CONSTANT COEFFICIENT TESTS FOR RANDOM COEFFICIENT REGRESSION *

Pedro Delicado
Universitat Pompeu Fabra

Juan Romo
Universidad Carlos III de Madrid

Abstract

Random coefficient regression models have been applied in different fields and they constitute a unifying setup for many statistical problems. The nonparametric study of this model started with Beran and Hall (1992) and it has become a fruitful framework. In this paper we propose and study statistics for testing a basic hypothesis concerning this model: the constancy of coefficients. The asymptotic behavior of the statistics is investigated and bootstrap approximations are used in order to determine the critical values of the test statistics. A simulation study illustrates the performance of the proposals.

Key words:
Goodness-of-fit, linear regression, random coefficients.

1 Introduction

Random coefficient regression models have been widely applied from biology to image compression to econometrics. From a theoretical point of view, they are a unifying frame for different important models as random effects in ANOVA, deconvolution models, heteroscedastic linear models or location-scale mixture models.

The nonparametric study of the random coefficient linear regression model has been recently consider by Beran and Hall (1992), Beran and Millar (1994), and Beran, Feuerverger, and Hall (1996). Let

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

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. $\{(A_i, B_i, X_i) : i \geq 1\}$ 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 observe a sample of n pairs (Y_i, X_i), $i \geq 1$. F_{AB} is the distribution of the coefficients (A, B) and F_X is that of X. The joint distribution of (Y_i, X_i) depends on these distributions and will be denoted $F_{YX} \equiv \mathcal{P}(F_{AB}, F_X)$. Let $P_n = n^{-1} \sum_{i=1}^n \delta_{(Y_i, X_i)}$ and $F_{X,n} = n^{-1} \sum_{i=1}^n \delta_{X_i}$ be the

*Research partially supported by DGES 96-0300 and DGICYT PB96-0111 (Spain).

empirical distributions associated to the observations (Y_i, X_i) and X_i, respectively.

A basic question about these models is to check the constancy of coefficients; this means to test if F_B, the distribution of B, is degenerated. In other words, to choose among

$$\begin{array}{lcl} H_0 & : & F_B = \delta_b, \quad \text{for some } b \in \mathbb{R}^q, \text{ and} \\ H_1 & : & P_{F_B}(B = b) < 1, \quad \text{for all } b \in \mathbb{R}^q. \end{array} \tag{1.2}$$

The existing tests to check for constant coefficients are based on the hypothesis of variances $\alpha_i = \operatorname{Var}(B_i)$, $i = 1, \ldots, q$, simultaneously equal to zero. The usual theory associated to maximum likelihood is not valid here because, under H_0, the parameter vector is on the parametric space boundary. Goodness-of-fit tests developed in Delicado and Romo (1997, 1998) would allow to test a less general hypothesis than (1.2): if we fix a parametric family for the distribution of A, we could test

$$H_0 \quad : \quad F_{AB} = F_A \otimes \delta_b, \quad F_A \in \boldsymbol{F}_{\boldsymbol{\Theta}}, b \in \mathbb{R}^q. \tag{1.3}$$

If there is no evidence to reject H_0, there will be no need to assume random coefficients in the model. However, if H_0 is rejected, it is not clear that it is exclusively due to the randomness of B; it could also happen that the distribution of A is far from belonging to $\boldsymbol{F}_{\boldsymbol{\Theta}}$.

This paper is organized as follows. In the remaining of this section we introduce some definitions and concepts from the theory of empirical processes and U-processes. Section 2 proposes a constant coefficient test based on U-processes. In Section 3 we explore two additional tests based on Kolmogorov-Smirnov's and Sukhatme's two-sample tests. Section 4 presents the results of a simulation study. Finally, all the proofs are collected in an Appendix.

1.1 Preliminaries

We recall now some concepts and results from the theory of empirical processes and U-processes which will be used below. For any signed measure μ and any measurable function f, we denote the integral of f with respect to μ by μf. We follow the general definition of *weak convergence* in Hoffmann-Jørgensen (1984).

Let $(S, \mathcal{S}, Q)$ be a probability space and let $\{Z_i\}_{i=1}^{\infty}$ be a sequence of independent and identically distributed random variables defined on S with common distribution Q. The random measure Q_n giving mass $1/n$ to each of these observations $Z_1, \ldots, Z_n$, $Q_n = (1/n) \sum_{i=1}^{n} \delta_{Z_i}$, is the corresponding empirical measure. Assume that $\mathcal{F}$ is a class of bounded functions on S such that $\sup_{f \in \mathcal{F}} |Qf| < \infty$. The *empirical process* $\{\nu_n^Q f : f \in \mathcal{F}\} = \{\sqrt{n}(Q_n f - Qf) : f \in \mathcal{F}\}$ has its sample paths in $l^{\infty}(\mathcal{F})$, the space of real bounded functions on $\mathcal{F}$; we consider on it the supremum norm. The definitions of Vapnik-Červonenkis classes of functions and euclidean classes of functions $\mathcal{F}$ can be found for instance in Dudley (1984) and Pollard (1984).

We will also need the theory and results on U-processes as it appears in Arcones and Giné (1994). Let m be a positive integer and let $k(z_1, \ldots, z_m)$ be a function symmetric in its arguments. The *U-statistic of order m with kernel k based on Q* is defined as

$$U_n^m(k,Q) = \frac{1}{\binom{n}{m}} \sum_{(i_1,\ldots,i_m)\in I_n^m} k(Z_{i_1},\ldots,Z_{i_m}),$$

where $I_n^m = \{A \in 2^{\{1,\ldots,n\}} : \#A = m\}$. Let $\mathcal{K}$ be a class of measurable functions with m variables, symmetric in their arguments. We define the *U-process of order m based on Q with kernel in the class $\mathcal{K}$* as

$$\Lambda_n^m(\mathcal{K},Q) = \{\sqrt{n}(U_n^m(k,Q) - Q^m k) : k \in \mathcal{K}\},$$

where $Q^m k = \int k dQ^m$ and Q^m is the product measure $Q \otimes \overset{(m)}{\cdots} \otimes Q$. If $m = 1$, U-processes are empirical processes.

A U-process Λ_n^m satisfies the Central Limit Theorem if there exists a gaussian process $\{G(k) : k \in \mathcal{K}\}$, with a version of bounded sample paths which are uniformly continuous for the pseudodistance d defined by $d^2(k_1,k_2) = \mathrm{Var}(Q^{m-1}(k_1 - k_2))$, such that

$$\sqrt{n}(U_n^m(k,Q) - Q^m k) \longrightarrow_w G(k) \ \text{ in } \ l^\infty(\mathcal{K}),$$

in the sense of Hoffmann-Jørgensen (1984). G is a centered gaussian process indexed by $\mathcal{K}$ with covariance function

$$E(G(k)G(h)) = m^2 Q[(Q^{m-1}k)(Q^{m-1}h)] - m^2(Q^m k)(Q^m h).$$

A U-process is degenerated if $Q^{m-1}k = 0$ for all $k \in \mathcal{K}$.

2 Tests based on U-processes

In this section we propose a constant coefficient test based on minimum distance and U-processes. First, we present a U-process related to the discrepancies between the observations of variables following a random coefficient regression model and the theoretical distribution of the variables under a model with constant coefficients. Consider model (1.1) and assume that the null hypothesis in (1.2) holds. If distributions F_X and F_A, and value b were known, a natural way of testing the constant coefficient hypothesis is by using some distance between the empirical distribution of the pairs (Y_i, X_i) and the theoretical distribution of (Y, X). This distance can be based on the empirical process

$$\sqrt{n}(P_n - \mathcal{P}(F_A \otimes \delta_b, F_X)), \tag{2.4}$$

indexed by the semiintervals of $\mathbb{R}^{p+pq}$.

The joint distribution of X and A is $F_X \otimes F_A$. If these distribution functions (or, equivalently, F_{YX} and b) were known, the test could be built by using the empirical process

$$\sqrt{n}(Q_n^b - F_X \otimes F_A) = \sqrt{n}(Q_n^b - F_X \otimes F_{Y-Xb}), \tag{2.5}$$

where $Q_n^b = \frac{1}{n}\sum_{i=1}^n \delta_{(X_i, A_i^b)}$, indexed by the semiintervals of $\mathbb{R}^{pq+p}$. If b is known, then pairs (X_i, A_i) are observable and it is possible to construct the empirical distributions Q_n^b, $F_{X,n} = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, and $F_{A,n}^b = \frac{1}{n}\sum_{i=1}^n \delta_{A_i^b}$, where $A_i^b = Y_i - X_i b$. If b is unknown but we know the populational distributions, (2.4) allows us to test the coefficients constancy. For each fixed b, the sup norm of the process (2.4) is a measure of the discrepancy between the observed empirical distribution and the theoretical one corresponding to b. The value b can be estimated by minimum distance (see, e.g., Pollard (1980)) as the quantity minimizing the sup norm of (2.4) and then use the minimum norm as the statistic for a goodness-of-fit test of the observations to model (1.1) with constant coefficient for X. An analogous strategy could be implemented from (2.5) but it is not so straightforward to frame it in the work of Pollard (1980) because in (2.5) we cannot separate an observed part, independent of the parameter b, and another one representing the different population distributions corresponding to each value b. Since all the population distributions in (2.4) and (2.5) are unknown, even the alternative of minimum distance estimation in (2.4) is not directly implementable. However, these ideas still allow us to define useful statistics to carry out our tests, just by estimating the unknown elements in (2.4) and (2.5). Consider the processes

$$\sqrt{n}(P_n - \mathcal{P}(F_{A,n}^b \otimes \delta_b, F_{X,n})), \tag{2.6}$$

and

$$\sqrt{n}(Q_n^b - F_{X,n} \otimes F_{A,n}^b), \tag{2.7}$$

indexed by the same semiintervals in $\mathbb{R}^{p+pq}$.

The following lemmas help to describe the asymptotic behavior of these statistics. The first one establishes that the processes in (2.6) and (2.7) are pointwise equal to U-processes defined on particular classes of functions. To simplify notation, we will write F instead of F_{YX}, and τ_n will be the sequence $\binom{n}{2}/n^2$.

Lemma 2.1 *Let*

$$\mathcal{L} = \{l_{bst} : b \in \mathbb{R}^q, s \in \mathbb{R}^p, t \in \mathbb{R}^{pq}\}$$

and

$$\mathcal{K} = \{k_{btv} : b \in \mathbb{R}^q, t \in \mathbb{R}^{pq}, v \in \mathbb{R}^p\}$$

be classes of real functions defined on $\mathbb{R}^{p+pq} \otimes \mathbb{R}^{p+pq}$ *(or* $\mathbb{R}^{pq+p} \otimes \mathbb{R}^{pq+p}$*) where*

$$\begin{aligned} l_{bst}((y,x),(\gamma,\alpha)) = {} & I_{(-\infty,s]}(y) I_{(-\infty,t]}(x) + I_{(-\infty,s]}(\gamma) I_{(-\infty,t]}(\alpha) - \\ & - I_{(-\infty,s]}(\gamma + (x-\alpha)b) I_{(-\infty,t]}(x) - I_{(-\infty,s]}(y + (\alpha - x)b) I_{(-\infty,t]}(\alpha), \end{aligned} \tag{2.8}$$

and

$$k_{btv}((y,x),(\gamma,\alpha)) = (I_{(-\infty,t]}(x) - I_{(-\infty,t]}(\alpha))(I_{(-\infty,v]}(y - xb) - I_{(-\infty,v]}(\gamma - b\alpha)), \tag{2.9}$$

and let U_n^1 *and* U_n^2 *be* U*-statistics of order 2 based on* F*. Then for all* $b \in \mathbb{R}^q, s \in \mathbb{R}^p, t \in \mathbb{R}^{pq}$ *and* $v \in \mathbb{R}^p$*, it holds that*

$$\begin{aligned} \sqrt{n}(P_n - \mathcal{P}(F_{A,n}^b \otimes \delta_b, F_{X,n}))(s,t) &= \sqrt{n}\tau_n U_n^1(l_{bst}, F), \quad \textit{and} \\ \sqrt{n}(Q_n^b - F_{X,n} \otimes F_{A,n}^b)(t,v) &= \sqrt{n}\tau_n U_n^2(k_{btv}, F). \end{aligned}$$

To construct U-processes from U_n^1 and U_n^2 it is necessary to know $F \otimes F(l_{bst})$ and $F \otimes F(k_{btv})$; we will calculate these expectations under the null hypothesis of constant coefficients, which we will keep through the rest of the paper. The following is a technical lemma that we will use later.

Lemma 2.2 *Let $(Y_i, X_i), i \geq 1$ be independent and identically distributed random variables with distribution $F = \mathcal{P}(F_A \otimes \delta_{b_0}, F_X)$. Then:*

(i) For all $b \in \mathbb{R}^q, s \in \mathbb{R}^p, t \in \mathbb{R}^{pq}$,

$$F \otimes F(l_{bst}) = 2\{P(Y_1 \leq s, X_1 \leq t) - P(X_2 b + (Y_1 - X_1 b) \leq s, X_2 \leq t)\}.$$

(ii) For all $b \in \mathbb{R}^q, t \in \mathbb{R}^{pq}, v \in \mathbb{R}^p$,

$$F \otimes F(k_{btv}) = 2\{P(X_1 \leq t, Y_1 - X_1 b \leq v) - P(X_1 \leq t)P(Y_1 - X_1 b \leq v)\}.$$

(iii) If $b = b_0$ then $F \otimes F(l_{bst}) = 0$, for all $s \in \mathbb{R}^p, t \in \mathbb{R}^{pq}$.

(iv) Let $q = 1$ and $V(A) < \infty$, $0 \neq V(X) < \infty$. If $F \otimes F(l_{bst}) = 0$ for all $s, t \in \mathbb{R}^p$, then $b = b_0$.

(v) Let $X^{(j)}$ be one component of X. If the distribution of $X^{(j)}\beta$ is non-degenerated for all $\beta \in \mathbb{R}^q$, $\beta \neq 0$, then

$$F \otimes F(k_{btv}) = 0, \text{ for all } t \in \mathbb{R}^{pq}, v \in \mathbb{R}^p \text{ if, and only if, } b = b_0.$$

Remark 1. Part *(iv)* in Lemma 2.2 does not hold for $q > 1$; a proof of this can be seen in the Appendix.

The following result establishes a key property of the classes of functions in Lemma 2.1.

Proposition 2.1 *$\mathcal{L}$ and $\mathcal{K}$ are euclidean classes of functions.*

Now, Theorem 4.8 in Arcones and Giné (1993) allows us to obtain the asymptotic behavior of the U-processes defined from the U-statistics U_n^1 and U_n^2 introduced in Lemma 2.1.

Theorem 2.1 *The U-processes*

$$\Lambda_n(\mathcal{L}, F) = \{\sqrt{n}(U_n^1(l_{bst}, F) - F \otimes F(l_{bst})) : l_{bst} \in \mathcal{L}\}$$

and

$$\Pi_n(\mathcal{K}, F) = \{\sqrt{n}(U_n^2(k_{btv}, F) - F \otimes F(k_{btv})) : k_{btv} \in \mathcal{K}\}$$

satisfy the Central Limit Theorem with gaussian limit processes G_1 and G_2, respectively.

The following Corollary shows that both the Kolmogorov-Smirnov and the Cramér-von Mises statistics related with the metric used in the minimum distance estimation, converge weakly to a non-degenerated random variable.

Corollary 2.1 *It holds that*

$$\delta_n^{(1)} = \inf_{b\in\mathbb{R}^q} \sup_{s,t} |\Lambda_n(l_{bst})| \longrightarrow_w \inf_b \sup_{s,t} |G_1(l_{bst})|$$

and

$$\delta_n^{(2)} = \inf_{b\in\mathbb{R}^q} \sup_{t,v} |\Pi_n(k_{btv})| \longrightarrow_w \inf_b \sup_{t,v} |G_2(k_{btv})|.$$

Moreover, if Q is a finite measure on $\mathbb{R}^{p+pq} \times \mathbb{R}^{p+pq}$ then

$$\gamma_n^{(1)} = \inf_{b\in\mathbb{R}^q} \left(\int \Lambda_n(l_{bst})^2 dQ \right)^{\frac{1}{2}} \longrightarrow_w \inf_b \left(\int G_1(l_{bst})^2 dQ \right)^{\frac{1}{2}}$$

and

$$\gamma_n^{(2)} = \inf_{b\in\mathbb{R}^q} \left(\int \Pi_n(k_{btv})^2 dQ \right)^{\frac{1}{2}} \longrightarrow_w \inf_b \left(\int G_2(k_{btv})^2 dQ \right)^{\frac{1}{2}}.$$

In the sequel we will study only the Kolmogorov-Smirnov type statistics. All the proposals for them can be straightforwardly extended to the Cramér-von Mises ones, $\gamma_n^{(1)}$ and $\gamma_n^{(2)}$.

The statistics $\delta_n^{(1)}$ and $\delta_n^{(2)}$ cannot be used in practice because F is unknown. The next result gives the asymptotic behavior of the corresponding statistic built by replacing F by its empirical version. An analogous result holds for Π_n.

Corollary 2.2 *Let the gaussian process G_3 be the weak limit in $l^\infty(\mathcal{L})$ of the empirical process $\{\sqrt{n}\,(F_n \otimes F_n(l_{bst}) - F \otimes F(l_{bst})) : l_{bst} \in \mathcal{L}\}$. Then:*

(i) The process $\hat{\Lambda}_n(\mathcal{L}, F) = \{\sqrt{n}\,(U_n^1(l_{bst}, F) - F_n \otimes F_n(l_{bst})) : l_{bst} \in \mathcal{L}\}$ converges weakly in $l^\infty(\mathcal{L})$ to $G_1 - G_3$.

(ii)

$$\hat{\delta}_n^{(1)} = \inf_{b\in\mathbb{R}^q} \sup_{s,t} |\hat{\Lambda}_n(l_{bst})| \longrightarrow_w \inf_b \sup_{s,t} |G_1(l_{bst}) - G_3(l_{bst})|.$$

In the following Corollary we study some properties of another natural distance to test the null hypothesis of constant coefficients.

Corollary 2.3 *Let*

$$d_n^{(1)} = \inf_b \sup_{s,t} |\sqrt{n} U_n^{(1)}(l_{bst})|$$

and

$$d_n^{(2)} = \inf_b \sup_{t,v} |\sqrt{n} U_n^{(2)}(k_{btv})|.$$

The sequences of random variables $d_n^{(1)}$ and $d_n^{(2)}$ are stochastically bounded.

Next, we present an algorithm for the bootstrap implementation of these tests. We will do it for the distance $d_n^{(2)}$. Calculating $d_n^{(2)}$ is faster than $d_n^{(1)}$ because the maximum must be calculated over a grid of n^2 points instead of the n^3 points for $d_n^{(1)}$.

Algorithm 2.1

1. *Calculate the distance* $d_n = d_n^{(2)} = \inf_b \sup_{t,v} |\sqrt{n} U_n^{(2)}(k_{btv})|$.

2. *Calculate a consistent estimate of* $b = E[B_i]$ *which will be denoted by* $\hat{b}_n$ *(the least squares estimate, the generalized least squares estimate, or any other) and construct the empirical distribution of the estimated* A_i *coefficients,* $\hat{F}_{A,n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\{Y_i - X_i \hat{b}_n\}}$.

3. *Obtain the bootstrap sample* $(Y_i^*, X_i^*), i = 1, \ldots, n,$ *where* $Y_i^* = A_i^* + X_i^* \hat{b}_n$ *and* (X_i^*, A_i^*) *are independent observations of a random variable with distribution* $F_{X,n} \otimes \hat{F}_{A,n}$.

4. *Repeat step* 1 *with the bootstrap sample* $(Y_i^*, X_i^*), i = 1, \ldots, n$ *and let* d_n^* *be the minimum distance.*

5. *Repeat steps* 3 *and* 4 B *times to obtain* B *bootstrap observations of* d_n^**:* $d_n^{*(j)}, j = 1, \ldots, B$.

6. *Compare* d_n *with the* $\alpha - th$ *upper quantile of the empirical distribution of* $d_n^{*(j)}$, $j = 1, \ldots, B$, *and reject* H_0 *if* d_n *is larger than this quantile.*

This algorithm can be modified to avoid the optimization in steps 1 and 4 diminishing the computational burden; to this end, we may approximate the minimum distance by the distance corresponding to the estimates $\hat{b}_n$ and $\hat{b}_n^*$ used in steps 2 and 4.

2.1 A test based on U-processes with coefficients prediction

We present now a modification of the previous test. In Algorithm 2.1, the resampling was based on the empirical distribution of X_i and the empirical distribution $\hat{F}_{A,n}$ of A_i (which we will call residuals) estimated by means of an estimate of b. Under H_0, $\hat{F}_{A,n}$ converges uniformly to F_A (see, e.g., Shorack and Wellner (1986), page 194). However, under H_1 each estimated residual is not an estimation of A_i, but of $A_i + X_i(B_i - E[B_i])$; it follows that this empirical distribution of the residuals won't approach F_A: if we assume that A and B are independent, that there exists a limit for this empirical distribution and that the second moments converge, then the limit variance will be $V(A+X(B-E[B])) = V(A)+E[X^2]V(B) > V(A)$. So, under H_1 the empirical distribution of the residuals comes from a variable with a larger variance than A. It is convenient to have an estimate of F_A which is appropriate both under H_0 and H_1. Let $\hat{b}_n$ an estimate of $b = E[B]$ and let $A_{in} = Y_i - X_i \hat{b}_n$. The differences between the empirical distribution of (X_i, A_{in}) and the product of the empiricals under the alternative hypothesis will be large not only due to the dependence between residuals and X_i, but also due to the different dispersion of the residuals.

Griffiths (1972) proposes estimates (or predictors) $\hat{B}_{in}$ of the values taken by the coefficients in any of the observed individuals, and he studies their properties assuming that the variance of (A_i, B_i) is a known diagonal matrix. If this matrix is unknown, it is possible to estimate it consistently (see, e.g., Hildreth and Houck (1968), Amemiya (1984) or Judge, Griffiths, Hill, Lütkepohl, and Lee (1985)). From the estimation of the coefficient values, we can obtain an estimation of the residuals: $\hat{A}_{in} = Y_i - X_i \hat{B}_{in}$. Under the null hypothesis of constant coefficients and when the covariance matrix is known, these estimates coincide with least squares estimates. Thus, under H_0 and with estimated covariance matrix, they are asymptotically equivalent to the least squares estimate and the empirical distribution of the corresponding residuals converges to F_A. Under the alternative hypothesis, the variance of the residuals $\hat{A}_{in}$ is smaller than that of the least squares residuals and, as a consequence, the distance used for the test will be larger; this improves the test power. The procedure is as follows. First, obtain the estimates of the coefficient actual values $\hat{B}_n = (\hat{B}_{1n}, \ldots, \hat{B}_{nn})'$ and the residuals $\hat{A}_{in} = Y_i - \hat{B}_{in}, i = 1, \ldots, n$. Let $\check{F}_{A,n}$ be the empirical distribution of these estimated residuals. Calculate the distance $\check{d}_n$ between the empirical distribution of the pairs (X_i, A_{in}) and the product distribution of the empiricals $F_{X,n}$ and $\check{F}_{A,n}$. The following resampling algorithm provides a way to obtain the critical point to be compared with $\check{d}_n$.

Algorithm 2.2

1. *Let $\hat{b}_n$ be an estimate of b. Calculate the distance $\check{d}_n$ between the empirical distribution of the pairs (X_i, A_{in}) and the product distribution of the empiricals $F_{X,n}$ and $\check{F}_{A,n}$, where $A_{in} = Y_i - \hat{b}_n X_i$, and $\check{F}_{A,n}$ is the empirical distribution of the estimations of A_i made from the predictions of B_i given by the method in Griffiths (1972).*

2. *Obtain the bootstrap sample $(Y_i^*, X_i^*), i = 1, \ldots, n$, where $Y_i^* = \hat{A}_{in}^* + X_i^* \hat{b}_n$ and $(X_i^*, \hat{A}_{in}^*)$ are independent observations of a random variable with distribution $F_{X,n} \otimes \check{F}_{A,n}$.*

3. *Calculate $\hat{b}_n^*$, the bootstrap version of $\hat{b}_n$ and $A_{in}^* = Y_i^* - \hat{b}_n^* X_i^*$. Calculate the distance $\check{d}_n^*$ between the empirical distribution of the pairs (X_i^*, A_{in}^*) and the product distribution of the empiricals $F_{X,n}^*$ and $\check{F}_{A,n}^*$, the later constructed following Griffiths (1972).*

4. *Repeat steps 2 and 3 B times to get $\check{d}_n^*$: $\check{d}_n^{*(j)}, j = 1, \ldots, B$.*

5. *Compare $\check{d}_n$ with the α-th upper quantile of the empirical distribution of $\check{d}_n^{*(j)}, j = 1, \ldots, B$, and reject H_0 if $\check{d}_n$ is larger than this quantile.*

We can rewrite Algorithm 2.2 in a way such that the test statistic is not the distance between two distributions for a fixed value of b (the estimate $\hat{b}_n$) but the infimum in b of the distances associated to each value of b. To this end, it is necessary to redefine the estimation of the predictions given by Griffiths (1972) in such a way that allows to fix a value b for the expectation of the coefficient of X.

3 Alternative tests

The ideas underlying the previous tests can be used to define two new tests: one of them based on the Kolmogorov-Smirnov two-sample test and the other one on Sukhatme's two-sample equal dispersion test.

In the first test presented in section 2 we considered the distance between the empirical distribution of the pairs (X_i, A_{in}) and the product distribution of the empiricals of X_i and A_{in}, respectively. Residuals A_{in} were obtained from the observations (Y_i, X_i) and an estimate of b. We have also argued that, under the alternative, the residuals distribution could be better approached when using the estimates of the actual values of the coefficients $\hat{B}_{in}$; because of that, in Section 2.1 we used as a statistic for the test the distance between the empirical of (X_i, A_{in}) and the product of the empiricals of X_i and $\hat{A}_{in}$.

Under H_0, the product distributions in these two tests are close because the estimates of the coefficient actual values and the estimate of b are close; however, these distributions can be very different under the alternative. Since the empirical distribution of X_i is common to both product distributions, the part responsible for the difference is the empirical of the residuals. Thus, the test of H_0 could be based on the differences between $\hat{F}_{A,n}$ and $\check{F}_{A,n}$; this allows to use the well-known methods to compare two distributions, but the information contained in the dependence structure of X_i and A_i is lost. The two-sample Kolmogorov-Smirnov statistics can be employed to test the equality of $\hat{F}_{A,n}$ and $\check{F}_{A,n}$. The coefficient variance estimates proposed by Hildreth and Houck (1968) are consistent. From this and the relationship between the ordinary residuals and the residuals proposed by Griffiths (1972),

$$\hat{A}_{in} - \hat{a} = \frac{\sigma_A^2}{\sigma_A^2 + \sum_{k=1}^{q} X_{ik}^2 \sigma_{B_k}^2} (A_{in} - \hat{a}), \qquad (3.10)$$

we can conclude that, under H_0, both $\hat{F}_{A,n}$ and $\check{F}_{A,n}$ are asymptotically the same if they are calculated from the same data. The empirical distribution of usual residuals converges in the sup norm to the true residuals distribution (see, e.g., Shorack and Wellner (1986), page 194). Thus, $\check{F}_{A,n}$ converges to the same distribution in the sup norm. This justifies the use of the two-sample Kolmogorov-Smirnov test to compare $\hat{F}_{A,n}$ and $\check{F}_{A,n}$.

This test could present an empirical level which is different to the nominal one. There exist two reasons for this. First, under H_0 and for small samples, the residuals estimated following Griffiths (1972) are, in general, smaller than the ordinary residuals. Then, the dispersion of the empirical distribution of the former ones is smaller than that of the latter ones. And second, the two distributions compared by the test are coming from dependent samples. Thus the asymptotic distribution of the proposed statistic under the null hypothesis differs from its asymptotic distribution under the usual conditions they were designed for. We may use a resampling technique to overcome this problem; then the following algorithm summarizes our proposals.

Algorithm 3.1

1. *Using the residuals obtained by generalized least squares, estimate the residuals distribution by* $\hat{F}_{A,n}$.

2. *Using the residuals obtained following Griffiths (1972), estimate the residuals distribution by* $\check{F}_{A,n}$.

3. *Apply the two-sample Kolmogorov-Smirnov test to the distributions* $\hat{F}_{A,n}$ *and* $\check{F}_{A,n}$. *Let* K_n *be the value of this statistic.*

4. *Construct an estimate* $\hat{b}_n$ *of* b *and estimate the residuals distribution by* $\check{F}_{A,n}$, *built from the residuals obtained following Griffiths (1972).*

5. *Obtain the bootstrap sample* $(Y_i^*, X_i^*), i = 1, \ldots, n$, *where* $Y_i^* = \hat{A}_{in}^* + X_i^* \hat{b}_n$ *and* $(X_i^*, \hat{A}_{in}^*)$ *are independent observations of a random variable with distribution* $F_{X,n} \otimes \check{F}_{A,n}$.

6. *Repeat steps* 1, 2 *and* 4 *with the bootstrap sample obtained in* 5 *to get* K_n^*.

7. *Repeat steps* 5 *and* 6 B *times to get* B *bootstrap observations of* K_n: $K_n^{*(j)}, j = 1, \ldots, B$.

8. *Compare* K_n *with the upper* α*-th quantile of the empirical distribution of* $K_n^{*(j)}, j = 1, \ldots, B$, *and reject* H_0 *if* K_n *is larger than this quantile.*

Finally, we propose another way of testing for constant coefficients. Following Griffiths (1972), the predictions of the residuals are given by (3.10). Thus, each residual prediction is the ordinary residuals multiplied by a quantity smaller than one, which is different for each observation. So, the main difference between $\hat{F}_{A,n}$ (adequate estimate of F_A under H_0) and $\check{F}_{A,n}$ (reasonable both under H_0 and H_1) relies on the dispersion of the distributions. This suggests to use a test for equal variances like Sukhatme's (see, e.g., Gibbons (1985), page 186). The only change needed in Algorithm 3.1 is to apply Sukhatme's test in step 3 instead of Kolmogorov-Smirnov's.

4 Simulation study

In this section we report the results of a simulation study carried out to compare different ways of testing the null hypothesis of constant coefficients in the random coefficient regression model. Data were generated using 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 $Y_i = A_i + X_i B_i, i = 1, \ldots, n$. The value b_0 is always equal to 1.

We label *normal* (or *N*) 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 σ_B. The collection of simulations labeled *Cauchy* (or C) is constructed from A with Cauchy distribution with zero median and interquantile semirange s_A 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. In our simulation study, we have considered each of this two situations with two sample sizes ($n = 50$ and $n = 100$) and two distributions for X ($N(0,1)$ and $N(2,1)$). The dispersion parameter of B, either σ_B or s_B (depending on the distribution of B) belongs to the set $\{0, 0.2, 0.4, 0.6, 0.8, 1, 1.3, 1.6, 2\}$ if $n = 50$ and to the set $\{0, 0.2, 0.4, 0.8, 1.3, 2\}$ if $n = 100$. For each of the possible combinations of these parameters, 500 samples were obtained. Also, when needed, 500 resamples were generated in each case. Two criteria were used to asses the results obtained: test empirical levels and estimated power functions.

We will compare five tests labelled Ci, $i = 1, \ldots, 5$. Tests $C1$ to $C4$ have been introduced in previous sections and $C5$ is the test in Koenker (1981). This test is preferred to the one proposed by Breusch and Pagan (1979) and it is practically equivalent to that of White (1980) for simple regression. Koenker's test is less sensitive than Breusch and Pagan's to nonnormality in the residuals; under normality, both are equivalent. On the other hand, the differences in calculating the statistics in Koenker's and White's tests are small and not relevant in the case we are considering: simple regression models where heteroscedasticity is due to random coefficients. Tests $Ci, i = 1, \ldots, 5$, have been applied as follows:

$C1$: Described in Algorithm 2.1, but instead of minimizing in b, we use an estimate $\hat{b}_n$ of location of B.

$C2$: As in $C1$, but using Algorithm 2.2. The coefficient values have been predicted as in Griffiths (1972). His method is adequate for normal (A, B); for Cauchy coefficients, we have predicted them as follows: in model (1.1) with $p = q = 1$ and Cauchy coefficients, we have that

$$S(Y - (m_A + Xm_B) \mid X = x) = S(A - m_A + x(B - m_B)) = s_A + |x| s_B,$$

where m_V is the median of a variable V and $S(V)$ (or s_V) is its interquartile range. From this,

$$\begin{aligned} S(A - m_A) &= \frac{s_A}{s_A + |x| s_B} S(Y - (m_A + x m_B)), \\ S(x(B - m_B)) &= |x| S(B - m_B) = \frac{|x| s_B}{s_A + |x| s_B} S(Y - (m_A + x m_B)), \\ S(B - m_B) &= \frac{s_B}{s_A + |x| s_B} S(Y - (m_A + x m_B)). \end{aligned}$$

In Griffiths (1972), the residuals estimated by generalized mean squares are assigned to the terms $x_{ik} B_{ik}$ proportionally to the variance of the error $A + x_i'(B - E(B))$ as it decomposes among the terms $x_{ik} V(B_{ik})$. Following this idea, the generalized minimum absolute deviations residuals $\hat{u}_i$ will be now distributed as:

$$\begin{aligned} \hat{A}_i &= \frac{\hat{s}_A}{\hat{s}_A + |x_i| \hat{s}_B} \hat{u}_i, \\ \hat{B}_i &= \frac{\hat{s}_B}{\hat{s}_A + |x_i| \hat{s}_B} \hat{u}_i, \end{aligned}$$

n	(A,B)	$E(X)$	C1	C2	C3	C4	C5
50	N	0	*.000*	*.050*	.014	.012	.014
			.002	*.076*	.048	.060	.062
			.006	.112	.078	.098	.106
		2	*.000*	*.068*	*.054*	*.030*	.004
			.004	*.126*	*.114*	*.094*	.050
			.008	*.180*	*.154*	*.146*	.124
	C	0	*.180*	.002	*.052*	.018	*.032*
			.414	*.012*	*.118*	.068	.060
			.568	*.044*	*.170*	*.136*	*.068*
		2	*.000*	.006	*.048*	.002	.018
			.004	*.018*	*.072*	*.014*	.052
			.010	*.048*	.086	*.026*	.084
100	N	0	—	*.032*	.018	.012	.008
			—	*.080*	.038	.048	.036
			—	.124	.074	.104	.082
		2	—	*.048*	*.060*	.018	.008
			—	*.102*	*.116*	*.080*	.050
			—	*.170*	*.156*	*.142*	.106
	C	0	—	.002	*.030*	*.024*	*.036*
			—	.038	*.108*	.066	.048
			—	.076	*.152*	*.140*	.076
		2	—	.004	.010	.008	*.028*
			—	*.014*	*.026*	*.016*	.060
			—	*.042*	*.034*	*.028*	.082

Table 1: Empirical levels for the five constant coefficients tests. Each cell shows empirical levels for theoretical levels .01, .05 and .1. Significant differences from nominal levels are indicated by italic types.

where $\hat{s}_A$ and $\hat{s}_B$ are respectively the estimates of s_A and s_B, introduced in Delicado and Romo (1997).

$C3$: The test proposed in Algorithm 3.1, using the two-sample Kolmogorov-Smirnov test. The coefficients prediction are as in $C2$.

$C4$: Analogous to $C3$, but using Sukhatme's test in Algorithm 3.1.

$C5$: Koenker's test.

Table 1 shows the empirical sizes for each of the eight models generated under the null hypothesis of dispersion of B equal to zero. It is clear the poor behavior of $C1$; we decided not to include this test for $n = 100$. The empirical levels for $C5$ are significantly different from the theoretical ones at level 95% in 4 out of 24 cases. This occurs more often in the remaining tests procedures. $C3$ and $C4$ provide also good results in the normal case, mainly if X is centered at 0. $C5$ behaves well for any combination of models for (A,B) and distributions for X. For $n = 100$, Cauchy distribution for A and $E(X) = 0$, $C2$ outperforms $C5$.

Figures 1 and 2 give the power functions for the theoretical level $\alpha = 0.05$ and sample sizes $n = 50$ and $n = 100$, respectively. Parts (a) and (c) in both figures correspond to X with zero mean; $E(X)=2$ in (b) and (d). Coefficients (A,B) are normal in (a) and (b), and Cauchy in (c) and (d). Results from $C1$ are not shown due to its poor behavior advanced in the comments about Table 1.

Observe that the four tests we have considered ($C2$, $C3$, $C4$ and $C5$) present empirical levels not too far from the nominal level, though some differences are statistically significant (see Table 1). For residuals A with normal distribution and for both $n=50$ (Figure 1 (a) and (b)) and $n=100$ (Figure 2 (a) and (b)), the results are highly satisfactory for these four tests. Tests $C2$ and $C5$ present very low power when (A,B) is Cauchy (the estimated powers essentially coincide with the nominal level). On the other hand, tests $C3$ and $C4$ performs well also under Cauchy models. In general $C3$ is more successful than $C4$ in power terms, although $C4$ fulfills better the empirical levels requirements.

As a conclusion we can say that the four tests we consider have a similar global behavior (i.e., considering jointly empirical level and power) under normal assumptions, and that $C3$ and $C4$ are preferred when the normality assumption does not hold. Regarding only performances under H_0, Koenker test ($C5$) seems to be the test of choice.

APPENDIX: Proofs

Proof of Lemma 2.1. Given $b\in I\!R^q$, $s\in I\!R^p$ and $t\in I\!R^{pq}$,

$$\sqrt{n}(P_n-\mathcal{P}(F^b_{A,n}\otimes\delta_b,F_{X,n}))(s,t)=$$

$$=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n I_{(-\infty,s]}(Y_i)I_{(-\infty,t]}(X_i)-\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n I_{(-\infty,s]}(X_ib+A^b_j)I_{(-\infty,t]}(X_i)\right)=$$

$$=\sqrt{n}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n(I_{(-\infty,s]}(Y_i)I_{(-\infty,t]}(X_i)-I_{(-\infty,s]}(Y_j+(X_i-X_j)b)I_{(-\infty,t]}(X_i))\right)=$$

$$=\sqrt{n}\left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j>i}^n\Big(I_{(-\infty,s]}(Y_i)I_{(-\infty,t]}(X_i)+I_{(-\infty,s]}(Y_j)I_{(-\infty,t]}(X_j)-\right.$$

$$\left.-I_{(-\infty,s]}(Y_j+(X_i-X_j)b)I_{(-\infty,t]}(X_i)-I_{(-\infty,s]}(Y_i+(X_j-X_i)b)I_{(-\infty,t]}(X_j)\Big)\right)=$$

$$=\sqrt{n}\frac{1}{n^2}\sum_{i=1}^n\sum_{j>i}^n l_{bst}((Y_i,X_i),(Y_j,X_j))=$$

$$=\frac{\binom{n}{2}}{n^2}\sqrt{n}\left(\frac{1}{\binom{n}{2}}\sum_{i=1}^n\sum_{j>i}^n l_{bst}((Y_i,X_i),(Y_j,X_j)))\right)=\sqrt{n}\tau_nU^1_n(l_{bst}),$$

where $l_{bst}((y,x),(\gamma,\alpha))$ is defined in (2.8). This proves the first part. Now, given $b\in I\!R^q$, $t\in I\!R^{pq}$ and $v\in I\!R^p$,

$$\sqrt{n}(Q^b_n-F_{X,n}\otimes F^b_{A,n})(t,v)=$$

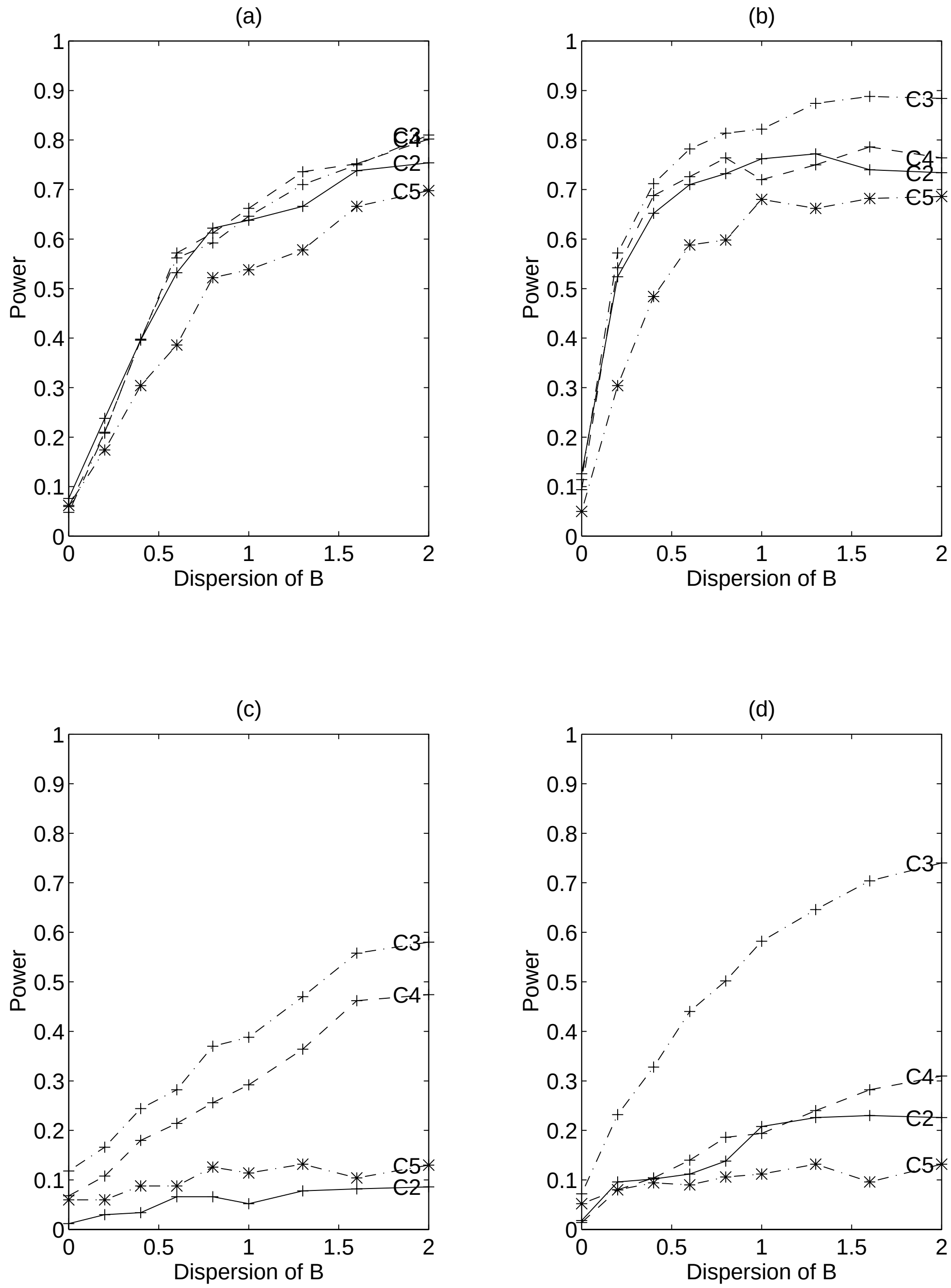

Figure 1: Power functions for the constant coefficients test, $n = 50$.
In (a) and (b) (A, B) is normal; in (c) and (d) they are Cauchy. In (a) and (c) $X \sim N(0,1)$, and in (b) and (d), $X \sim N(2,1)$.
$\text{—+—}\ C_2\ -\cdot+\cdot-\ C_3,\ --+--\ C_4,\ -\cdot*\cdot-\ C_5.$

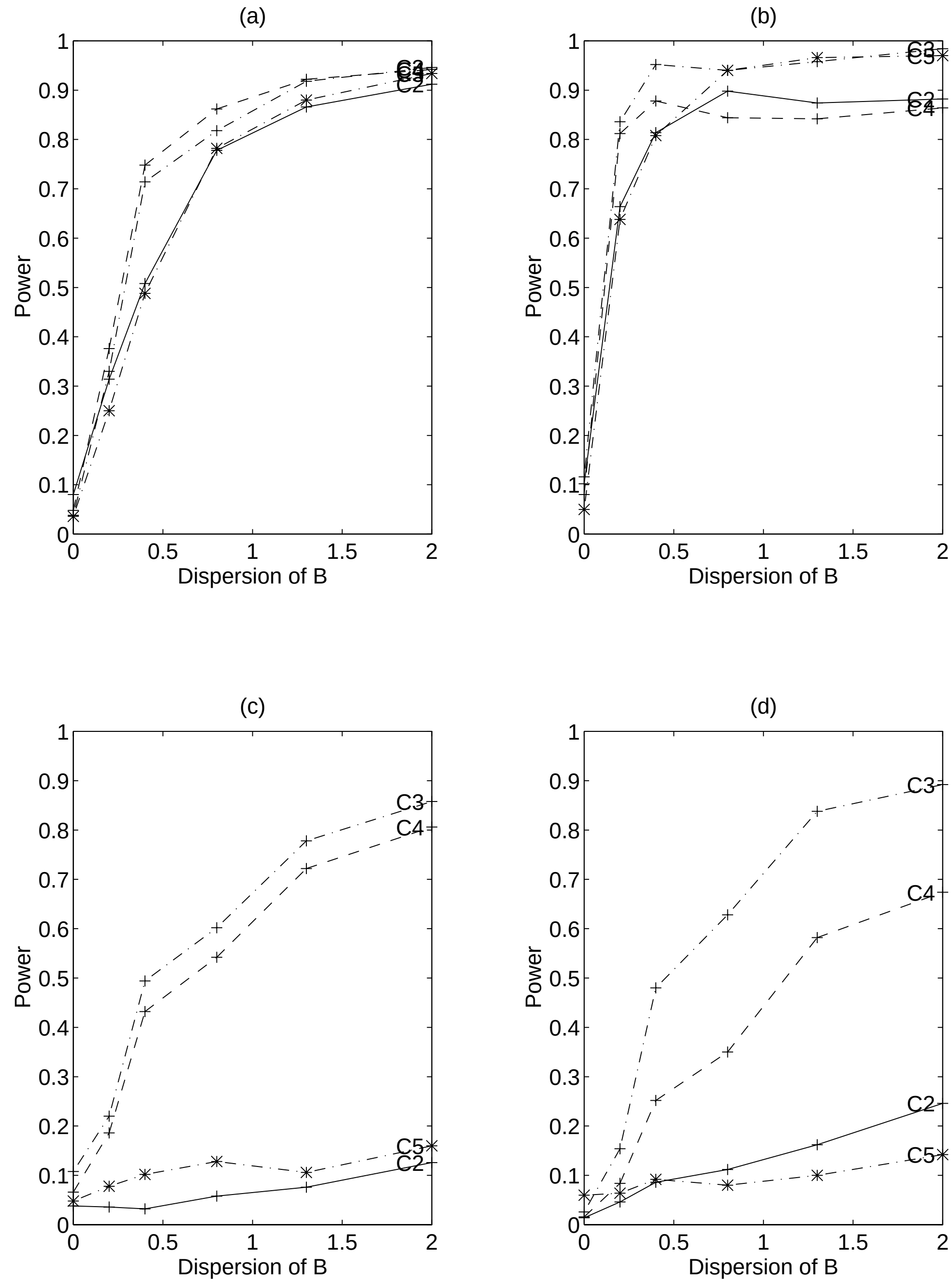

Figure 2: Power functions for the constant coefficients test, $n = 100$.
In (a) and (b) (A, B) *is normal; in (c) and (d) they are Cauchy. In (a) and (c)* $X \sim N(0,1)$, *and in (b) and (d),* $X \sim N(2,1)$.
—+— C_2 $-\cdot + \cdot -$ C_3, $- - + - -$ C_4, $- \cdot * \cdot -$ C_5.

$$= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,t]}(X_i) I_{(-\infty,v]}(A_i^b) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{(-\infty,t]}(X_i) I_{(-\infty,v]}(A_j^b)\right) =$$

$$= \sqrt{n}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{(-\infty,t]}(X_i)(I_{(-\infty,v]}(A_i^b) - I_{(-\infty,v]}(A_j^b))\right) =$$

$$= \sqrt{n}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j>i}^{n}(I_{(-\infty,t]}(X_i) - I_{(-\infty,t]}(X_j))(I_{(-\infty,v]}(A_i^b) - I_{(-\infty,v]}(A_j^b))\right) =$$

$$= \sqrt{n}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j>i}^{n}(I_{(-\infty,t]}(X_i) - I_{(-\infty,t]}(X_j))(I_{(-\infty,v]}(Y_i - X_i b) - I_{(-\infty,v]}(Y_j - X_j b))\right) =$$

$$= \sqrt{n}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j>i}^{n} k_{btv}((Y_i, X_i), (Y_j, X_j)),$$

$$= \frac{\binom{n}{2}}{n^2}\sqrt{n}\left(\frac{1}{\binom{n}{2}}\sum_{i=1}^{n}\sum_{j>i}^{n} k_{btv}((Y_i, X_i), (Y_j, X_j))\right) = \sqrt{n}\tau_n U_n^2(k_{btv}),$$

with $k_{btv}((y,x),(\gamma,\alpha))$ as defined in (2.9). This proves the Lemma. □

To establish Lemma 2.2 we need the following previous result.

Lemma 4.1 *Let A and X be random variables with values in $\mathbb{R}$ and $\mathbb{R}^p$, respectively. Assume that for all $\beta \in \mathbb{R}^p - \{0\}$, the random variable $X'\beta$ is non-degenerated. If the random variables A and X are independent, and also are $A + X'b$ and X, then $b = 0$.*

Proof of Lemma . By independence (of $A + X'b$ and X, and of A and X), for all $u \in \mathbb{R}$ and $v \in \mathbb{R}^p$,

$$\phi_{(A+X'b,X)}(u,v) = E\left[e^{i(u(A+X'b)+v'X)}\right] = E\left[e^{iu(A+X'b)}\right] E\left[e^{iv'X}\right] =$$

$$= E\left[e^{iuA}\right] E\left[e^{iuX'b}\right] E\left[e^{iv'X}\right]. \tag{4.11}$$

On the other hand, also by independence,

$$\phi_{(A+X'b,X)}(u,v) = E\left[e^{iuA}\right] E\left[e^{i(uX'b+v'X)}\right]. \tag{4.12}$$

From (4.11) and (4.12), it follows that for all $u \in \mathbb{R}$ and $v \in \mathbb{R}^p$,

$$\phi_{(X'b,X)}(u,v) = E\left[e^{i(uX'b+v'X)}\right] =$$

$$= E\left[e^{iu(A+X'b)}\right] E\left[e^{iv'X}\right] = \phi_{X'b}(u)\phi_X(v), \tag{4.13}$$

and thus X and $X'b$ are independent. Consider $w \in \mathbb{R}$. From (4.13),

$$\phi_{(X'b,X'b)}(u,w) = \phi_{(X'b,X)}(u,wb) = \phi_{X'b}(u)\phi_X(wb) = \phi_{X'b}(u)\phi_{X'b}(w),$$

and $X'b$ is independent of itself. Then $b=0$ because we have assumed that for any $\beta \neq 0$ the variable $X'\beta$ is nondegenerate. □

Proof of Lemma 2.2. We establish *(i) and (iii)* first. It holds that

$$F \otimes F(l_{bst}) = \int\int \Big(I_{(-\infty,s]}(y) I_{(-\infty,t]}(x) +$$

$$+ I_{(-\infty,s]}(\gamma) I_{(-\infty,t]}(\alpha) - I_{(-\infty,s]}(\gamma + (x-\alpha)b) I_{(-\infty,t]}(x) -$$

$$- I_{(-\infty,s]}(y + (\alpha - x)b) I_{(-\infty,t]}(\alpha) \Big) \, dF(\gamma,\alpha) dF(y,x) =$$

$$= P(Y \leq s, X \leq t) + P(Y \leq s, X \leq t) -$$

$$- \int \left(\int I_{(-\infty,s]}(\gamma + (x-\alpha)b) I_{(-\infty,t]}(x) dF(\gamma,\alpha) \right) dF(y,x) -$$

$$- \int \left(\int I_{(-\infty,s]}(y + (\alpha - x)b) I_{(-\infty,t]}(\alpha) dF(y,x) \right) dF(\gamma,\alpha) =$$

$$= 2\{P(Y_1 \leq s, X_1 \leq t) - P(X_2 b + (Y_1 - X_1 b) \leq s, X_2 \leq t)\}.$$

Note that $F \otimes F(l_{bst})$ can be also written as

$$2\{P(Y_1 \leq s, X_1 \leq t) - P((Y_1 - X_1 b_0) + X_2 b - X_1(b - b_0) \leq s, X_2 \leq t)\} =$$

$$= 2\{P(A_1 + X_1 b_0 \leq s, X_1 \leq t) - P(A_1 + X_2 b - X_1(b-b_0) \leq s, X_2 \leq t)\},$$

where A_1, X_1, X_2 are independent, and X_1 and X_2 have the same distribution. Obviously, if $b = b_0$ then $F \otimes F(l_{bst}) = 0$.

Let us now show *(iv)*. Under the assumptions, $F \otimes F(l_{bst}) = 0$. Thus, for all $s \in I\!R^p, t \in I\!R^p$,

$$P(A_1 + X_1 b_0 \leq s, X_1 \leq t) = P(A_1 + X_2 b - X_1(b - b_0) \leq s, X_2 \leq t);$$

this implies that for all $s \in I\!R^p$,

$$P(A_1 + X_1 b_0 \leq s) = P(A_1 + X_2 b - X_1(b - b_0) \leq s),$$

and it follows that the distribution of the first component of $A_1 + X_1 b_0$ and $A_1 + X_2 b - X_1(b - b_0)$ coincide and so they have the same variance (if $V(X^{(1)}) = 0$, we can use a different component of X with non-null variance):

$$V(A^{(1)}) + b_0^2 V(X^{(1)}) = V(A^{(1)}) + b^2 V(X^{(1)}) + (b - b_0)^2 V(X^{(1)}),$$

thus $b(b - b_0)V(X^{(1)}) = 0$; then either $b = 0$ or $b = b_0$. If $b = 0$, the hypotheses in *(iv)* imply that $P(A_1 + X_1 b_0 \leq s, X_1 \leq t) = P(A_1 + X_1 b_0 \leq s, X_2 \leq t)$, with X_1 and X_2 independent. Then

$$P(A_1 + X_1 b_0 \leq s, X_1 \leq t) =$$

$$= P(A_1 + X_1 b_0 \leq s) P(X_2 \leq t) = P(A + X b_0 \leq s) P(X \leq t),$$

and this implies independence of X and $(A + X b_0)$, which can only hold if $b_0 = 0$ because A and X are independent and the previous Lemma applies. Thus, $b = b_0$ and *(iv)* holds.

To prove *(v)*, we calculate $F \otimes F(k_{btv})$. It holds that

$$\begin{aligned} F \otimes F(k_{btv}) &= \int \Big(\int (I_{(-\infty,t]}(x) - I_{(-\infty,t]}(\alpha)) \\ & \quad (I_{(-\infty,v]}(y - xb) - I_{(-\infty,v]}(\gamma - \alpha b)) dF(\gamma,\alpha) \Big) dF(y,x) = \\ &= P(X \le t, \quad -Xb \le v) - P(X \le t) P(Y - Xb \le v) - \\ & \quad -P(\{X \le t) P(Y - Xb \le v) + P(X \le t, \quad -Xb \le v) = \\ &= 2\{P(X \le t, \quad -Xb \le v) - (P(X \le t) P(-Xb \le v)\}. \end{aligned}$$

If, for all $t \in \mathbb{R}^{pq}$, $v \in \mathbb{R}^p$, $F \otimes F(k_{btv}) = 0$ then $(Y - Xb)$ and X are independent. We have $Y - Xb = Y - Xb_0 - X(b - b_0) = A - X(b - b_0)$, with A and X independent. If $p = 1$, the previous Lemma gives $b = b_0$. If $p > 1$, the marginals are independent. Let $X^{(j)}$ be a row of X such that the distribution of $X^{(j)}\beta$ is nondegenerated for all $\beta \in \mathbb{R}^q - \{0\}$. Now, the previous auxiliary Lemma applies to $X^{(j)}\beta$ and the corresponding marginal of A to give $b = b_0$. The reciprocal is straightforward. □

Proof of Remark 1. Let A and X be random variables with dimensions 1 and $q > 1$, respectively. Let X_1 and X_2 be variables with the same distribution as X, and such that A, X_1 and X_2 are independent. Assume also that the distributions of $A + X'b_0$ and $A + X_1'b - X_2'(b - b_0)$ coincide. Thus, $V(A) + b_0'V(X)b_0 = V(A) + b'V(X)b + (b - b_0)'V(X)(b - b_0)$ and so $b'V(X)(b - b_0) = 0$.

We construct now a counterexample where $A + X'b_0$ and $A + X_1'b - X_2'(b - b_0)$ have the same distribution for a value of b different from 0 and b_0, such that $b'V(X)(b - b_0) = 0$. Let $q = 2$ and $X \sim N_2(\mu, \Sigma)$ with $\mu = (0,0)'$, $\Sigma = I_2$, and let $A \sim N(0,1)$. Take $b_0 = (1,0)'$ and $b = (-1/2, 1/2)'$. Define X_1 and X_2 as in the previous paragraph. On the one hand, $A + X'b_0 \sim N(0, \sigma_1^2)$ with $\sigma_1^2 = V(A) + V(X'b_0) = 2$; on the other hand, $A + X_1'b - X_2'(b - b_0) \sim N(0, \sigma_2^2)$, where $\sigma_2^2 = V(A) + V(X'b) + V(X_2'(b - b_0)) = 1 + b'b + (b - b_0)'(b - b_0) = 2$. This shows that part *(iv)* in Lemma 2.2 does not necessarily hold for $q > 1$. □

Proof of Proposition 2.1. It is enough to prove it for $\mathcal{L}$; the proof for $\mathcal{K}$ is analogous. The elements of the class $\mathcal{L}$ are sums and products of indicator functions of semiintervals and functions of the form $g_{sb}(y, x, \alpha) = I_{(-\infty,s]}(y + (\alpha - x)b)$. The class of indicator functions of semiintervals is a Vapnik-Červonenkis class. The class $\mathcal{G} = \{g_{sb} : s \in \mathbb{R}^p, b \in \mathbb{R}^q\}$ is also Vapnik-Červonenkis; so is $\mathcal{G} = \{I_C : C \in \mathcal{C}\}$, where

$$\mathcal{C} = \{\{(y, x, \alpha) : \tilde{g}_{sb}(y, x, \alpha) = y + (\alpha - x)b - s \le 0\} : s \in \mathbb{R}^p, b \in \mathbb{R}^q\}.$$

Let us show that $\mathcal{C}$ is a Vapnik-Červonenkis class of sets. If $q > 1$, the class $\mathcal{C}$ is the product of q classes which are analogous to it, built for $q = 1$; so, by Theorem 9.2.6 in Dudley (1984) (the product of Vapnik-Červonenkis classes is Vapnik-Červonenkis), it is enough to show that each of them is a Vapnik-Červonenkis class, that is, to prove that $\mathcal{C}$ is Vapnik-Červonenkis for $q = 1$. Let $\mathcal{C}^c$ be the class of complements of the sets in $\mathcal{C}$; then $\mathcal{C}^c = \text{pos}(\tilde{\mathcal{G}})$, where $\tilde{\mathcal{G}} = \{\tilde{g}_{sb} : s \in \mathbb{R}^p, b \in \mathbb{R}^q\}$. The functions in this class are the sum of a function $f(y, x, \alpha) = y$ and of functions in the vector space $\{(\alpha - x)b - s : s \in \mathbb{R}^p, b \in \mathbb{R}^q\}$ of dimension $p+q$. Thus, Theorem 9.2.1 in Dudley (1984) shows that $\mathcal{C}$ is a Vapnik-Červonenkis class and, by Lemma II.2.5 in Pollard (1984), a Vapnik-Červonenkis class of functions is a euclidean class. Then $\mathcal{L}$ is the sum of euclidean classes and Corollary 17 in Nolan and Pollard (1987) gives that $\mathcal{L}$ is also a euclidean class. □

Proof of Theorem 2.1. It is enough to check conditions in Theorem 4.8 in Arcones and Giné (1993). Given our Proposition 2.1 and Corollary 21 in Nolan and Pollard (1987), it suffices to show that Λ_n and Π_n are nondegenerate U-processes. For Λ_n,

$$F(l_{bst}((Y,X),(\gamma,\alpha)) = \int l_{bst}((y,x),(\gamma,\alpha))dF(y,x) =$$

$$= F(s,t) - I_{(-\infty,s]}(\gamma)I_{(-\infty,t]}(\alpha) - P(\gamma+(X-\alpha)b \leq s, X \leq t) +$$
$$+ P(Y+(\alpha-x)b \leq s)I_{(-\infty,t]}(\alpha) \neq 0,$$

and for Π_n,

$$F(k_{btv}((Y,X),(\gamma,\alpha))) = \int k_{btv}((y,x),(\gamma,\alpha))dF(y,x) =$$

$$= P(\{X \leq t\} \text{ and } \{-Xb \leq v\}) - P(X \leq t)I_{(-\infty,v]}(\gamma-\alpha b) -$$
$$-P(Y-Xb \leq v)I_{(-\infty,t]}(\alpha) + I_{(-\infty,t]}(\alpha)I_{(-\infty,v]}(\gamma-\alpha b) \neq 0.$$

□

Proof of Corollary 2.1. The proof is similar in all four cases and it is based on the Continuous Mapping Theorem and on the inequalities

$$|\inf_{x\in\mathcal{X}} f(x) - \inf_{x\in\mathcal{X}} g(x)| \leq \sup_{x\in\mathcal{X}} |f(x)-g(x)|,$$
$$|\sup_{x\in\mathcal{X}} f(x) - \sup_{x\in\mathcal{X}} g(x)| \leq \sup_{x\in\mathcal{X}} |f(x)-g(x)|,$$

for real functions f, g defined on $\mathcal{X}$. We will establish it for $\delta_n^{(1)}$. It is enough to show that the functional $\Psi(M) = \inf_b \sup_{s,t} |M(l_{bst})|$, $M \in l^\infty(\mathcal{L})$ is continuous. For $M_1, M_2 \in l^\infty(\mathcal{L})$,

$$|\Psi(M_1) - \Psi(M_2)| = \left| \inf_b \sup_{s,t} |M_1(l_{bst})| - \inf_b \sup_{s,t} |M_2(l_{bst})| \right| \leq$$

$$\sup_b \left| \sup_{s,t} |M_1(l_{bst})| - \sup_{s,t} |M_2(l_{bst})| \right| \leq \sup_b \sup_{s,t} ||M_1(l_{bst})| - |M_2(l_{bst})|| \leq$$
$$\leq \sup_b \sup_{s,t} |M_1(l_{bst}) - M_2(l_{bst})| = \|M_1 - M_2\|_\infty.$$

This proves the Corollary. □

Proof of Corollary 2.2. First part is trivial because $\mathcal{L}$ is a euclidean class. To prove *(i)* from it, for any continuous and bounded function H on $l^\infty(\mathcal{L})$,

$$\int Hd\sqrt{n}((F_n \otimes F_n(l_{bst}) - F \otimes F(l_{bst})) \longrightarrow \int HdG_3.$$

Also, from Theorem 2.1,

$$\int H d\sqrt{n}\left(U_n^1(l_{bst}, F) - F \otimes F(l_{bst})\right) \longrightarrow \int H dG_1.$$

By subtracting both expressions,

$$\int H d\sqrt{n}\left(U_n^1(l_{bst}, F) - F_n \otimes F_n(l_{bst})\right) \longrightarrow \int H d(G_1 - G_3),$$

and *(i)* follows. Part *(ii)* holds because we just have already shown in the proof of Corollary 2.1 that the functional Ψ is continuous in $l^\infty(\mathcal{L})$. □

Proof of Corollary 2.3. The proof is analogous for $d_n^{(1)}$ and $d_n^{(2)}$. We establish the result only for $d_n^{(1)}$. For any $b \in I\!R^q$, the functional $,_b(M) = \sup_{s,t} |M(l_{bst})|$, $M \in l^\infty(\mathcal{L})$, is continuous. Indeed, if M_1 and M_2 are in $l^\infty(\mathcal{L})$ then

$$\begin{aligned} |,_b(M_1) - ,_b(M_2)| &= \left| \sup_{s,t} |M_1(l_{bst})| - \sup_{s,t} |M_2(l_{bst})| \right| \leq \\ &\leq \sup_{s,t} ||M_1(l_{bst})| - |M_2(l_{bst})|| \leq \\ &\leq \sup_{b} \sup_{s,t} |M_1(l_{bst}) - M_2(l_{bst})| = \|M_1 - M_2\|_\infty. \end{aligned}$$

It follows that $\sup_{s,t} |\Lambda_n(l_{bst})| \longrightarrow_w \sup_{s,t} |G_1(l_{bst})|$, for all $b \in I\!R^q$. Now, using *(iii)* in Lemma 2.2,

$$d_n^{(1)} = \inf_b \sup_{s,t} |\sqrt{n} U_n^{(1)}(l_{bst})| \leq \sup_{s,t} |\sqrt{n} U_n^{(1)}(l_{b_0 st})| =$$

$$\sup_{s,t} |\Lambda_n(l_{b_0 st})| \longrightarrow_w \sup_{s,t} |G_1(l_{b_0 st})|,$$

and this proves the result. □

References

Amemiya, T. (1984). *Advanced Econometrics*. Basil Blackwell.

Arcones, M. A. and E. Giné (1993). Limit thorems for U-processes. *Annals of Probability*, **29**, 1494–1542.

Arcones, M. A. and E. Giné (1994). U-processes indexed by Vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of U-statistics with estimated parameters. *Stochastic Processes and their Applications*, **52**, 17–83.

Beran, R., A. Feuerverger, and P. Hall (1996). On nonparametric estimation of intercept and slope distributions in random coefficient regression. *Annals of Statistics*, **24**, 2569–2592.

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

Beran, R. and P. W. Millar (1994). Minimum distance estimation in random coefficient regression. *Annals of Statistics*, **22**, 1976–1992.

Breusch, T. S. and A. R. Pagan (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, **47**, 1287–1294.

Delicado, P. and J. Romo (1997). Random coefficient regressions: Parametric goodness-of-fit tests. Technical report.

Delicado, P. and J. Romo (1998). Goodness-of-fit tests in random coefficient regression models. *Annals of the Institute of Statistical Mathematics*. (To appear).

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*, pp. 2–142. Springer-Verlag, New York.

Gibbons, J. D. (1985). *Nonparametric Statistical Inference* (Second ed.). Marcel Dekker, Inc.

Griffiths, W. E. (1972). Estimation of actual response coefficients in the Hildreth-Houck random coefficient model. *J. Amer. Statist. Assoc.*, **67**, 633–635.

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

Hoffmann-Jørgensen, J. (1984). *Stochastic processes on Polish spaces*, Volume 39 of *Various Publications Series*. Matematisk Inst., Aarhus U.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T. C. Lee (1985). *The theory and practice of Econometrics* (Second ed.). Wiley, New York.

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, **17**, 107–112.

Nolan, D. and D. Pollard (1987). U-processes: rates of convergence. *Annals of Statistics*, **15**, 780–799.

Pollard, D. (1980). The minimum distance method of testing. *Metrika*, **24**, 215–227.

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

Shorack, G.R. and J.A. Wellner (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, **48**, 817–838.