using results by Giné and Zinn on bootstrap empirical processes. A simulation study illustrates the properties of these tests.

Key words:
Goodness of fit, linear regression, random coefficient, empirical processes, Vapnik-Červonenkis classes.

[1]Research partially supported by DGICYT PB93-0232 (Spain).

1 Introduction.

Random coefficient regression models have raised a growing interest in recent years. From a statistical point of view, they constitute a unifying frame for several important models, including random effects in ANOVA (see, e.g. Scheffé (1959)), deconvolution models (Fan (1991), van Es (1991)), heteroscedastic linear models or location-scale mixture models. The scope of their application ranges from biology to image compression to econometrics. Raj and Ullah (1981), Chow (1983), Nicholls and Quinn (1982) and Nicholls and Pagan (1985) are surveys of this work. A common feature of this literature is that interest is focused on moments estimation, essentially mean and variance.

A nonparametric approach to inference in these models has been started by Beran and Hall (1992) by addressing the estimation of the random parameter joint distribution; they solved it by introducing consistent estimators based on estimated moments. Beran (1991) developed nonparametric prediction intervals for the dependent variable and introduced a minimum distance estimate. Beran and Millar (1991) have studied its consistency and proved that it is a $n^{1/2}$-consistent estimator of the coefficient distribution in a particular case.

In this paper, we consider nonparametric goodness of fit tests for the distribution of the random coefficients. In section 1.1 we establish the model and we propose the corresponding test statistics. The one-dimensional response case is studied in section 2.1. We use results from empirical processes theory to assess the asymptotic behaviour of the statistics. Then we present different bootstrap resampling strategies to approach the unknown limiting distribution and we prove their validity. Section 2.2 extends these results to linear models with p-dimensional dependent variable. A simulation study on the performance of these tests is carried out in section 3. Finally, all the proofs are collected in an appendix.

1.1 Preliminaries.

Let us write the random coefficient regression model as

$$Y_i = A_i + X_i B_i, \quad i \geq 1, \tag{1.1}$$

where Y_i and A_i are p-dimensional random variables, B_i is a q-dimensional random vector 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 (A_i, B_i) is independent of X_i. The distribution of (A_i, B_i, X_i) is unknown and we can observe the n pairs (Y_i, X_i), $1 \leq i \leq n$. Let F_{AB} be the distribution of (A_i, B_i) and let F_X be the distribution of X_i, both unknown. The joint distribution of (Y_i, X_i) depends on both distributions and will be denoted by $F_{YX} = \mathcal{P}(F_{AB}, F_X)$. 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. This is the model considered by Beran and Millar (1991). Our goal is a goodness of fit test for the distribution F_{AB}, i.e.,

$$H_0 \; : \; F_{AB} = G \tag{1.2}$$

$$H_1 \; : \; F_{AB} \neq G \tag{1.3}$$

for a specified distribution G.

We will assume *identifiability* in the model (1.1), i.e., $\mathcal{P}(F_{AB}, F_X) = \mathcal{P}(\tilde{F}_{AB}, F_X)$ implies $F_{AB} = \tilde{F}_{AB}$. Beran (1991) and Beran and Millar (1991) give sufficient conditions for identifiability and also for ***strong identifiability***, a locally uniform version of identifiability. If identifiability does not hold, one can consider the equivalence classes $C(F_{AB}) = \{\tilde{F}_{AB} \mid \mathcal{P}(F_{AB}, F_X) = \mathcal{P}(\tilde{F}_{AB}, F_X)\}$ and carry out the test

$$H_0 \; : \; F_{AB} \in C(G) \tag{1.4}$$

$$H_1 \; : \; F_{AB} \notin C(G) \tag{1.5}$$

instead.

We will base our test on the empirical process $D_n = \sqrt{n}\,(P_n - \mathcal{P}(F_{AB}, F_X))$. Since F_X is not known and is not specified under H_0, we consider the "estimated" empirical process

$$J_n = \sqrt{n}\,(P_n - \mathcal{P}(F_{AB}, F_{X,n}))$$

indexed by the class $\mathcal{J} = \{I_{st} = (-\infty, s] \times (-\infty, t] : s \in \mathbf{R}^p, t \in \mathbf{R}^{pq}\}$ of $(p+pq)$-dimensional semiintervals. When convenient, we will also express $J_n(I_{st})$ as $J_n(s,t)$, $s \in \mathbf{R}^p$, $t \in \mathbf{R}^{pq}$.

Given $s \in \mathbf{R}^p, t \in \mathbf{R}^{pq}$,

$$\begin{aligned} J_n(I_{st}) &= \sqrt{n}\left(\int_{\mathbf{R}^{p+pq}} I_{(-\infty,(s,t)]}(y,x)\,dP_n(y,x) - \right. \\ &\quad \left. - \int_{\mathbf{R}^{pq}} P_{F_{AB}}(A + xB \leq s) I_{(-\infty,t]}(x)\,dF_{X,n}(x)\right) = \\ &= \sqrt{n}\int_{\mathbf{R}^{p+pq}} \left(I_{(-\infty,s]}(y) - P_{F_{AB}}(A + xB \leq s)\right) I_{(-\infty,t]}(x)\,dP_n(y,x) = \\ &= \sqrt{n}\int_{\mathbf{R}^{p+pq}} f_{st}(y,x)\,dP_n(y,x) = \sqrt{n}P_n(f_{st}), \end{aligned}$$

where $f_{st}(y,x) = \left(I_{(-\infty,s]}(y) - P_{F_{AB}}(A + xB \leq s)\right) I_{(-\infty,t]}(x)$. Observe that $\mathcal{P}(F_{AB}, F_X)(f_{st})$ is equal to zero. Thus, it turns out that, for each s, t, $J_n(s,t) = D_n(f_{st})$.

Let

$$\mathcal{F} = \{f_{st} \mid s \in \mathbf{R}^p, t \in \mathbf{R}^{pq}\}. \tag{1.6}$$

Our test statistics will be Kolmogorov-Smirnov or Cramér-von Mises functionals of J_n; we will get their asymptotic distribution from the convergence of $\{D_n : n \in \mathbf{N}\}$ as empirical processes indexed by $\mathcal{F}$. We refer to Giné and Zinn (1986) for the definitions of weak convergence in the space $l^\infty(\mathcal{F})$ of bounded functions on $\mathcal{F}$, and the concepts of Vapnik-Červonenkis and Donsker classes of functions.

For any class $\mathcal{F}$ of functions $f : \mathcal{X} \to \mathbf{R}$, if $A \subset \mathcal{X}$ is a finite set and $\varepsilon > 0$, let

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

and define $D(\varepsilon, F, \mathcal{F}) = \sup\{D(\varepsilon, A, F, \mathcal{F}) | A \subseteq \mathcal{X}, A \text{ finite}\}$. The corresponding entropy is $H(\varepsilon, F, \mathcal{F}) = \log D(\varepsilon, F, \mathcal{F})$. As usual, if V is a random variable, we will represent by Supp(V) (the support of V) the smallest closed set S such that $P\{V \notin S\} = 0$.

2 Main results.

2.1 Univariate dependent variable.

In this section we establish the asymptotic distribution for the Kolmogorov-Smirnov and the Cramér-von Mises statistics based on the empirical process D_n when the dependent variable Y is one-dimensional, i.e., $p = 1$. First, we give sufficient conditions for the class $\mathcal{F}$ defined in (1.6) to be a permissible class of functions in the sense of Pollard (1984).

PROPOSITION 2.1. *Consider the conditions:*

(a) $P_{F_{AB}}(A + xB = s) = 0$ *for all* $x \in$ *Supp*(X) *and for all* $s \in \mathbf{R}$.

(b) The distribution of (A, B) *is discrete.*

If either (a) or (b) hold then $\mathcal{F}$ *is a permissible class of functions.*

Note that the proposition holds if, e.g., for any fixed value of X, the distribution of Y has no atoms and this follows if the distribution of A is absolutely continuous, except for extreme dependence between A and B. This measurability requirement implies the one used by Giné and Zinn in theorems that we will need below (see Giné and Zinn (1990), p. 854). The next result gives a bound for the entropy of $\mathcal{F}$.

PROPOSITION 2.2. *There exist positive constants* A *and* w *such that* $D(\varepsilon, F, \mathcal{F}) \le A\varepsilon^{-2w}$, *where* $F = 1$ *is an envelope for* $\mathcal{F}$.

The asymptotic behaviour of the sequence $\{D_n : n \in \mathbf{N}\}$ of empirical processes is obtained in the following theorem.

THEOREM 2.1. *Under the conditions in Proposition 2.1,*

$$D_n \longrightarrow_w Z_{\mathcal{P}(F_{AB},F_X)} \tag{2.7}$$

in $l^\infty(\mathcal{F})$, *where* $Z_{\mathcal{P}(F_{AB},F_X)}$ *is the* $\mathcal{P}(F_{AB},F_X)$*-brownian bridge with covariance structure given by*

$$\begin{aligned} Cov(f_{st}, f_{uv}) &= \int_{\{x\le t\wedge v\}} (P_{F_{AB}}(A+xB\le s\wedge u)- \\ &\quad -P_{F_{AB}}(A+xB\le s)P_{F_{AB}}(A+xB\le u))\, dF_X(x). \end{aligned}$$

From this result, we can obtain the asymptotic distribution of the corresponding Kolmogorov-Smirnov statistic,

$$K_n = \|D_n\|_{\mathcal{F}} = \sup_{f_{st}\in\mathcal{F}} \mid D_n(f_{st}) \mid = \|J_n\|_{\mathcal{J}} = \sup_{I_{st}\in\mathcal{J}} \mid J_n(I_{st}) \mid$$

and the Cramér-von Mises one,

$$M_n = \left(\int_{\mathbf{R}^{1+q}} (D_n(f_{st}))^2 dQ(s,t)\right)^{\frac{1}{2}} = \left(\int_{\mathbf{R}^{1+q}} (J_n(s,t))^2 dQ(s,t)\right)^{\frac{1}{2}},$$

where Q is a finite measure on $\mathbf{R}^{1+q}$.

COROLLARY 2.1. *Under the conditions in Proposition 2.1,*

$$\begin{aligned} K_n &\longrightarrow_w \|Z_{\mathcal{P}(F_{AB},F_X)}\|_{\mathcal{F}} \text{ and} \\ M_n &\longrightarrow_w \left(\int_{\mathbf{R}^{1+q}} (Z_{\mathcal{P}(F_{AB},F_X)}(f_{st}))^2 dQ(s,t)\right)^{\frac{1}{2}}. \end{aligned}$$

Finally, the limit distributions of D_n, K_n and M_n under $H_0 : F_{AB} = G$ can be obtained by replacing F_{AB} by G in the corresponding expressions in Theorem 2.1 and Corollary 2.1. These asymptotic distributions are depending both on G and the unknown F_X, so, it is important to provide a way of approaching them and this can be done through bootstrapping. We will study two different resampling strategies.

Let $(Y_i^{*1}, X_i^{*1}), i = 1, \ldots, n$, be a random sample from the distribution P_n and let P_n^{*1} be the corresponding empirical distribution. Define $D_n^{*1} = \sqrt{n}\,(P_n^{*1} - P_n)$, $n \in \mathbf{N}$, indexed by $\mathcal{F}$, as the standard bootstrap version of D_n. Consider now a random sample (Y_i^{*2}, X_i^{*2}), $i = 1, \ldots, n$, from the distribution $\mathcal{P}(F_{AB}, F_{X,n})$ and let P_n^{*2} the empirical distribution obtained from them. Define the corresponding bootstrap empirical process $D_n^{*2} = \sqrt{n}\,(P_n^{*2} - \mathcal{P}(F_{AB}, F_{X,n}))$, indexed by $\mathcal{F}$. Finally, let us introduce J_n^{*2}, the bootstrap analogous of J_n, indexed by the lower semiintervals in $\mathbf{R}^{1+q}$,

$$J_n^{*2}(s,t) = \sqrt{n}\left(P_n^{*2}(s,t) - \mathcal{P}(F_{AB}, F_{X,n}^{*2})(s,t)\right),$$

where $F^{*2}_{X,n}$ is the second marginal distribution of P^{*2}_n, i.e., the empirical distribution corresponding to X^{*2}_i, $i = 1, \ldots, n$. It's straightforward to check now that $J^{*2}_n(s,t) = D^{*2}_n(f_{st})$ and obviously $J^{*2}_n(s,t) \neq D^{*1}_n(f_{st})$. Our next result gives the asymptotic validity of these bootstrap statistics.

THEOREM 2.2. *Assume the conditions in Proposition 2.1. Then the bootstrap empirical process D^{*1}_n converges weakly in $l^\infty(\mathcal{F})$ to the limit process of D_n for almost all sample sequence $\{(Y_i, X_i) : i \in \mathbb{N}\}$ satisfying model (1.1):*

$$D^{*1}_n \longrightarrow_w Z_{\mathcal{P}(F_{AB}, F_X)} \quad a.s., \quad n \longrightarrow \infty.$$

Also D^{*2}_n converges weakly in $l^\infty(\mathcal{F})$ to the same limit for almost all sample sequence $\{(Y_i, X_i) : i \in \mathbb{N}\}$ satisfying model (1.1):

$$D^{*2}_n \longrightarrow_w Z_{\mathcal{P}(F_{AB}, F_X)} \quad a.s., \quad n \longrightarrow \infty.$$

The proof of the second part relies on Corollary 2.7 in Giné and Zinn (1991) and it can be seen in the Appendix. Now, the continuous mapping theorem provides the asymptotic behaviour of the bootstrap versions of K_n and M_n.

COROLLARY 2.2. *Under the conditions in Proposition 2.1,*

$$K^{*1}_n = \|D^{*1}_n\|_{\mathcal{F}} \longrightarrow_w \|Z_{\mathcal{P}(F_{AB},F_X)}\|_{\mathcal{F}} \quad a.s.,$$

$$K^{*2}_n = \|D^{*2}_n\|_{\mathcal{F}} \longrightarrow_w \|Z_{\mathcal{P}(F_{AB},F_X)}\|_{\mathcal{F}} \quad a.s.,$$

$$M^{*1}_n = \left(\int_{\mathbb{R}^{1+q}} (D^{*1}_n(f_{st}))^2 dQ(s,t)\right)^{\frac{1}{2}} \longrightarrow_w \left(\int_{\mathbb{R}^{1+q}} (Z_{\mathcal{P}(F_{AB},F_X)}(f_{st}))^2 dQ(s,t)\right)^{\frac{1}{2}} \quad a.s.,$$

$$\text{and } M^{*2}_n = \left(\int_{\mathbb{R}^{1+q}} (D^{*2}_n(f_{st}))^2 dQ(s,t)\right)^{\frac{1}{2}} =$$

$$= \left(\int_{\mathbb{R}^{1+q}} (J^{*2}_n(s,t))^2 dQ(s,t)\right)^{\frac{1}{2}} \longrightarrow_w \left(\int_{\mathbb{R}^{1+q}} (Z_{\mathcal{P}(F_{AB},F_X)}(f_{st}))^2 dQ(s,t)\right)^{\frac{1}{2}} \quad a.s.$$

as $n \longrightarrow \infty$.

Note that here we can obtain the convergence of D^{*1}_n and D^{*2}_n because now the data come from the distribution used to center D^{*2}_n and this allows us to use the Giné and Zinn (1991) corollary, which cannot be used to obtain the convergence of J_n.

These results allow us to propose two bootstrap algorithms for testing H_0 based on D^{*1}_n and D^{*2}_n, respectively. By replacing F_{AB} by G, the resampling scheme in D^{*2}_n provides data (Y^{*2}_i, X^{*2}_i) satisfying the null hypothesis; on the contrary, this does not necessarily happen with D^{*1}_n: the sample $(Y_i, X_i), i = 1, \ldots, n$, may not come from model (1.1) with $F_{AB} = G$.

However, we can take independent observations $(A_i^G, B_i^G), i = 1, \ldots, n$, from the distribution G and construct the pairs

$$(Y_i^G, X_i), \text{ with } Y_i^G = A_i^G + X_i B_i^G, \quad i = 1, \ldots, n \tag{2.8}$$

which come from model (1.1) with $F_{AB} = G$.

So, our resampling algorithms are the following. The one based on D_n^{*1} proceeds in four steps (suppose that we are using the statistic K_n):

1. Obtain the value of the statistic K_n from the sample (Y_i, X_i), with $F_{AB} = G$.
2. Construct the sample (Y_i^G, X_i) as indicated in (2.8).
3. Draw B bootstrap values $K_{n,b}^{*1}$, $b = 1, \ldots, B$ of the statistic K_n^{*1}, defined in Corollary 2.2, from the values (Y_i^G, X_i) in step 2.
4. Reject H_0 if K_n is larger than the α-th quantile of the empirical distribution of $K_{n,b}^{*1}$, $b = 1, \ldots, B$.

The second method, based on D_n^{*2}, modifies steps 2 and 3 to the following:

(2. and 3.)' For $b = 1, \ldots, B$, obtain (Y_i^{*2}, X_i^{*2}), $i = 1, \ldots, n$ i.i.d. from $\mathcal{P}(G, F_{X,n})$, construct J_n^{*2} and calculate $K_n^{*2} = \sup_{s,t} |J_n^{*2}(s,t)|$.

2.2 p-variate dependent variable.

We consider now the model (1.1) when the dependent variable Y is p-dimensional, with $p > 1$, and our goal is to prove similar results to the previous ones. Checking the proofs in 2.1, the fact of Y_i being univariate is only used to show that the class

$$\mathcal{F}_1 = \{f_s : \mathbf{R}^q \longrightarrow \mathbf{R} \mid f_s(x) = P_{F_{AB}}(A + xB \leq s), s \in \mathbf{R}\}$$

is a Vapnik-Červonenkis class of functions and so, it has a small entropy. We have to deal now with the class

$$\mathcal{F}_p = \{f_s : \mathbf{R}^{p \times q} \longrightarrow \mathbf{R} \mid f_s(x) = P_{F_{AB}}(A + xB \leq s), s \in \mathbf{R}^p\}, \ p > 1,$$

whose envelope is the function constantly equal to 1.

Our next theorem describes the properties of $\mathcal{F}_p$ under some conditions on the distribution of (A, B, X). We will use the function h defined as

$$\begin{aligned} h \ : \ \mathbf{R}^{p \times q} \times \mathbf{R}^p \ &\longrightarrow \ \mathbf{R} \\ (x, s) \ &\longmapsto \ f_s(x). \end{aligned}$$

THEOREM 2.3. *Assume that Supp*(X) *is compact in model (1.1) and that* $Y_x = A + xB$ *is absolutely continuous for all* $x \in$ *Supp*(X). *Suppose also that* h *has uniformly bounded partial derivatives:*

$$\left\| \frac{\partial h}{\partial x}(x,s) \right\| \leq M_1 \quad , \quad x \in \mathbf{R}^{pq},\ s \in \mathbf{R}^p,$$
$$\left\| \frac{\partial h}{\partial s}(x,s) \right\| \leq M_2 \quad , \quad x \in \mathbf{R}^{pq},\ s \in \mathbf{R}^p.$$

Then

(i) The family of probability measures $\{P_{Y_x},\ x \in \mathit{Supp}(X)\}$ *is tight: for all* $\varepsilon > 0$ *there exists a compact* $C(\varepsilon)$ *such that* $P_{Y_x}(C(\varepsilon)) \geq 1 - \varepsilon$, $x \in \mathit{Supp}(X)$. $C(\varepsilon)$ *can be chosen to be of the form*

$$C(\varepsilon) = [l(\varepsilon), u(\varepsilon)]_p \equiv \{s \in \mathbf{R}^p \mid l(\varepsilon) \leq s \leq u(\varepsilon)\}.$$

(ii) $D(\varepsilon, F, \mathcal{F}_p) \leq p^p \left(\frac{\varepsilon}{2M_2}\right)^{-p} \mathit{Vol}([l(\varepsilon), u(\varepsilon)]_p)$.

(iii) If

$$\int_0^1 \log\left(\mathit{Vol}([l(\varepsilon/2), u(\varepsilon/2) + 1_p \frac{\varepsilon}{M_2}]_p)\right) d\varepsilon < \infty \tag{2.9}$$

then

$$\int_0^1 H(\varepsilon, F, \mathcal{F}_p) d\varepsilon < \infty \tag{2.10}$$

and also

$$\int_0^1 H(\varepsilon, F, \mathcal{F}) d\varepsilon < \infty, \tag{2.11}$$

where $\mathcal{F}$ *is the class of functions defined in (1.6).*

Next we provide two important situations where Theorem 2.3 applies: first, in those cases where (A, B) is compactly supported and second, when (A, B) has a $(p+q)$-variate normal distribution.

PROPOSITION 2.3. *Assume that Supp*(X) *is compact. It holds that*

(i) if Supp(A, B) *is compact then*

$$Q = \bigcup_{x \in \mathit{Supp}(X)} \{y \in \mathbf{R}^p | y = a + xb,\ (a,b) \in \mathit{Supp}(A,B)\}$$

is compact and $P(Y_x \in Q) = 1$, *for all* $x \in \mathit{Supp}(X)$.

(ii) if, moreover, the hypotheses of Theorem 2.3 on Y_x *and on the partial derivatives of* $h(x,s)$ *hold, then*

$$\int_0^1 H(\varepsilon, F, \mathcal{F}) d\varepsilon < \infty.$$

The corresponding result for normally distributed coefficients is the following:

PROPOSITION 2.4. *Assume that the variable (A, B) in model (1.1) is normally distributed, that Supp(X) is compact and that, for all $x \in$ Supp(X), the variable Y_x is absolutely continuous. Then*

(i) The function $h(x, s)$ has uniformly bounded partial derivatives.

(ii) Condition (2.9) holds.

As a consequence of (i) and (ii), the conclusions (2.10) and (2.11) in Theorem 2.3 follow.

Observe that whenever Theorem 2.3 holds, all the conclusions we obtained for Y univariate can be carried out to the p-dimensional case. In particular, the goodness of fit test strategy applies straightforward to model (1.1) for any $p \geq 1$.

3 A simulation study.

To study the size and the power of these tests in practice, we have conducted a Monte-Carlo experiment. The data have been generated in the following way. 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 then construct $B_i = b_0 + \rho A_i + e_i, i = 1, \ldots, n$. Second, take independent $X_i, i = 1, \ldots, n$ with distribution F_X and, finally, calculate the observations $Y_i = A_i + X_i B_i, i = 1, \ldots, n$.

The first set of simulations (labelled *normal*) corresponds to a model generated using 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 σ_B. The second collection of simulations (labelled *Cauchy*) is built from A with a Cauchy distribution with zero median and interquantile semirange equal to one and B is obtained from a Cauchy variable e independent from A such that the interquantile semirange of B is a fixed value s_B. The last series of simulations (labelled *exponential*) has A and e with shifted exponential distributions. In this case, A and e are centered at 0, variance of A is 1 and the dispersion of e is chosen to get a fixed value of σ_B.

The parameter ρ takes three values (0, 0.4 and 0.8) when our goal is the test size. Three distributions for X ($N(0,1)$, $N(2,1)$ and $Exp(\lambda = 1)$) have beed considered. The sample size may be $n = 20, 50$ and 100. So, we have 81 different situations to study the empirical sizes. The Monte-Carlo experiment was carried out 500 times for each particular scheme. The number of bootstrap replications was $B = 500$.

Table 1 summarizes the obtained results on the performance of the three statistics we used: $M_n^{*1}, M_n^{*2}, K_n^{*2}$. We have not considered K_n^{*1} due to the complexity of the requiered optimizations. To calculate M_n^{*1} and M_n^{*2} we have used the measure Q specified by H_0. The nominal size was $\alpha = 0.05$. The values typed in italic are significatively (95%) different from

α. They are 4.9 % of the total. In general, we observed good performances and a small advantage for the bootstrap schemes based on D_n^{*2}.

[Table 1 about here]

For the power study, we have considered the *normal* and *Cauchy* cases with three sample sizes ($n = 20, 50, 100$) and two distributions for X ($N(0,1)$ and $N(2,1)$). So, we observed twelve different situations under H_0; we have studied the test power against alternative hypotheses established in terms of b_0 and ρ.

The data were simulated under two sets of alternative hypotheses. First, with $b_0 = 1$, allowing ρ to take values of the form $0 \pm 0.1h, h = 1, \ldots, 9$ and σ_B or s_B equal to one. Second, with $\rho = 0$, with values $1 \pm 0.1h, h = 1, \ldots, 9$, for b_0 and σ_B or s_B equal to 0.5.

Some results are displayed in Figures 1 to 4. Graph (a) in each of these figures represents the dispersion of Y for a given value x of X when (Y, X) follows model (1.1); more precisely, in Figures 1 and 2 they are 95% prediction bands and in Figures 3 and 4, we draw interquartile semiranges around the median. Graphs (b) and (c) give some of the power functions obtained for $n = 100$ and test size $\alpha = 0.05$. Power functions for other values of n and α have also been studied and they are considered in the following comments.

[Figures 1, 2, 3 and 4 about here]

For $n = 20$ the results are not satisfactory but there is an important improvement for $n = 50$. M_n^{*1} and M_n^{*2} behave similarly. As usual in goodness of fit tests the power for Kolmogorov-Smirnov statistics is generally lower than for those based on Cramér-von Mises ones; however, in some situations, K_n^{*2} is clearly the best one. The asymmetry of some power functions can be explained in terms of the different conditional distributions of Y given X under the alternatives (see part (a) in Figures 1 to 4).

Finally, we may remark the relevance of the identifiability idea in these models. In graph 3.(a) can be seen that if the variable X takes only positive values (respectivaly, negative) then the distribution of (Y, X) is the same for all positive values of ρ (respectively, negative). This is reflected in 3.(c) where one of the branches of each power function is practically constant and equal to the theoretic size since, for those models, $X \sim N(2,1)$ and it takes positive values with high probability.

4 APPENDIX: Proofs of Theorems.

PROOF OF PROPOSITION 2.1. Since $\mathcal{F}$ can be written as $\mathcal{F} = \{f_{st}(\cdot,\cdot) = f(\cdot,\cdot,s,t) \mid (s,t) \in \mathbf{R}^{1+q}\}$, condition *(ii)* in Definition 1 in Pollard (1984), pages 195-196, follows. Only remains to check that

$$f(y,x,s,t) = \left(I_{(-\infty,s]}(y) - P_{F_{AB}}(A + xB \leq s)\right) I_{(-\infty,t]}(x),$$

is $\mathcal{B}(\mathbf{R}^{1+q}) \times \mathcal{B}(\mathbf{R}^{1+q})$-measurable, and this follows because

$$\begin{aligned} P_{F_{AB}}(A + xB \leq s) &= \int_{\mathbf{R}^{1+q}} I_{(-\infty,s]}(a + xb) dF_{AB}(a,b) = \\ &= \lim_{n\longrightarrow\infty} \int_{\mathbf{R}^{1+q}} I_{(-\infty,s]}(a + xb) dF_{AB,n}(a,b) = \\ &= \lim_{n\longrightarrow\infty} P_{F_{AB,n}}(A + xB \leq s),\ x \in \mathbf{R}^q, s \in \mathbf{R}, \end{aligned}$$

for a sequence $\{(a_n,b_n) : n \in \mathbf{N}\}$ such that $F_{AB,n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{(a_i,b_i)}$, since $P_{F_{AB}}(A+xB = s) = 0$ if *(a)* holds. Under condition *(b)*, measurability of $h(x,s) = P_{F_{AB}}(A + xB \leq s)$ is trivial. □

PROOF OF PROPOSITION 2.2. Since the functions in $\mathcal{F}$ are of the form

$$\begin{aligned} f_{st}(y,x) &= \left(I_{(-\infty,s]}(y) - P_{F_{AB}}(A + xB \leq s)\right) I_{(-\infty,t]}(x) = \\ &= I_{(-\infty,s]}(y) I_{(-\infty,t]}(x) - P_{F_{AB}}(A + xB \leq s) I_{(-\infty,t]}(x), \end{aligned}$$

the result will follow from the next three claims:

CLAIM 4.1. $\mathcal{F}_1 = \{f_s : \mathbf{R}^q \longrightarrow \mathbf{R} \mid f_s(x) = P_{F_{AB}}(A + xB \leq s), s \in \mathbf{R}\}$ *is a Vapnik-Červonenkis class of functions. Moreover, the Vapnik-Červonenkis index of the graph class is 2.*

Proof. We have to check that the graph class of functions in $\mathcal{F}_1$ does not shatter two elements sets in $\mathbf{R}^q \times \mathbf{R}$ (i.e., each two points subset A in $\mathbf{R}^q \times \mathbf{R}$ contains a subset that cannot be obtained as the intersection of A and the graph of a function in $\mathcal{F}_1$).

Let $A = \{(x_1,t_1),(x_2,t_2)\} \subset \mathbf{R}^q \times \mathbf{R}$. If $\mathcal{C} = \{G_f, f \in \mathcal{F}_1\}$ shatters A then there exists $s_1, s_2 \in \mathbf{R}^q$ such that

$$A \cap G_{f_{s_1}} = \{(x_1,t_1)\}, \text{ and} \tag{4.12}$$

$$A \cap G_{f_{s_2}} = \{(x_2,t_2)\}. \tag{4.13}$$

Since all $f \in \mathcal{F}_1$ are nonnegative functions, from (4.13) it follows that

$$(x_1,t_1) \in G_{f_{s_1}} \implies f_{s_1}(x_1) \geq t_1, \text{ and} \tag{4.14}$$

$$(x_2,t_2) \notin G_{f_{s_1}} \implies f_{s_1}(x_2) < t_2. \tag{4.15}$$

Analogously, from (4.13),

$$(x_1,t_1) \notin G_{f_{s_2}} \implies f_{s_2}(x_1) < t_1, \text{ and} \tag{4.16}$$

$$(x_2,t_2) \in G_{f_{s_2}} \implies f_{s_2}(x_2) \geq t_2. \tag{4.17}$$

From the properties of univariate distribution functions, we have that $f_s(x) \leq f_t(x)$ if, and only if, $s \leq t$ and $f_s(x) < f_t(x)$ implies $s < t$.

This, combined with (4.14) and (4.16), gives $s_1 > s_2$, and combined with (4.15) and (4.17), leads to $s_1 < s_2$. This contradiction proves the claim.

CLAIM 4.2. *If $\mathcal{F} = \{f : D_2 \to \mathbf{R}\}$ is a Vapnik-Červonenkis class of functions then $\mathcal{G} = \{g_f : D_1 \times D_2 \to \mathbf{R} \mid g_f(y,x) = f(x), f \in \mathcal{F}\}$ is also a Vapnik-Červonenkis class.*

Proof. Let $\mathcal{C}_\mathcal{F}, \mathcal{C}_\mathcal{G}$ be, respectively, the graph classes of $\mathcal{F}$ and $\mathcal{G}$. Let us see that $\mathcal{C}_\mathcal{F}$ shatters a set of k elements if, and only if, $\mathcal{C}_\mathcal{G}$ does.

Assume that $\mathcal{C}_\mathcal{F}$ shatters a set $C = \{(x_i,t_i) \mid i = 1,\ldots,k\} \subset D_2 \times \mathbf{R}$. Let us check that $\mathcal{C}_\mathcal{G}$ shatters $C_1 = \{y_0\} \times C$, $y_0 \in D_1$.

Consider $A_1 \subseteq C_1$ and define $A = \{(x,t) \in C \mid (y_0,x,t) \in A_1\}$. Since $\mathcal{C}_\mathcal{F}$ shatters C, there exists $f \in \mathcal{F}$ such that $G_f \cap C = A$; then

$$\begin{aligned} G_{g_f} \cap C_1 &= \{(y,x,t) \in C_1 \mid 0 \leq t \leq g_f(y,x) \ \text{ or } \ g_f(y,x) \leq t \leq 0\} = \\ &= \{(y,x,t) \in C_1 \mid 0 \leq t \leq f(x) \ \text{ or } \ f(x) \leq t \leq 0\} = \\ &= \{(y,x,t) \in C_1 \mid (x,t) \in G_f\} = \\ &= \{y_0\} \times \{(x,t) \in C \mid (x,t) \in G_f\} = \\ &= \{y_0\} \times \{C \cap G_f\} = \{y_0\} \times A = A_1 \end{aligned}$$

and $\mathcal{C}_\mathcal{G}$ shatters some set of k elements.

For the reciprocal, suppose that $\mathcal{C}_\mathcal{G}$ shatters a set $C = \{(y_i,x_i,t_i) : i = 1,\ldots,k\} \subset (D_1 \times D_2) \times \mathbf{R}$. Define $C_0 = \{(x,t) \mid \exists y \in D_1 \text{ such that } (y,x,t) \in C\}$; this set has k elements: if for $i \neq j$, $y_i \neq y_j$ and $(x_i,t_i) = (x_j,t_j)$ then C is not shattered by $\mathcal{C}_\mathcal{G}$. Let $A_0 \subseteq C_0$ and let $A = \{(y,x,t) \in C \mid (x,t) \in A_0\}$; there exists $g_f \in \mathcal{G}$ such that

$$\begin{aligned} A = C \cap G_{g_f} &= \{(y,x,t) \in C \mid 0 \leq t \leq g_f(y,x) \ \text{ or } \ g_f(y,x) \leq t \leq 0\} = \\ &= \{(y,x,t) \in C \mid (x,t) \in C_0, 0 \leq t \leq f(x) \ \text{ or } \ f(x) \leq t \leq 0\} = \\ &= \{(y,x,t) \in C \mid (x,t) \in C_0 \cap G_f\} \Longrightarrow A_0 = C_0 \cap G_f \end{aligned}$$

and so $\mathcal{C}_\mathcal{F}$ shatters a k elements set. The claim follows.

CLAIM 4.3. *Let $\mathcal{F}$ and $\mathcal{G}$ be classes of functions with envelopes F and G, respectively. Let $\mathcal{F} + \mathcal{G} = \{f + g \mid f \in \mathcal{F}, g \in \mathcal{G}\}$. Then $F + G$ is an envelope for $\mathcal{F} + \mathcal{G}$ and*

$$D(\varepsilon, F+G, \mathcal{F}+\mathcal{G}) \leq D(\varepsilon, F, \mathcal{F}) D(\varepsilon, G, \mathcal{G}).$$

Let $\mathcal{FG} = \{fg \mid f \in \mathcal{F}, g \in \mathcal{G}\}$. If F and G are constant functions then FG is an envelope for $\mathcal{FG}$ and

$$D(\varepsilon, FG, \mathcal{FG}) \leq D(\varepsilon/2, F, \mathcal{F}) D(\varepsilon/2, G, \mathcal{G}).$$

The first part appears in Pollard (1984), p. 40, and the second part follows in a similar way as in Pollard (1989). This proves the claim.

Our class $\mathcal{F}$ is obtained from classes $\mathcal{F}_1$, $\mathcal{J}_q$ and $\mathcal{J}_1$, where $\mathcal{J}_d$ is the class of semiinterval indicators in $\mathbf{R}^d$ which is a Vapnik-Červonenkis class with envelope $F = 1$ (see, for instance, Corollary 9.2.15. in Dudley (1984)). From Claim 4.1 and Claim 4.2, $\mathcal{F}_1$ is also a Vapnik-Červonenkis class and so, there exist positive constants A_i, w_i such that

$$D(\varepsilon, F, \mathcal{F}_1) \leq A_1 \varepsilon^{-2w_1}, \; D(\varepsilon, F, \mathcal{I}_q) \leq A_2 \varepsilon^{-2w_2}, \; D(\varepsilon, F, \mathcal{I}_1) \leq A_3 \varepsilon^{-2w_3}, \quad 0 < \varepsilon \leq 1.$$

Now, Claim 4.3 gives that

$$D(\varepsilon, F^*, \mathcal{F}) \leq (D(\varepsilon/2, F, \mathcal{F}_1) D(\varepsilon/2, F, \mathcal{I}_1)) D(\varepsilon/2, F, \mathcal{I}_q),$$

where $F^* = (F + F)F = 2$ is an envelope for $\mathcal{F}$, and so there are constants A and w such that

$$D(\varepsilon, F, \mathcal{F}) \leq A\varepsilon^{-2w}, \quad 0 < \varepsilon \leq 1.$$

□

PROOF OF THEOREM 2.1. From Propositions 2.1 and 2.2 and Pollard's central limit theorem (Pollard (1982)), it follows that $\mathcal{F}$ is a Donsker class for $\mathcal{P}(F_{AB}, F_X)$. For the covariance function, since

$$\begin{aligned} E_{\mathcal{P}(F_{AB},F_X)} f_{st} &= \int \left(I_{(-\infty,s]}(y) - P_{F_{AB}}(A + xB \leq s)\right) I_{(-\infty,t]}(x) d\mathcal{P}(F_{AB}, F_X)(y,x) = \\ &= \mathcal{P}(F_{AB}, F_X)(s,t) - \int_{(-\infty,t]} P_{F_{AB}}(A + xB \leq s) dF_X(x) = \\ &= \mathcal{P}(F_{AB}, F_X)(s,t) - \int_{\{x \leq t\} \cap \{a + xb \leq s\}} d(F_{AB} \times dF_X)(a,b,x) = \\ &= \mathcal{P}(F_{AB}, F_X)(s,t) - \mathcal{P}(F_{AB}, F_X)(s,t) = 0, \quad s \in \mathbf{R}, \; t \in \mathbf{R}^q, \end{aligned}$$

we have that

$$\begin{aligned} \operatorname{Cov}(f_{st}, f_{uv}) = E f_{st} f_{uv} - E f_{st} E f_{uv} = E f_{st} f_{uv} = \\ = \mathcal{P}(F_{AB}, F_X) \left(I_{(-\infty,(s,t)]}(Y,X) I_{(-\infty,(u,v)]}(Y,X)\right) - \\ -\mathcal{P}(F_{AB}, F_X) \left(P_{F_{AB}}(A + XB \leq s) I_{(-\infty,t]}(X) I_{(-\infty,(u,v)]}(Y,X)\right) - \\ -\mathcal{P}(F_{AB}, F_X) \left(P_{F_{AB}}(A + XB \leq u) I_{(-\infty,v]}(X) I_{(-\infty,(s,t)]}(Y,X)\right) + \\ +\mathcal{P}(F_{AB}, F_X) \left(P_{F_{AB}}(A + XB \leq s) P_{F_{AB}}(A + XB \leq u) I_{(-\infty,t]}(X) I_{(-\infty,v]}(Y,X)\right) = \end{aligned}$$

$$= S_1 - S_2 - S_3 + S_4.$$

Obviously, $S_1 = \mathcal{P}(F_{AB}, F_X)(s \wedge u, t \wedge v)$. For the second term,

$$\begin{aligned} S_2 &= \int_{\{a+xb\leq u\}\cap\{x\leq t\wedge v\}} P_{F_{AB}}(A + xB \leq s)d(F_{AB} \times dF_X)(a,b,x) = \\ &= \int_{\{x\leq t\wedge v\}} P_{F_{AB}}(A + xB \leq s)\left(\int_{\{a+xb\leq u\}} dF_{AB}(a,b)\right) dF_X(x) = \\ &= \int_{\{x\leq t\wedge v\}} P_{F_{AB}}(A + xB \leq s)P_{F_{AB}}(A + xB \leq u)dF_X(x). \end{aligned}$$

Moreover, $S_3 = S_2$ and $S_4 = S_2$. So,

$$\text{Cov}(f_{st}, f_{uv}) =$$

$$= \mathcal{P}(F_{AB}, F_X)(s \wedge u, t \wedge v) - \int_{\{x\leq t\wedge v\}} P_{F_{AB}}(A + xB \leq s)P_{F_{AB}}(A + xB \leq u)dF_X(x) =$$

$$= \int_{\{x\leq t\wedge v\}} (P_{F_{AB}}(A + xB \leq s \wedge u) - P_{F_{AB}}(A + xB \leq s)P_{F_{AB}}(A + xB \leq u))\, dF_X(x).$$

□

PROOF OF COROLLARY 2.1. $\|\cdot\|_{\mathcal{F}}$ is trivially continuous in $l^\infty(\mathcal{F})$. Let

$$\|\Psi\|_{2,Q} = \left(\int_{\mathbf{R}^{1+q}} (\Psi(f_{st}))^2 dQ(s,t)\right)^{\frac{1}{2}}$$

for all $\Psi \in l^\infty(\mathcal{F})$. Thus, $M_n = \|D_n\|_{2,Q}$. To check that the norm $\|\cdot\|_{2,Q}$ is a continuous functional in $l^\infty(\mathcal{F})$, take a sequence $\{\Psi_n\}$ such that $\|\Psi_n - \Psi_0\|_{\mathcal{F}} \longrightarrow 0$. Then

$$|\|\Psi_n\|_{2,Q} - \|\Psi_0\|_{2,Q}| \leq \|\Psi_n - \Psi_0\|_{2,Q} =$$

$$= \left(\int_{\mathbf{R}^{1+q}} | \Psi_n(f_{st}) - \Psi_0(f_{st}) |^2 \, dQ(s,t)\right)^{\frac{1}{2}} \leq \|\Psi_n - \Psi_0\|_{\mathcal{F}} Q(\mathbf{R}^{1+q}) \longrightarrow 0.$$

□

PROOF OF THEOREM 2.2. The asymptotic behaviour of D_n^{*1} follows from Theorem 2.4 in Giné and Zinn (1990) because $\mathcal{F}$ is a Donsker class of functions for $\mathcal{P}(F_{AB}, F_X)$.

To establish that D_n^{*2} tends to the Brownian bridge $Z_{\mathcal{P}(F_{AB},F_X)}$, we use Corollary 2.7 in Giné and Zinn (1991). With their notation, taking $D_n^{*2} = \nu_n^{R_n}$, with $R_n = \mathcal{P}(F_{AB}, F_{X,n})$ and $R_0 = \mathcal{P}(F_{AB}, F_X)$, we have to prove that $\|R_n - R_0\|_{\mathcal{G}} \longrightarrow 0$, with $\mathcal{G} = \mathcal{F} \cup \mathcal{F}' \cup \mathcal{F}^2 \cup (\mathcal{F}')^2$, where $\mathcal{F}'$ is the class of differences of functions in $\mathcal{F}$, and $\mathcal{F}^2$ is the set of squares of the elements of $\mathcal{F}$. It is enough to see that the supremum on each of $\mathcal{G} = \mathcal{F}$, $\mathcal{F}'$, $\mathcal{F}^2$ and $(\mathcal{F}')^2$ tends to zero.

We have that $R_n(f_{st}) = F_{X,n}(r_{st})$ and $R_0(f_{st}) = F_X(r_{st})$, with

$$\begin{aligned} r_{st}(x) &= \int f_{st}(y,x)d\mathcal{P}(F_{AB}, \delta_x)(y,x) = E_{F_{AB}}[f_{st}(A + xB, x)] = \\ &= (P_{F_{AB}}(A + xB, x) - P_{F_{AB}}(A + xB, x))\, I_{(-\infty,t]}(x) = 0, \quad s \in \mathbf{R},\ t \in \mathbf{R}^q. \end{aligned}$$

Thus, $\|R_n - R_0\|_{\mathcal{F}} = 0$. Similarly $R_n(f_{st} - f_{uv}) = R_0(f_{st} - f_{uv}) = 0$ and $\|R_n - R_0\|_{\mathcal{F}'} = 0$.

For the convergence of $\|R_n - R_0\|_{\mathcal{F}^2}$, note that $\|R_n - R_0\|_{\mathcal{F}^2} \leq \|R_n - R_0\|_{\mathcal{F}\mathcal{F}}$. Now, let

$$\begin{aligned} r_{stuv}(x) &= \int f_{st}(y,x) f_{uv}(y,x) d\mathcal{P}(F_{AB}, \delta_x)(y,x) = \\ &= E_{F_{AB}}[f_{st}(A + xB, x) f_{uv}(A + xB, x)]. \end{aligned}$$

Since

$$f_{st}(y,x) f_{uv}(y,x) = \Big(I_{(-\infty, u\wedge s]}(y) - P_{F_{AB}}(A + xB \leq u) I_{(-\infty,s]}(y) -$$

$$- P_{F_{AB}}(A + xB \leq s) I_{(-\infty,u]}(y) + P_{F_{AB}}(A + xB \leq u) P_{F_{AB}}(A + xB \leq s)\Big) I_{(-\infty, t\wedge v]}(x),$$

we get that

$$r_{stuv}(x) = (P_{F_{AB}}(A + xB \leq u \wedge s) -$$

$$- P_{F_{AB}}(A + xB \leq u) P_{F_{AB}}(A + xB \leq s))\, I_{(-\infty, t\wedge v]}(x).$$

As we saw along the proof of Theorem 2.1, the class $\mathcal{R}$ of functions r_{stuv} is a Donsker class for F_X. Since

$$(R_n - R_0)(f_{st} f_{uv}) = \frac{1}{\sqrt{n}} \nu_n^{F_X}(r_{stuv}),$$

we have that $\|R_n - R_0\|_{\mathcal{F}\mathcal{F}} = n^{-1/2} \|\nu_n^{F_X}\|_{\mathcal{R}} \longrightarrow 0$, because $\nu_n^{F_X} \longrightarrow_w Z_{F_X}$ in $l^\infty(\mathcal{R})$. Analogously, $\|R_n - R_0\|_{(\mathcal{F}')^2} \leq 4 \|R_n - R_0\|_{\mathcal{F}\mathcal{F}} \longrightarrow 0$. □

PROOF OF THEOREM 2.3. Let us prove *(i)* first. From the Mean Value Theorem and since the partial derivatives of h are bounded, it follows that $\mathcal{F}_p$ is uniformly equicontinuous: for any $\varepsilon > 0$ there exists $\delta = \varepsilon / M_1$ such that $\|x - x'\| \leq \delta$ implies $|f_s(x) - f_s(x')| < \varepsilon$, for all $s \in \mathbf{R}^p$.

Let V be a p-dimensional random variable and let $[a,b]_p = \{v \in \mathbf{R}^p \mid a \leq v \leq b\}$. We can write

$$P_V([a,b]_p) = \sum_{j=1}^{r} \alpha_j F_V(m_j(a,b)),$$

where $m_j(a,b)$ are vertices of $[a,b]_p$. The coefficients α_j, the vertices $m_j(a,b)$ and the value r only depend on the dimension p.

Given $\varepsilon > 0$, let $\delta = \varepsilon (2M_1 \sum_{j=1}^r |\alpha_j|)^{-1}$. Since $\mathrm{Supp}(X)$ is totally bounded, there exists $n = n(\delta)$ and points $x_1, \ldots, x_n \in \mathrm{Supp}(X)$, such that for all $x \in \mathrm{Supp}(X)$, $\min_{i=1,\ldots,n} \|x - x_i\| \leq \delta$. For each x_i there exists a set $[l_i, u_i]_p$ with $P(Y_{x_i} \in [l_i, u_i]_p) \geq 1 - \varepsilon/2$. Let $l = l(\varepsilon)$ (respectively, $u = u(\varepsilon)$) the point in $\mathbf{R}^p$ whose j-th coordinate is the smallest (respectively, the largest) of the j-th coordinates of the points $l_i, i = 1, \ldots, n$ (respectively, $u_i, i = 1, \ldots, n$). Thus, $P(Y_{x_i} \in [l,u]_p) \geq 1 - \frac{\varepsilon}{2}$, $i = 1, \ldots, n$. For $x \in \mathrm{Supp}(X)$, $P(Y_x \in$

$[l,u]_p) = \sum_{j=1}^r \alpha_j F_{Y_x}(m_j(l,u))$. Let $i = i(x)$ be the index of the closest point to x among $x_1, \ldots, x_n$. Then

$$|P(Y_x \in [l,u]_p) - P(Y_{x_i} \in [l,u]_p)| \leq \sum_{j=1}^r |\alpha_j||F_{Y_x}(m_j(l,u)) - F_{Y_{x_i}}(m_j(l,u))| \leq$$

$$\leq \sum_{j=1}^r |\alpha_j| \frac{\varepsilon}{2\sum_{j=1}^r |\alpha_j|} = \frac{\varepsilon}{2}$$

and, so $P(Y_x \in [l,u]_p) \geq 1 - \varepsilon$, for all $x \in \text{Supp}(X)$. This proves *(i)*.

Let us now establish *(ii)*. From uniform boundedness of the partial derivatives of h, for each $\varepsilon > 0$, there exists $\gamma = \varepsilon/(2M_2)$ such that $\|s - s'\| \leq \gamma$ implies $|h(x,s) - h(x,s')| \leq \varepsilon/2$ for all $x \in \text{Supp}(X)$.

Using γ and the compact $[l(\varepsilon/2), u(\varepsilon/2)]_p$, we define the following points in $\mathbf{R}^p$:

$$s(i_1, \ldots, i_p) = l(\varepsilon/2) + \left(\frac{i_1(u_1 - l_1)}{n_1}, \ldots, \frac{i_p(u_p - l_p)}{n_p}\right), \quad n_j = \left[\frac{u_j - l_j}{\gamma/p}\right] + 1,$$

$$i_j = 1, \ldots, n_j, \quad j = 1, \ldots, p.$$

If $N(\varepsilon)$ is the number of points just defined, we have

$$N(\varepsilon) = \prod_{j=1}^p (n_j + 1) = \prod_{j=1}^p \left(\left[\frac{u_j - l_j}{\gamma/p}\right] + 2\right) \leq$$

$$\leq \prod_{j=1}^p \left(\frac{p(u_j - l_j)}{\gamma} + 2\right) \leq p^p \gamma^{-p} \text{Vol}([l(\varepsilon/2), u(\varepsilon/2) + 2\gamma 1_p]_p)$$

where 1_p is the vector with all coordinates equal to one.

Since $\|s - t\| \leq p \max_{j=1,\ldots,p} |s_j - t_j|$, for all $s \in [l(\varepsilon/2), u(\varepsilon/2)]_p$ there exists s' among the $N(\varepsilon)$ points with $\|s - s'\| \leq \gamma$ and so, $|h(x,s) - h(x,s')| \leq \varepsilon/2$ for all $x \in \text{Supp}(X)$. If $s \notin [l(\varepsilon/2), u(\varepsilon/2)]_p$, let $\bar{s}$ be the closest point to s in $[l(\varepsilon/2), u(\varepsilon/2)]_p$, and let s' be the point closest to $\bar{s}$ among the $N(\varepsilon)$ just defined. We have that

$$|h(x,s) - h(x,s')| \leq |h(x,s) - h(x,\bar{s})| + |h(x,\bar{s}) - h(x,s')| \leq$$

$$\leq |h(x,s) - h(x,\bar{s})| + \frac{\varepsilon}{2} = |F_{Y_x}(s) - F_{Y_x}(\bar{s})| + \frac{\varepsilon}{2}.$$

Note that

$$F_{Y_x}(s) = P(Y_x \in (-\infty, s]) =$$

$$= P(Y_x \in (-\infty, s] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p) + P(Y_x \in (-\infty, s] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p^c)$$

The difference between the sets $(-\infty, s] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p$ and $(-\infty, \bar{s}] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p$ is included in the boundary of $[l(\varepsilon/2), u(\varepsilon/2)]_p$ and so, it has zero probability under the absolutely continuous distribution of Y_x, for all $x \in \text{Supp}(X)$. Thus

$$\begin{aligned} F_{Y_x}(s) &= P(Y_x \in (-\infty, \bar{s}] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p) + \\ &\quad + P(Y_x \in (-\infty, s] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p^c) = \\ &= F_{Y_x}(\bar{s}) + P(Y_x \in (-\infty, s] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p^c) - \\ &\quad - P(Y_x \in (-\infty, \bar{s}] \cap [l(\varepsilon/2), u(\varepsilon/2)]_p^c). \end{aligned}$$

Then, for all $x \in \text{Supp}(X)$,

$$|h(x,s) - h(x,s')| \leq \frac{\varepsilon}{2} + P(Y_x \notin [l(\varepsilon/2), u(\varepsilon/2)]_p) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

It follows that for a finite $A \subset \text{Supp}(X)$,

$$\min_{i=1,\ldots,N(\varepsilon)} \sum_{x \in A} (f_s(x) - f_{s_i}(x))^2 \leq \varepsilon^2 \text{card}(A), \ s \in \mathbf{R}^p,$$

and then

$$\begin{aligned} D(\varepsilon, F, \mathcal{F}_p) \leq N(\varepsilon) &\leq p^p \gamma^{-p} \text{Vol}([l(\varepsilon/2), u(\varepsilon/2) + 2\gamma 1_p]_p) = \\ &= (2M_2 p)^p \varepsilon^{-p} \text{Vol}([l(\varepsilon/2), u(\varepsilon/2) + \frac{\varepsilon}{M_2} 1_p]_p). \end{aligned}$$

This proves *(ii)*, and *(iii)* follows. □

PROOF OF PROPOSITION 2.3. The compactness of Q follows from a standard subsequences argument. Moreover, it follows that

$$\begin{aligned} P(Y_x \in Q) \geq P(Y_x \in K_x) &= P(A + xB \in K_x) = \\ &= \int_{\{(a,b) | a + xb \in K_x\}} dF_{AB}(a,b) \geq 1, \end{aligned}$$

since $\text{Supp}(A,B) \subseteq \{(a,b) \mid a + xb \in K_x\}$ and the result follows. □

PROOF OF PROPOSITION 2.4. Let $(A,B)^t \sim N_{p+q}((\mu_A, \mu_B)^t, \Sigma)$ and let $Y_x = A + xB \sim N_p(\mu_x = \mu_A + x\mu_B, \Sigma_x = (I_p, x)\Sigma(I_p, x)^t)$ with density function

$$f_{Y_x}(y) = f_x(y) = (2\pi)^{-p/2} |\Sigma_x|^{-1/2} \exp\{-\frac{1}{2}(y - \mu_x)^t \Sigma_x^{-1} (y - \mu_x)\}.$$

We have to show that the partial derivatives with respect to $x_i, i = 1, \ldots, p \times q$ and $s_j, j = 1, \ldots, p$ of the function $h(x,s) = \int_{\{y \leq s\}} f_x(y) dy$ are uniformly bounded for all $x \in \text{Supp}(X)$ and all $s \in \mathbf{R}^p$.

Let $V = (V_1, \ldots, V_k)$ be an absolutely continuous random variable with distribution and density functions F_V and f_V, respectively. It is straightforward to check that, under regularity conditions (fulfilled by the normal distribution) which allow interchanging integrals and derivatives

$$\frac{\partial F_V}{\partial v_k}(v_1, \ldots, v_k) = F_{V_1,\ldots,V_{k-1}|V_k = v_k}(v_1, \ldots, v_{k-1}) f_{V_k}(v_k),$$

and analogously for j, $j = 1, \ldots, k-1$. Applying this to $V = Y_x$, we have

$$\left|\frac{\partial h(x,s)}{\partial s_j}\right| \le f_{V_j}(\mu_{x,j}) = \frac{1}{\sqrt{2\pi}\sigma_{x,j}} = g_j(x),\ j = 1, \ldots, p.$$

Since the j-th component variance is a continuous function of x, $g_j(x)$ is bounded on the compact set $\mathrm{Supp}(X)$.

Now, let us turn to the partial derivatives with respect to x_i which, under regularity conditions —fulfilled by the normal distribution—, are

$$\frac{\partial h(x,s)}{\partial x_i} = \int_{\{y \le s\}} \frac{\partial f_x(y)}{\partial x_i} dy,\ i = 1, \ldots, p \times q.$$

We have that

$$\frac{\partial f_x(y)}{\partial x_i} = (2\pi)^{-p/2} \sum_{r=1}^{p}\sum_{l=1}^{p} \left[\left(\frac{\partial}{\partial \sigma_{rl}} |\Sigma_x|^{-1/2} \right) \exp\{-\frac{1}{2}(y-\mu_x)^t \Sigma_x^{-1}(y-\mu_x)\} - \right.$$

$$\left. -\frac{1}{2}|\Sigma_x|^{-1/2} \exp\{-\frac{1}{2}(y-\mu_x)^t\Sigma_x^{-1}(y-\mu_x)\} \frac{\partial}{\partial \sigma_{rl}} \left((y-\mu_x)^t \Sigma_x^{-1}(y-\mu_x) \right) \right] \frac{\partial \sigma_{rl}}{\partial x_i} -$$

$$-\frac{(2\pi)^{-p/2}}{2} |\Sigma_x|^{-1/2} \exp\{-\frac{1}{2}(y-\mu_x)^t\Sigma_x^{-1}(y-\mu_x)\} \sum_{j=1}^{p} \left(\frac{\partial}{\partial \mu_j} ((y-\mu_x)^t \left(\Sigma_x^{-1}(y-\mu_x) \right)) \right) \frac{\partial \mu_j}{\partial x_i} =$$

$$= S_1(y) + S_2(y) + S_3(y).$$

Let us study each of these three terms.

(a)

$$S_1(y) = \sum_{r=1}^{p}\sum_{l=1}^{p} \left(\frac{\partial}{\partial \sigma_{rl}} |\Sigma_x|^{-1/2} \right) \frac{\partial \sigma_{rl}}{\partial x_i} |\Sigma_x|^{1/2} f_x(y).$$

For a non-singular matrix, the partial derivatives of the determinant and the partial derivatives of the inverse matrix are continuous functions of the matrix elements (see, e.g., Mardia, Kent, and Bibby (1979)). So $|S_1(y)| \le K_1 f_x(y)$ where

$$K_1 = \max_{x \in \mathrm{Supp}(X)} \sum_{r=1}^{p}\sum_{l=1}^{p} |\Sigma_x|^{1/2} \left| \frac{\partial}{\partial \sigma_{rl}} |\Sigma_x|^{-1/2} \right| \left| \frac{\partial \sigma_{rl}}{\partial x_i} \right| < \infty.$$

(b)

$$S_2(y) = -\frac{1}{2} \sum_{r=1}^{p}\sum_{l=1}^{p} \left[\frac{\partial}{\partial \sigma_{rl}} \left((y-\mu_x)^t \Sigma_x^{-1}(y-\mu_x) \right) \right] \frac{\partial \sigma_{rl}}{\partial x_i} f_x(y).$$

As before, $|S_2(y)| \le \frac{1}{2}\sum_{r=1}^{p}\sum_{l=1}^{p} \left| \frac{\partial}{\partial \sigma_{rl}} \left((y-\mu_x)^t\Sigma_x^{-1}(y-\mu_x) \right) \right| K_2 f_x(y)$ with

$$K_2 = \max_{r,l} \max_{x \in \mathrm{Supp}(X)} \left| \frac{\partial \sigma_{rl}}{\partial x_i} \right| < \infty.$$

Let $A = \Sigma_x^{-1}$ with elements a_{hk}. Since Σ_x is non-singular, the partial derivatives of a_{hk} with respect to σ_{rl} are continuous function of x and so, bounded in Supp(X); let K_3 be an upper bound of this quantity for all h, k, r, l. Then

$$\begin{aligned}\int_{\{y\leq s\}} |S_2(y)|dy &\leq \frac{K_2K_3}{2}\sum_r\sum_l\sum_h\sum_k \int_{\{y\leq s\}} |(y-\mu_x)(y-\mu_x)^t|_{hk} f_x(y)dy \leq \\ &\leq \frac{K_2K_3p^2}{2}\sum_h\sum_k \int |(y-\mu_x)(y-\mu_x)^t|_{hk} f_x(y)dy \\ &\leq \frac{K_2K_3p^2}{2}\sum_h\sum_k (\sigma^2_{x,k}+\sigma^2_{x,k}) \leq \frac{K_2K_3K_4p^2}{2},\end{aligned}$$

where $K_4 = \max_{h,k}\max_{x\in \mathrm{Supp}(X)}(\sigma^2_{x,k}+\sigma^2_{x,k}) < \infty$.

(c)

$$S_3(y) = -\frac{1}{2}(2\pi)^{p/2}\sum_{j=1}^{p}\left[\frac{\partial}{\partial\mu_j}\left((y-\mu_x)^t\Sigma_x^{-1}(y-\mu_x)\right)\right]\frac{\partial\mu_j}{\partial x_i}f_x(y).$$

As before,

$$|S_3(y)| \leq \frac{K_5}{2}\sum_{j=1}^{p}\left|\frac{\partial}{\partial\mu_j}\left((y-\mu_x)^t\Sigma_x^{-1}(y-\mu_x)\right)\right| f_x(y)$$

with $K_5 = \max_j \max_{x\in \mathrm{Supp}(X)} |\frac{\partial\mu_j}{\partial x_i}| < \infty$. Now,

$$\begin{aligned}\int_{\{y\leq s\}} |S_3(y)|dy &\leq K_5\sum_j \int |\Sigma_x^{-1}(y-\mu_x)|_j f_x(y)dy \leq \\ &\leq K_5K_6\sum_{j=1}^{p}\sum_{h=1}^{p}\int |(y_h-\mu_h)| f_x(y)dy \leq \\ &\leq K_5K_6\sum_{j=1}^{p}\left[p+\sum_{h=1}^{p}\sigma^2_{x,h}\right] < \infty,\end{aligned}$$

where $K_6 = \max_{r,l}\max_{x\in \mathrm{Supp}(X)}|\Sigma_x^{-1}|_{rl}$.

This proves part *(i)*.

Let us see now the proof of *(ii)*. The set $\{\mu_x = \mu_A + x\mu_B \mid x \in \mathrm{Supp}(X)\}$ is compact (since Supp(X) is) and so there exists $M > 0$ such that $\|\mu_x\| \leq M$ for all $x \in \mathrm{Supp}(X)$. For each x, define $R_x(\varepsilon) = \{y \in \mathbf{R}^p \mid f_x(y) \geq \gamma\} = \{y \in \mathbf{R}^p \mid (y-\mu_x)^t\Sigma_x^{-1}(y-\mu_x) \leq \delta^2\}$, where $\delta^2 = \delta^2(\varepsilon) = F^{-1}_{\chi^2_p}(1-\varepsilon)$ and thus $P(Y_x \in R_x(\varepsilon)) = 1-\varepsilon$. Let λ_x be the largest eigenvalue of Σ_x; since the implicit function theorem, λ_x is a continuous function of x and so bounded on Supp(X); let λ be an upper bound. For the hypercube $C(\varepsilon) = \prod_{i=1}^{p}[-M-\delta\sqrt{\lambda}, M+\delta\sqrt{\lambda}+\varepsilon/M_2]$, $R_x(\varepsilon) \subset C(\varepsilon)$ and then $P(Y_x \in C(\varepsilon)) \geq 1-\varepsilon$, for all $x \in \mathrm{Supp}(X)$. Moreover,

$$v(\varepsilon) = \mathrm{Vol}(C(\varepsilon)) \leq (2M + 2\sqrt{\lambda}\delta(\varepsilon) + 1/M_2)^p \leq (a_0 + a_1\delta(\varepsilon))^p$$

for some $a_0 > 1$, a_1 not depending on ε. Thus

$$\int_0^1 \log v(\varepsilon/2)d\varepsilon \leq p\int_0^1 \log(a_0 + a_1\delta(\varepsilon/2))d\varepsilon.$$

Changing variable to $u = \delta^2(\varepsilon/2) = F^{-1}_{\chi^2_p}(1-\varepsilon/2)$, this last integral equals

$$2\int_{u_0}^{\infty} \log(a_0 + a_1\sqrt{u}) f_{\chi^2_p}(u)du \leq 2\int_0^{\infty} (a_0 + a_1(1+u)) f_{\chi^2_p}(u)du = 2a_0 + 2a_1(1+E(\chi^2_p)) < \infty.$$

This ends the proof. □

References

Beran, R. (1991). *Prediction in random coefficient regression.* Technical Report 299, Dpt. Stat., Univ. California, Berkeley.

Beran, R. and Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions. *Annals of Statistics*, 20:1970–1984.

Beran, R. and Millar, P. W. (1991). *Minimum distance estimation in random coefficient regression.* Technical Report 326, Dpt. Stat., Univ. California, Berkeley.

Chow, G. C. (1983). Random and changing coefficient models. In Z. Griliches, M. D. Intriligator, editor, *Handbook of Econometrics, Vol. II*, p. 1213–1245, North-Holland, Amsterdam.

Dudley, R. M. (1984). A course on empirical processes. In *École d'Été de Probabilités de Saint-Flour XII—1982. Lecture Notes in Mathematics 1097*, p. 2–142, Springer-Verlag, New York.

van Es, A. J. (1991). Uniform deconvolution: nonparametric maximum likelihood and inverse estimation. In *Nonparametric Functional Estimation and Related Topics*, Kluwer Academic Publishers, Dordrecht.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, 19:1275–1272.

Giné, E. and Zinn, J. (1986). Lectures on the central limit theorem for empirical processes. In *Probabily and Banach Spaces. Lecture Notes in Mathematics 1221*, p. 50–113, Springer-Verlag, Berlin.

Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability*, 18:851–869.

Giné, E. and Zinn, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Annals of Probability*, 19:758–782.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Statistics.* Academic Press.

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

Nicholls, D. F. and Quinn, B. N. (1982). *Random Coefficient Autoregressive Models: An Introduction.* Springer-Verlag, New York.

Pollard, D. (1982). A central limit theorem for empirical processes. *Journal of Australian Mathematical Society, Ser. A*, 33:235–248.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science*, 4:341–366.

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

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

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
28903 GETAFE, Madrid, Spain

(A,B)	ρ	$X \sim N(0,1)$			$X \sim N(2,1)$			$X \sim Exp(1)$		
		20	50	100	20	50	100	20	50	100
		.076	.060	.054	.050	.046	.056	.054	.054	.062
	0	.050	.066	.054	.038	.050	.048	.046	.056	.058
		.054	.064	.050	.044	.042	.042	.048	.044	.056
		.054	.042	.048	.052	.048	.060	.064	.060	*.070*
N	.4	.044	.044	.048	.040	.050	.048	.046	.068	.064
		.058	.034	.062	.050	.056	.050	.046	.050	*.084*
		.054	.032	.048	.060	.050	*.030*	.054	.060	.058
	.8	.040	*.030*	.046	.042	.044	.032	.042	.060	.050
		.038	.034	*.028*	.054	.042	.054	.058	.050	.050
		.054	.062	.058	.054	.042	.048	.050	.054	*.028*
	0	.042	.054	.046	.046	.044	.044	.040	.048	.036
		.040	.044	*.030*	.050	.062	.054	.046	.048	.054
		.054	.062	.048	.058	.054	.050	.064	.050	.056
C	.4	.044	.066	.048	.044	.054	.056	.046	.058	.066
		.040	.052	.060	.046	.050	.064	.052	.066	.054
		.078	*.082*	.042	.060	.056	.054	.066	.068	.056
	.8	.046	.064	.038	.046	.058	.054	.056	.056	.052
		.052	.054	.066	.038	.058	.052	.050	.062	.062
		.078	.058	.052	.050	.056	.048	.038	.056	.046
	0	.060	.054	.052	.042	.066	.046	.044	.046	.046
		.052	.060	.062	.056	.060	.058	.036	.042	.058
		.060	.054	.048	.056	.044	.060	*.072*	.060	.054
E	.4	.052	.044	.046	.046	.048	.060	.050	.052	.046
		.050	.050	.034	.046	.052	.062	.042	.054	.058
		.066	.054	.052	.048	.038	.052	.064	*.074*	.048
	.8	.056	.060	.056	.032	*.026*	.050	.052	.062	.046
		.046	.054	.046	.052	.046	.042	.062	.056	.062

Table 1: Tests sizes. $\alpha = .05, B = 500$, 500 simulations. The three values appearing in each cell are α_M^1, α_M^2 and α_K^2, from top to bottom.

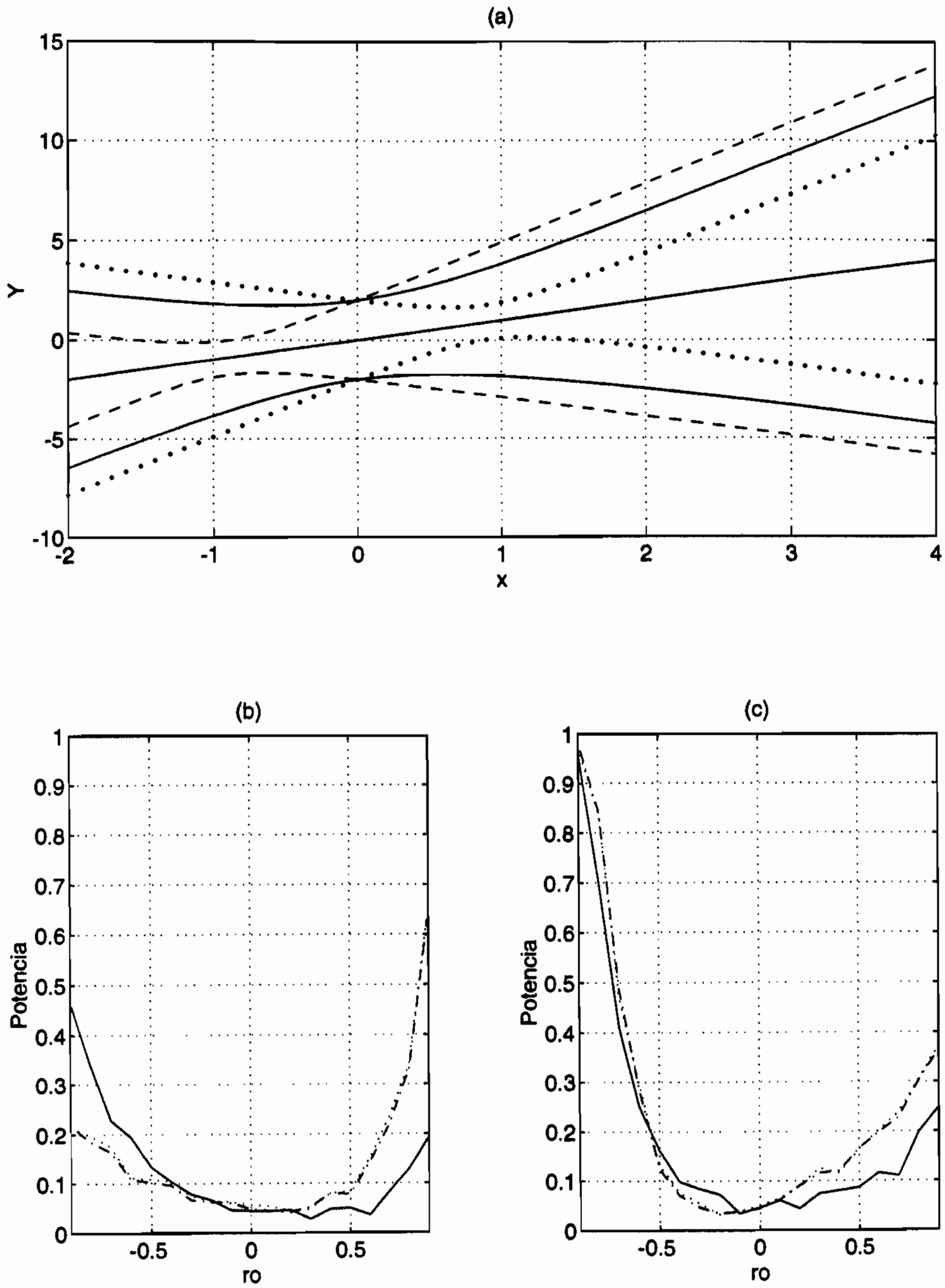

Figure 1: *Normal* model varying ρ in the alternative.

(a) Dispersion of Y given values x of X under H_0 and some alternative hypotheses.

$\cdots\cdots$ $\rho = -0.9$, —— $\rho = 0$, - - - - $\rho = 0.9$

(b), (c) Power function for $X \sim N(0,1)$ and $X \sim N(2,1)$, respectively.

$\cdots\cdots$ M_n^{*1}, $- \cdot - \cdot -$ M_n^{*2}, —— K_n^{*2}

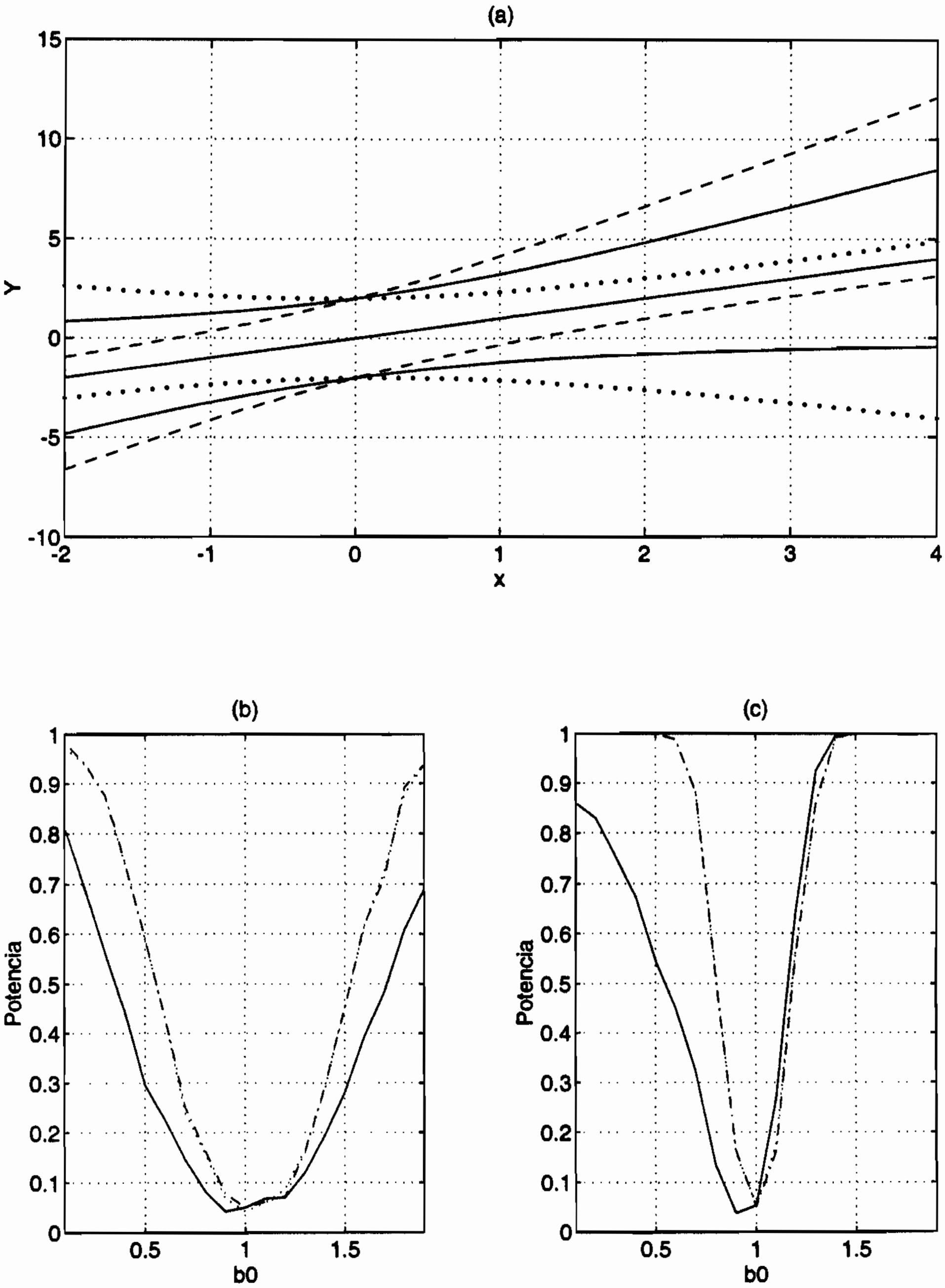

Figure 2: *Normal* model varying b_0 (mean of B) in the alternative.

(a) Dispersion of Y given values x of X under H_0 and some alternative hypotheses.

$$\cdots\cdots\ b_0 = 0.1, \text{———}\ b_0 = 1, \text{- - - -}\ b_0 = 1.9$$

(b), (c) Power function for $X \sim N(0,1)$ and $X \sim N(2,1)$, respectively.

$$\cdots\cdots\ M_n^{*1}, -\cdot-\cdot-\ M_n^{*2}, \text{———}\ K_n^{*2}$$

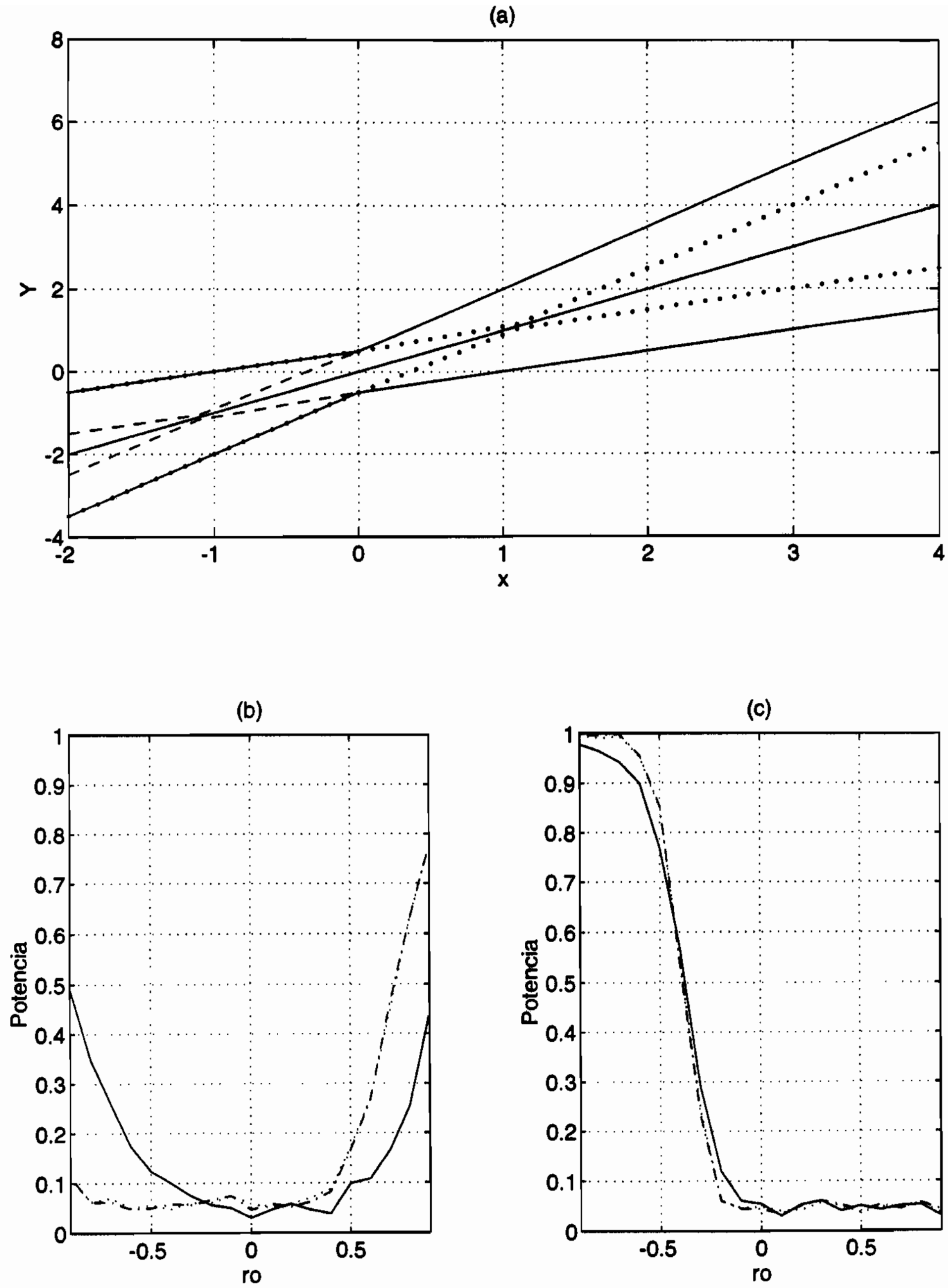

Figure 3: *Cauchy* model varying ρ in the alternative.

(a) Dispersion of Y given values x of X under H_0 and some alternative hypotheses.

$\cdots\cdots\ \rho = -0.9$, —— $\rho = 0$, - - - - $\rho = 0.9$

(b), (c) Power function for $X \sim N(0,1)$ and $X \sim N(2,1)$, respectively.

$\cdots\cdots\ M_n^{*1}$, $- \cdot - \cdot -\ M_n^{*2}$, —— K_n^{*2}

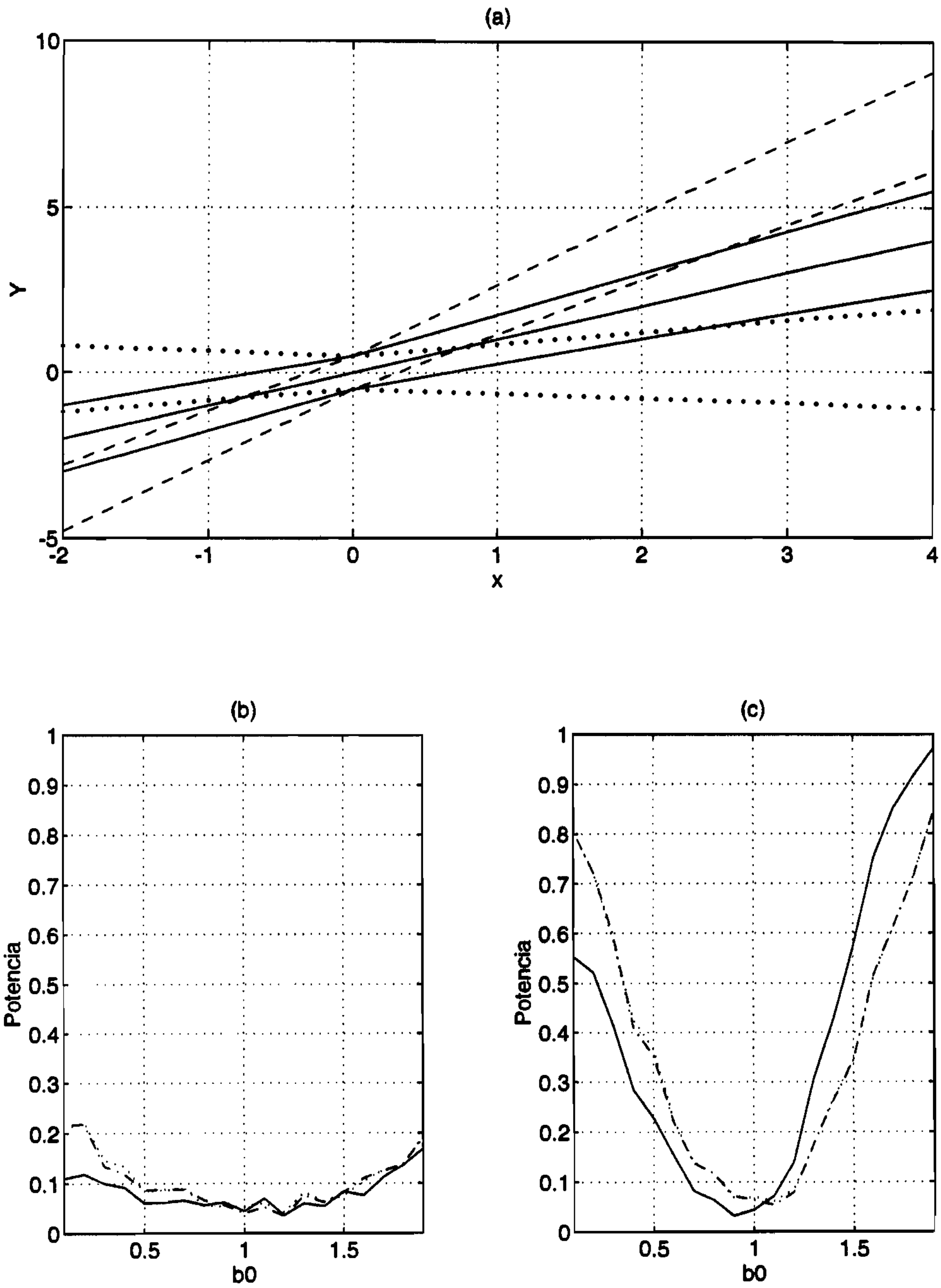

Figure 4: *Cauchy* model varying b_0 (median of B) in the alternative.

(a) Dispersion of Y given values x of X under H_0 and some alternative hypotheses.

$$\cdots\cdots\ b_0 = 0.1, \text{———}\ b_0 = 1, \text{- - - -}\ b_0 = 1.9$$

(b), (c) Power function for $X \sim N(0,1)$ and $X \sim N(2,1)$, respectively.

$$\cdots\cdots\ M_n^{*1}, \text{— · — · —}\ M_n^{*2}, \text{———}\ K_n^{*2}$$