Testing Constancy in Varying Coefficient Models

Miguel A. Delgado, Luis A. Arteaga-Molina

uc3m | Universidad **Carlos III** de Madrid

Testing Constancy in Varying Coefficient Models[1]

Miguel A. Delgado[2]	&	Luis A. Arteaga-Molina
Universidad Carlos III de Madrid		Universidad de Cantabria

Abstract

This article proposes tests for constancy of coefficients in semi-varying coefficients models. The testing procedure resembles in spirit the union-intersection parameter stability tests in time series, where observations are sorted according to the explanatory variable responsible for the coefficients varying. The test can be applied to model specification checks of interactive effects in linear regression models. Because test statistics are not asymptotically pivotal, critical values and p-values are estimated using a bootstrap technique. The finite sample properties of the test are investigated by means of Monte Carlo experiments, where the new proposal is compared to existing tests based on smooth estimates of the unrestricted model. We also report an application to returns of education modeling.

Keywords: Varying coefficient models; Model checks; U-I tests; Concomitants; Interactive effects model checks.

JEL Codes: C12, C14, C52.

[2] Corresponding author: e-mail: delgado@est-econ.uc3m.es

1. INTRODUCTION

This article proposes a methodology for testing coefficients constancy in semi-varying coefficient models. Let (Y, Z, X_1, X_2) be a $\mathbb{R}^{2+k_1+k_2} - valued$ random vector defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$\mathbb{E}\left(Y \mid X, Z\right) = \boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{\beta}_0\left(Z\right) + \boldsymbol{X}_2^{\mathrm{T}} \boldsymbol{\delta}_0 \ a.s., \tag{1}$$

where "T" means transpose, $\boldsymbol{\beta}_0 = \left(\beta_{00}, \beta_{01}, ..., \beta_{0k_1}\right)^{\mathrm{T}}$, $\boldsymbol{X}_1 = \left(1, X_{11}, ..., X_{1k_1}\right)^{\mathrm{T}}$ and $\boldsymbol{\delta}_0 = \left(\delta_{01}, ..., \delta_{0k_2}\right)^{\mathrm{T}}$ and $\boldsymbol{X}_2 = \left(X_{21}, ..., X_{2k_2}\right)^{\mathrm{T}}$, $\boldsymbol{\beta}_0 : \mathbb{R} \rightarrow \mathbb{R}^{1+k_1}$ is a vector of unknown functions, and $\boldsymbol{\delta}_0$ is an unknown parameter vector in $\mathbb{R}^{k_2}$. Henceforth, the discussion is centered on the case where the constant term is in $\boldsymbol{X}_1$, but the procedure also applies to the case where there is a constant intercept, i.e. when $\boldsymbol{X}_1 = \left(X_{11}, ..., X_{1k_1}\right)^{\mathrm{T}}$ and $\boldsymbol{X}_2 = \left(1, X_{21}, ..., X_{2k_2}\right)^{\mathrm{T}}$. The model with constant slopes, i.e. $Var\left(\beta_{0j}(Z)\right) = 0$ all $j = 1, ..., k_1$, is known as partly linear model, and inferences on $\beta_{00}\left(\cdot\right)$ and $\boldsymbol{\delta}_0$ have been justified under different regularity conditions by Shiller (1984), Wahba (1985), Engle et al. (1986), Heckman (1986), Schick (1986), Speckman (1988), Chen (1988) and Robinson (1988) among others. This requires estimating the nonparametric regression functions of Y given Z and of each $\boldsymbol{X}_2$ component given Z. Inferences when all the coefficients are varying, i.e. when $\boldsymbol{\delta}_0 = 0$, have been proposed by Cleveland et al. (1991), Hastie and Tibshirani (1993), Chan and Tsay (1998), McCabe and Tremayne (1995), Wu et al. (1998), Fan and Zhang (1999, 2000), Chiang et al. (2001), Hoover et al. (1998), Cai et al. (2000), Kim (2007), Hoderlain and Sherman (2015) or Feng et al. (2017). The semi-varying coefficient model, with $\boldsymbol{\delta}_0 \neq 0$, has been studied by Zhang et al. (2002), Xia et al. (2004), Ahmad et al. (2005), Fan and Huang (2005), G. Li et al. (2011), D. Li et al. (2011), Hu and Xia (2012), or K. Li et al. (2017) among others. All these methods use smooth estimators of the underlying nonparametric functions, generally Nadaraya-Watson kernel regression.

Model (1) nests discontinuous regression models where

$$\beta_0(z) = \bar{\beta}_{00} + \bar{\beta}_{01} 1_{\{z \leq z_0\}}, \tag{2}$$

for parameter vectors $\bar{\beta}_{00}$ and $\bar{\beta}_{01}$, where the discontinuity is explained by the variable Z, which is the typical alternative to parameter stability hypothesis in time series analysis, with parameters changing at an unknown time point. It is not possible consistently estimating β_0 in model (2) using smoothing based methods.

2

The goal of this article consists of testing that the varying coefficients in model (1) are constant in the direction of nonparametric alternatives, i.e. testing

$$H_0 : Var\left(\beta_{0j}\right) = 0 \text{ for all } j = 0, 1, ..., k_1 \text{ vs. } H_1 : Var\left(\beta_{0j}\right) \neq 0 \text{ for some } j = 0, 1, ..., k_1. \tag{3}$$

Model (1) also nests a model with

$$\boldsymbol{X}_2 = \left(g_0^{\mathrm{T}}(Z), X_{11}g_1^{\mathrm{T}}(Z), ..., X_{1k_1}g_{k_1}^{\mathrm{T}}(Z)\right)^{\mathrm{T}}, \quad \boldsymbol{\delta}_0 = \left(\delta_{00}^{\mathrm{T}}, \delta_{01}^{\mathrm{T}}, ..., \delta_{0k_1}^{\mathrm{T}}\right)^{\mathrm{T}} \quad \text{and} \quad k_2 = \sum_{j=0}^{k_1} m_j, \tag{4}$$

where δ_{0j} are unknown $m_j \times 1$ parameter vectors, and $g_j : \mathbb{R} \to \mathbb{R}^{m_j}$ are known functions, $j = 0, ..., k_1$. In this case, (1) can be expressed as

$$\mathbb{E}\left(Y \mid X, Z\right) = \boldsymbol{X}_1^{\mathrm{T}}\left[\boldsymbol{\beta}_0(Z) + \boldsymbol{\mu}_0(Z)\right] \ a.s. \tag{5}$$

with nonparametric $\boldsymbol{\beta}_0$ and parametric

$$\boldsymbol{\mu}_0(\cdot) = \left(g_0^{\mathrm{T}}(\cdot)\delta_{00}, \ g_1^{\mathrm{T}}(\cdot)\delta_{01}, ..., g_{k_1}^{\mathrm{T}}(\cdot)\delta_{0k_1}\right)^{\mathrm{T}},$$

for some $\boldsymbol{\delta}_0 = \left(\delta_{00}^{\mathrm{T}}, \delta_{01}^{\mathrm{T}}, ..., \delta_{0k_1}^{\mathrm{T}}\right)^{\mathrm{T}} \in \mathbb{R}^{k_2}$. Therefore, under the maintained hypothesis (3) and (5) are equivalent to omnibus model checking that the marginal effects of $\boldsymbol{X}_1$ are $\boldsymbol{\mu}_0(Z)$, which implies a particular parameterization of the interactive effects. Needless to say that model (5) is not identifiable in many circumstances, but the test we propose does not need estimating the model under the alternative hypothesis. In particular, our test is in fact a directional specification test for the linear in parameters regression model in the direction of a semi-varying coefficient model. It can also be applied as an omnibus specification test of a simple regression model with explanatory variable Z, i.e. $k_1 = 0$ in (5).

Kauermann and Tutz (1999), Cai et al. (2000), Fan and Zhang (2000), Fan et al. (2001), Fan and Huang (2005), Qu and Li (2006) and Cai et al. (2017) have considered testing (3) based on the discrepancy between restricted and unrestricted sum of squared residuals using smooth estimates of the varying coefficients. In these proposals, smooth estimates of β_{0j} are needed and, hence, situations like (2) are ruled out. Also, these tests are not applicable when the model on the alternative is not identified.

In this paper we adapt classical parameter stability tests in time series (e.g. Quandt, 1958, 1960; Chernoff and Zacks, 1964; Battacharyya and Johnson, 1968; Hinkley, 1970;

3

Brown et al., 1975; Sen and Srivastava, 1975; Hawkins, 1977, 1989; Nyblon, 1989; Andrews, 1993; Csörgő and Hortváth, 1988, 1997; Aue et al., 2008 among many others.)

Given $(Y_i, Z_i, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i})_{i=1}^n$ $i.i.d.$ as $(Y, Z, \boldsymbol{X}_1, \boldsymbol{X}_2)$, we interpret $(Y_i, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i})_{i=1}^n$ as sequentially observed with respect to the ordered values of $\{Z_i\}_{i=1}^n$. That is, denote by $\left(Y_{[i:n]}, \boldsymbol{X}_{1[i:n]}, \boldsymbol{X}_{2[i:n]}\right)_{i=1}^n$ the $Z-$concomitants, or induced order statistics, of $(Y_i, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i})_{i=1}^n$, i.e. $\left(Y_{[i:n]}, \boldsymbol{X}_{1[i:n]}, \boldsymbol{X}_{2[i:n]}\right) = (Y_j, \boldsymbol{X}_{1j}, \boldsymbol{X}_{2j})$ iff $Z_{(n:i)} = Z_j$, where $Z_{(n:1)} \leq Z_{(n:2)} \leq ... \leq Z_{(n:n)}$ are the ordered statistics of $\{Z_i\}_{i=1}^n$. We propose to adapt union-intersection (U-I) type tests in time series to our context. See Hawkins (1989), Andrews (1993), Horváth and Shao (1995) or Csörgő and Hortváth (1997, Section 3.1.5.) The test consists of comparing ordinary least squares (OLS) estimators of $\boldsymbol{X}_1$ coefficients using subsamples $\left(Y_{[i:n]}, \boldsymbol{X}_{1[i:n]}, \boldsymbol{X}_{2[i:n]}\right)_{i=1}^j$ and $\left(Y_{[i:n]}, \boldsymbol{X}_{1[i:n]}, \boldsymbol{X}_{2[i:n]}\right)_{i=j+1}^n$ at each $j-th$ sample $Z-$quantile.

The rest of the article is organized as follows. Next section discusses and justifies the testing procedure. Section 3 studies the finite sample performance of the test in the context of a Monte Carlo experiment. We report comparisons of existing tests for coefficient constancy based on smooth β_0 estimates, as well as specification CUSUM type tests, as proposed by Stute (1997) and Andrews (1997), which are omnibus, i.e. designed to detect any alternative, much broader than H_1 in (1). In Section 4 we apply the testing procedure to modeling interactive effects of IQ when studying education returns. Section 5 is devoted to conclusions. Mathematical proofs can be found in an appendix at the end of the article.

2. TESTING METHOD

Define $M_{\ell j}(u) = \mathbb{E}\left(\boldsymbol{X}_\ell \boldsymbol{X}_j^{\mathrm{T}} 1_{\{F_Z(Z) \leq u\}}\right)$ and $S_j(u) = \mathbb{E}\left(\boldsymbol{X}_j Y 1_{\{F_Z(Z) \leq u\}}\right)$, $j, \ell = 1, 2$, where F_Z is the cdf of Z. Assume that

A.1 F_Z is continuous.

A.2
$$Rank\left\{ \begin{bmatrix} M_{11}(u) & M_{12}(u) \\ M_{21}(u) & M_{22}(1) \end{bmatrix} \right\} = k_1 + k_2 + 1 \text{ for all } u \in [0, 1].$$

For the sake of exposition assume w.l.o.g. that Z is uniformly distributed on $[0, 1]$. An U-I test of H_0 is based on the sample version of $\eta_0(u) = \left(\boldsymbol{\theta}_0^- - \boldsymbol{\theta}_0^+\right)(u)$, where $\boldsymbol{\theta}_0(u) =$

$$\left(\boldsymbol{\theta}_0^{-\mathrm{T}}(u), \boldsymbol{\theta}_0^{+\mathrm{T}}(u), \boldsymbol{\theta}_0^{o\mathrm{T}}(u)\right)^{\mathrm{T}},$$

$$\boldsymbol{\theta}_0(u) = \underset{\boldsymbol{\theta}^-,\boldsymbol{\theta}^+,\boldsymbol{\theta}^o}{\arg\min} \left\{ \mathbb{E}\left[\left(Y - \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\theta}^- - \boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{\theta}^o\right)1_{\{Z \le u\}}\right]^2 \right. \tag{6}$$

$$\left. +\mathbb{E}\left[\left(Y - \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\theta}^+ - \boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{\theta}^o\right)1_{\{Z > u\}}\right]^2 \right\}$$

$$= \boldsymbol{M}^{-1}(u)\boldsymbol{S}(u), \ u \in [0,1],$$

$$\boldsymbol{M}(u) = \begin{bmatrix} M_{11}(u) & 0_{k_1+1} & M_{12}(u) \\ 0_{k_1+1} & M_{11}(1) - M_{11}(u) & M_{12}(1) - M_{12}(u) \\ M_{21}(u) & M_{21}(1) - M_{21}(u) & M_{22}(1) \end{bmatrix},$$

0_m is a $m \times m$ matrix of zeroes, and $\boldsymbol{S}(u) = \left(S_1^{\mathrm{T}}(u), [S_1(1) - S_1(u)]^{\mathrm{T}}, S_2^{\mathrm{T}}(1)\right)^{\mathrm{T}}.$

Obviously, $Var\left(\beta_{0j}(Z)\right) = 0$ for all $j = 0, ..., k_1$ implies that $\eta_0(u) = 0$ for all $u \in [0,1]$. In relevant circumstances, discussed below, also $Var\left(\beta_{0j}(Z)\right) = 0 \ \ iff \ \ \eta_0(u) = 0$ for all $u \in [0,1]$.

Remark 1 *Consider $M_{11}(u) = uM_{11}(1)$, which is always satisfied in the partly linear model, and $M_{12}(u) = uM_{12}(1)$ for all $u \in [0,1]$, which is equivalent to $\mathbb{E}\left(\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{T}}|Z\right) = M_{11}(1)$ a.s. and $\mathbb{E}\left(\boldsymbol{X}_1\boldsymbol{X}_2^{\mathrm{T}}|Z\right) = M_{12}(1)$ a.s. Therefore, $S_1(u) = M_{11}(1)\mathbb{E}\left(\beta_0(Z)1_{\{Z \le u\}}\right)$, and applying Lemma A.5 in Andrews (1993),*

$$\begin{aligned} \eta_0(u) &= \frac{\mathbb{E}\left(\beta_0(Z)1_{\{Z \le u\}}\right) - u\mathbb{E}\left(\beta_0(Z)\right)}{u(1-u)} \\ &= \frac{1}{u(1-u)} \int_{\{Z \le u\}} [\beta_0(Z) - \mathbb{E}\left(\beta_0(Z)\right)] \, d\mathbb{P} \\ &= 0 \text{ for all } u \in [0,1] \Leftrightarrow \beta_0(Z) = \mathbb{E}\left(\beta_0(Z)\right) \ a.s. \end{aligned}$$

Remark 2 *Consider $\boldsymbol{\delta}_0 = 0$, i.e. a pure varying coefficient model. Then, for all $u \in [0,1]$,*

$$\begin{aligned} \eta_0(u) &= M_{11}^{-1}(u)S_1(u) - [M_{11}(1) - M_{11}(u)]^{-1}[S_1(1) - S_1(u)] \\ &= [M_{11}(1) - M_{11}(u)]^{-1} M_{11}(1) \left[M_{11}^{-1}(u)S_1(u) - M_{11}^{-1}(1)S_1(1)\right], \end{aligned}$$

and

$$\begin{aligned} \eta_0(u) = 0 \text{ all } u \in [0,1] \ &\Leftrightarrow \ S_1(u) - M_{11}(u)M_{11}^{-1}(1)S_1(1) = 0 \text{ all } u \in [0,1] \\ &\Leftrightarrow \ \int_{\{Z \le u\}} J(Z)\left[\beta_0(Z) - \mathbb{E}\left(J(Z)\right)^{-1}\mathbb{E}\left(J(Z)\beta_0(Z)\right)\right] \, d\mathbb{P} = 0. \end{aligned}$$

with $J(Z) = \mathbb{E}\left(\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{T}}|Z\right).$ Hence, if $J(Z)$ is non-singular a.s.,

$$\begin{aligned} \eta_0(u) = 0 \text{ all } u \in [0,1] \ &\Leftrightarrow \ \beta_0(Z) = \mathbb{E}\left(J(Z)\right)^{-1}\mathbb{E}\left(J(Z)\beta_0(Z)\right) \ a.s. \\ &\Leftrightarrow \ Var\left(\beta_{0j}(Z)\right) = 0 \text{ for all } j = 0, ..., k_1. \end{aligned}$$

The above two remarks show that testing

$$\bar{H}_0 : \eta_0\left(u\right) = 0 \text{ all } u \in [0,1] \text{ vs. } \bar{H}_1 : \eta_0\left(u\right) \neq 0 \text{ some } u \in [0,1].$$

is equivalent to (3) in pure varying coefficient regression models, as well as in situations where the elements in $\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{T}}$ are mean independent of Z. Also, rejecting $\bar{H}_0$ in the direction of $\bar{H}_1$ implies rejecting H_0 in the direction of H_1 for many other semi-varying coefficient models, as discussed in Section 3.

The sample analog of (6) is $\hat{\boldsymbol{\theta}}_n(u) = \left(\hat{\boldsymbol{\theta}}_n^{-\mathrm{T}}(u), \hat{\boldsymbol{\theta}}_n^{+\mathrm{T}}(u), \hat{\boldsymbol{\theta}}_n^{o\mathrm{T}}(u)\right)^{\mathrm{T}}$, with

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_n(u) &= \underset{\boldsymbol{\theta}^-,\boldsymbol{\theta}^+,\boldsymbol{\theta}^o}{\arg\min} \left\{ \sum_{i=1}^{\lfloor nu \rfloor} \left(Y_{[i:n]} - \boldsymbol{X}_{1[i:n]}^{\mathrm{T}}\boldsymbol{\theta}^- - \boldsymbol{X}_{2[i:n]}^{\mathrm{T}}\boldsymbol{\theta}^o\right)^2 \right. \\
&\qquad\qquad \left. + \sum_{i=1+\lfloor nu \rfloor}^{n} \left(Y_{[i:n]} - \boldsymbol{X}_{1[i:n]}^{\mathrm{T}}\boldsymbol{\theta}^+ - \boldsymbol{X}_{2[i:n]}^{\mathrm{T}}\boldsymbol{\theta}^o\right)^2 \right\} \\
&= \hat{\boldsymbol{M}}_n^{-1}(u)\,\hat{\boldsymbol{S}}_n(u), \ u \in [0,1],
\end{aligned}
$$

where $\lfloor\cdot\rfloor$ means smallest nearest integer, $\hat{\boldsymbol{S}}_n(u) = \left(\hat{S}_{n1}^{\mathrm{T}}(u), \hat{S}_{n1}^{\mathrm{T}}(1) - \hat{S}_{n1}^{\mathrm{T}}(u), \hat{S}_{n2}^{\mathrm{T}}(1)\right)^{\mathrm{T}}$, $\hat{S}_{nj}(u) = \sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{j[i:n]} Y_{[i:n]}$, $j = 1, 2$,

$$
\hat{\boldsymbol{M}}_n\left(u\right) = \begin{bmatrix} \hat{M}_{11n}(u) & 0_{k_1+1} & \hat{M}_{12n}(u) \\ 0_{k_1+1} & \hat{M}_{11n}(1) - \hat{M}_{11n}(u) & \hat{M}_{12n}(1) - \hat{M}_{12n}(u) \\ \hat{M}_{21n}(u) & \hat{M}_{21n}(1) - \hat{M}_{21}(u) & \hat{M}_{n22}(1) \end{bmatrix},
$$

and $\hat{M}_{n\ell j}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor}\boldsymbol{X}_{\ell[i:n]}\boldsymbol{X}_{j[i:n]}^{\mathrm{T}}$, $\ell, j = 1, 2$. Similar expressions can be found in time series parameter stability testing. This suggests test statistics for $\bar{H}_0$ based on suitable functionals of

$$\hat{\eta}_n(u) = \left(\hat{\boldsymbol{\theta}}_n^- - \hat{\boldsymbol{\theta}}_n^+\right)(u) = R\hat{\boldsymbol{M}}_n^{-1}(u)\,\hat{\boldsymbol{S}}_n(u), \tag{7}$$

with $R = \left[I_{k_1+1} \vdots -I_{k_1+1} \vdots 0_{k_2}\right]$ and I_m is a $m \times m$ identity matrix, which is the difference between (OLS) estimators of $\boldsymbol{X}_1$ coefficients under H_0 using subsamples $\left(Y_{[i:n]}, \boldsymbol{X}_{1[i:n]}, \boldsymbol{X}_{2[i:n]}\right)_{i=1}^{j}$ and $\left(Y_{[i:n]}, \boldsymbol{X}_{1[i:n]}, \boldsymbol{X}_{2[i:n]}\right)_{i=j+1}^{n}$.

Notice that

$$\hat{\boldsymbol{\theta}}_n(u) = \boldsymbol{\theta}_0(u) + \hat{\boldsymbol{M}}_n^{-1}(u)\hat{\boldsymbol{N}}_n(u),$$

$\hat{\boldsymbol{N}}_n(u) = \left(\hat{N}_{n1}^{\mathrm{T}}(u), \hat{N}_{n1}^{\mathrm{T}}(1) - \hat{N}_{n1}^{\mathrm{T}}(u), \hat{N}_{n2}^{\mathrm{T}}(u)\right)^{\mathrm{T}}$, and $\hat{N}_{nj}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor}\boldsymbol{X}_{j[i:n]}U_{[i:n]}$, $j = 1, 2$ and $U_i = Y_i - \boldsymbol{X}_{1i}^{\mathrm{T}}\boldsymbol{\beta}_0(Z) - \boldsymbol{X}_{2i}^{\mathrm{T}}\boldsymbol{\delta}_0,$. The asymptotic distribution of $\hat{N}_n$ is obtained

applying results for partial sums of concomitants in Bathacharya (1974, 1976), extended by Sen (1976), Stute (1993, 1997) or Davydov and Egorov (2000), among others. Define $\boldsymbol{N}_\infty(u) = (N_{\infty 1}^{\mathrm{T}}(u), N_{\infty 1}^{\mathrm{T}}(1) - N_{\infty 1}^{\mathrm{T}}(u), N_{\infty 2}^{\mathrm{T}}(u))^{\mathrm{T}}$, where $N_{\infty j}$ be $k_j \times 1$, $j = 1, 2$, vectors of centered Gaussian processes with $\mathbb{E}\left(N_{\infty \ell}(u) N_{\infty j}^{\mathrm{T}}(v)\right) = \mathbb{E}\left(\boldsymbol{X}_\ell \boldsymbol{X}_j^{\mathrm{T}} U^2 1_{\{Z \leq u \wedge v\}}\right)$, $\ell, j = 1, 2$. Next assumption suffices to show weak convergence of $\sqrt{n} \hat{\boldsymbol{N}}_n$ and uniform convergence of $\hat{\boldsymbol{M}}_n$.

A3: $\mathbb{E} \|\boldsymbol{X} U\|^2 < \infty$.

Henceforth, for any matrix A, $\|A\| = \sqrt{\bar{\lambda}(A^{\mathrm{T}} A)}$ is the spectral norm, where $\bar{\lambda}(C)$ is the maximum eigenvalue of the matrix C, and "$\to_d$" means convergence in distribution of random variables, random vectors or random elements in a Skorohov's space $D[a, b]$, $0 \leq a < b \leq 1$.

Proposition 1: *Assuming A1, A2 and A3,*

$$\sqrt{n}\left(\hat{N}_{n1}^{\mathrm{T}}, \hat{N}_{n2}^{\mathrm{T}}\right)^{\mathrm{T}} \to_d (N_{\infty 1}^{\mathrm{T}}, N_{\infty 2}^{\mathrm{T}})^{\mathrm{T}} \text{ in } D[0, 1], \tag{8}$$

and

$$\lim_{n \to \infty} \sup_{u \in [0,1]} \left\|\left(\hat{M}_{n\ell j} - M_{\ell j}\right)(u)\right\| = 0 \text{ a.s., } \ell, j = 1, 2. \tag{9}$$

Therefore, since

$$\hat{\eta}_n(u) = \left(\boldsymbol{\theta}_0^- - \boldsymbol{\theta}_0^+\right)(u) + R\hat{\boldsymbol{M}}_n^{-1}(u) \hat{\boldsymbol{N}}_n(u),$$

under $\bar{H}_0$ and conditions in Proposition 1,

$$\sqrt{n}\hat{\eta}_n \to_d \eta_\infty \text{ in } D[\epsilon, 1 - \epsilon], \ \epsilon \in (0, 1),$$

where $\eta_\infty(u) \overset{d}{=} R^{\mathrm{T}} \boldsymbol{M}^{-1}(u) \boldsymbol{N}_\infty^0(u)$ with

$$\boldsymbol{N}_\infty^0(u) = \left(N_{\infty 1}^{0\mathrm{T}}(u), N_{\infty 1}^{0\mathrm{T}}(1) - N_{\infty 1}^{0\mathrm{T}}(u), N_{\infty 2}^{0\mathrm{T}}(u)\right)^{\mathrm{T}},$$

and $N_{\infty \ell}^0$, $\ell = 1, 2$ is a vector of mean zero Gaussian processes with $\mathbb{E}\left(N_{\infty \ell}^0(u) N_{\infty j}^{0\mathrm{T}}(v)\right) = \mathbb{E}\left(\boldsymbol{X}_\ell \boldsymbol{X}_j^{\mathrm{T}} V^2 1_{\{Z \leq u \wedge v\}}\right) =: \Omega_{0\ell j}(u \wedge v)$, with $V = Y - \boldsymbol{X}_1^{\mathrm{T}} \bar{\boldsymbol{\beta}}_0 - \boldsymbol{X}_2^{\mathrm{T}} \bar{\boldsymbol{\delta}}_0$ uncorrelated with the components of $(\boldsymbol{X}_1^{\mathrm{T}}, \boldsymbol{X}_2^{\mathrm{T}})^{\mathrm{T}}$ and $\left(\bar{\boldsymbol{\beta}}_0^{\mathrm{T}}, \bar{\boldsymbol{\delta}}_0^{\mathrm{T}}\right)^{\mathrm{T}} \in \mathbb{R}^{1+k_1+k_2}$. That is $(\bar{\boldsymbol{\beta}}_0, \bar{\boldsymbol{\delta}}_0)$ are the parameters of the best linear predictor of Y given $(\boldsymbol{X}_1, \boldsymbol{X}_2)$, and $V = U$ a.s. under $\bar{H}_0$. Weak convergence of $\sqrt{n}\hat{\eta}_n$ in $D[0, 1]$ is not possible even assuming that Z is independent of X

and U, as shown by Chibisov (1964) for the standard empirical process (see subsection 2.5 in Gaenssler and Stute, 1979 for discussion).

Therefore, for $\epsilon \in (0,1)$,

$$\mathbb{E}\left(\eta_\infty(u)\eta_\infty^{\mathrm{T}}(v)\right) = \boldsymbol{\Sigma}_0(u \wedge v) = R^{\mathrm{T}}\boldsymbol{M}^{-1}(u)\boldsymbol{\Omega}_0(u \wedge v)\boldsymbol{M}^{-1}(v)R, \ u,v \in (\epsilon, 1-\epsilon),$$

with $\boldsymbol{\Omega}_0(u) = \mathbb{E}\left(\boldsymbol{N}_\infty^0(u)\boldsymbol{N}_\infty^{0\mathrm{T}}(u)\right)$. Applying the U-I testing principle, this suggests tests based on functionals of the empirical process,

$$\hat{\alpha}_n(u) = \hat{\eta}_n^{\mathrm{T}}(u)\,\hat{\boldsymbol{\Sigma}}_n^{-1}(u)\hat{\eta}_n(u)\,, u \in (\epsilon, 1-\epsilon)\,.$$

where

$$\hat{\boldsymbol{\Sigma}}_n(u) = R^{\mathrm{T}}\hat{\boldsymbol{M}}_n^{-1}(u)\hat{\boldsymbol{\Omega}}_n(u)\hat{\boldsymbol{M}}_n^{-1}(u)R,$$

estimates $\boldsymbol{\Sigma}_0(u)$, and

$$\hat{\boldsymbol{\Omega}}_n(u) = \begin{bmatrix} \hat{\Omega}_{n11}(u) & 0_{k_1+1} & \hat{\Omega}_{n12}(u) \\ 0_{k_1+1} & \hat{\Omega}_{n11}(1) - \hat{\Omega}_{n11}(u) & \hat{\Omega}_{n12}(1) - \hat{\Omega}_{n12}(u) \\ \hat{\Omega}_{n21}(u) & \hat{\Omega}_{n21}(1) - \hat{\Omega}_{n21}(u) & \hat{\Omega}_{n22}(1) \end{bmatrix}$$

estimates $\boldsymbol{\Omega}_0(u)$, with $\hat{\Omega}_{n\ell j}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor}\boldsymbol{X}_{\ell[i:n]}\boldsymbol{X}_{j[i:n]}^{\mathrm{T}}\hat{V}_{[i:n]}^2$, $\ell, j = 1, 2$, and $\hat{V}_i = Y_i - \boldsymbol{X}_{1i}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}_n^+(1) - \boldsymbol{X}_{2i}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}_n^o(1)$ are the OLS residuals under $\bar{H}_0$. A sufficient condition for consistency of $\hat{\boldsymbol{\Omega}}_n(u)$ is

A4: $\mathbb{E}\|\boldsymbol{X}\|^4 < \infty$ and $\mathbb{E}\|V\|^4 < \infty$.

This condition can be relaxed by assuming that $\mathbb{E}\left(V^2\,|\,X, Z\right) = \mathbb{E}\left(V^2\right) = \sigma^2$ a.s., which implies that $\boldsymbol{\Omega}_0(u) = \sigma^2\boldsymbol{M}_n(u)$ and $\hat{\boldsymbol{\Sigma}}_n(u) = \sigma^2 R^{\mathrm{T}}\hat{\boldsymbol{M}}_n^{-1}(u)R$. Consider the U-I type test,

$$\hat{\varphi}_{n\epsilon} = \max_{\lfloor n\epsilon \rfloor + K \leq j \leq \lfloor n(1-\epsilon)\rfloor - K} n \cdot \hat{\alpha}_n\left(\frac{j}{n}\right) \text{ for small } \epsilon \in \left(0, \frac{1}{2} - \frac{K}{n}\right], \text{ with } K < \frac{n}{2},$$

and $K = k_1 + k_2 + 1$. The trimming parameter ϵ is introduced to avoid boundary points and should be chosen as close to zero as possible in order to detect any possible coefficient variation on all its domain, including those close to the boundary. However, too small ϵ values can produce serious size distortions (see section 3). The asymptotic distribution of $\hat{\varphi}_{n\epsilon}$ is derived as an immediate consequence of proposition 1 after showing uniform consistency of $\hat{\boldsymbol{\Sigma}}_n$. Define the vector of random processes,

$$\{\alpha_\infty(u)\}_{u\in[\epsilon, 1-\epsilon]} \stackrel{d}{=} \left\{\boldsymbol{N}_\infty^{0\mathrm{T}}(u)\boldsymbol{M}^{-1}(u)R^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(u)R\boldsymbol{M}^{-1}(u)\boldsymbol{N}_\infty^0(u)\right\}_{u\in[\epsilon, 1-\epsilon]}.$$

Next proposition establishes the asymptotic distribution of $\hat{\varphi}_{n\epsilon}$ as a consequence of Proposition 1, after providing consistency of $\hat{\boldsymbol{\Sigma}}_n(u)$ uniformly on $u \in [\epsilon, 1-\epsilon]$.

Proposition 2: *Assume $A1 - A4$. Under $\bar{H}_0$, for any small fixed $\epsilon \in (0, 1/2 - K/n]$,*

$K < n/2$

$$\hat{\varphi}_{n\epsilon} \to_d \varphi_{\infty\epsilon} \overset{d}{=} \sup_{u \in [\epsilon, 1-\epsilon]} \alpha_\infty(u).$$

Therefore, a test with α significance level is given by the binary random variable $\hat{\Phi}_{n\epsilon}(\alpha) = 1_{\{\hat{\varphi}_{n\epsilon} > c_\epsilon(\alpha)\}}$, where $c_\epsilon(\alpha)$ is the $(1-\alpha) - th$ quantile of $\varphi_{\infty\epsilon}$.

These U-I tests in time series are asymptotically distribution-free under suitable regularity conditions, which has a counterpart in our context assuming that

A5 Z is independent of X and U.

Of course, this assumption is not acceptable in practice, but it is worth discussing to illustrate the relation of our proposal with related ones in time series parameter instability testing and the behavior of our test statistic when ϵ is too small. Consider the $\boldsymbol{\delta}_0 = 0$ case for simplicity. Under $A.5.$, $M_{1j}(u) = uM_{1j}(1)$, $\Omega_{1j}(u) = \sigma^2 \cdot u \cdot M_{1j}(1)$, $j = 1, 2$, $\{N_{\infty 1}(u)\}_{u \in [0,1]} \overset{d}{=} \left\{ M_{11}^{1/2}(1) \cdot W_0(u) \right\}_{u \in [0,1]}$, W_0 is a $(1 + k_1) \times 1$ vector of independent Wiener's processes and

$$\boldsymbol{\Sigma}_0(u) = \sigma^2 R^{\mathrm{T}} \begin{bmatrix} uM_{11}(1) & 0 \\ 0 & (1-u)M_{11}(1) \end{bmatrix}^{-1} R = \sigma^2 \frac{M_{11}(1)}{u(1-u)}. \tag{10}$$

Therefore, under $A5$,

$$\varphi_{\infty\epsilon} \overset{d}{=} \sup_{u \in [\epsilon, 1-\epsilon]} \frac{B_0(u)}{u(1-u)}, \tag{11}$$

where $B_0(u) = [W_0(u) - uW_0(1)]^{\mathrm{T}} [W_0(u) - uW_0(1)]$ is the sum of $1 + k_1$ squared independent Brownian bridges. The distribution of $\varphi_{\infty\epsilon}$ has been tabulated by James et al. (1987) for B_0 scalar and different values of ϵ, and by Andrews (1993) in the general case.

Under $A5$, one can exploit the information in (10) and, after estimating σ^2 by $\tilde{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \hat{V}_{ni}^2$, use as test statistic,

$$\tilde{\varphi}_n^{(0)} = n \cdot \max_{K \le j \le n-K} \tilde{\alpha}_n\left(\frac{j}{n}\right),$$

with

$$\tilde{\alpha}_n(u) = \hat{\eta}_n^{\mathrm{T}}(u) \frac{\hat{M}_{11n}(1)u(1-u)}{\tilde{\sigma}_n^2} \hat{\eta}_n(u), \ u \in [0, 1],$$

which resembles the classical U-I tests avoiding any trimming. This statistics, suitably standardized, converges to a extremum distribution, which is proved applying Darling and Erdős (1956) type results for normalized partial sums. To this end, we need the alternative conditions that replaces $A3$ and $\bar{A}4$ by,

9

A3' $\mathbb{E}\,|U|^{2+\delta} < \infty$ and $\mathbb{E}\,\|\boldsymbol{X}\|^{2+\delta} < \infty$ for some $\delta > 0$.

This implies, under $A5$, that $\mathbb{E}\,\|\boldsymbol{X}U\|^{2+\delta} < \infty$, which is stronger than $A3$. These type of moment conditions was proposed by Shorak (1979) to extend Darling and Erdős (1956) result to allow less than three moments. These can be further relaxed using Einmahl (1989) moment condition. Henceforth, $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$, , and E is a random variable such that $\mathbb{P}\,(E \leq x) = \exp(-2\exp(-x))$, $a(x) = \sqrt{2\log x}$ and $b_m(x) = 2\log x + (m/2)\log\log x - \log\Gamma(m/2)$. The convergence of $\tilde{\varphi}_n^{(0)}$ is slow, which results in a poor size accuracy, and some alternatives may be preferred when A5 is satisfied.

We can also consider the Cramér-v Mises type statistic

$$\tilde{\varphi}_n^{(1)} = \sum_{j=K}^{n-K} \tilde{\alpha}_n\left(\frac{j}{n}\right)\ ,$$

and the unweighted statistic

$$\tilde{\varphi}_n^{(2)} = \max_{K \leq j \leq n-K} \frac{j(n-j)}{n}\,\tilde{\alpha}_n\left(\frac{j}{n}\right)\ ,$$

which both have better size accuracy under A5 than the test based on $\tilde{\varphi}_n^{(0)}$. Next proposition provides the limiting distribution of $\tilde{\varphi}_n^{(j)}$, $j = 0, 1, 2$ under H_0.

Proposition 3: *Assume $\boldsymbol{\delta}_0 = 0$, A1, A2, A3' and A5, under $\bar{H}_0$,*

$$a(\log n)\sqrt{\tilde{\varphi}_n^{(0)}} - b_{1+k_1}(\log n) \xrightarrow{d} E, \tag{12}$$

$$\tilde{\varphi}_n^{(1)} \xrightarrow{d} \int_0^1 \frac{B_0(u)}{u(1-u)}du, \tag{13}$$

$$\tilde{\varphi}_n^{(2)} \xrightarrow{d} \sup_{u \in [0,1]} B_0(u). \tag{14}$$

This suggests that, because the rate of convergence of $\hat{\varphi}_{n\epsilon}$ changes suddenly at $\epsilon = 0$, tests based on critical values of the asymptotic approximation (12) are expected to exhibit poor size accuracy. See simulations in section 4.

Next, we study the power of the test in the direction of sequences of local alternatives of the form,

$$\bar{H}_{n1} : \beta(Z) = \bar{\beta}_0 + \frac{\tau(Z)}{\sqrt{n}}\ a.s.,$$

for constant $\bar{\beta}_0$ and a function $\tau : \mathbb{R} \to \mathbb{R}^{1+k_1}$ such that $T(u) = \mathbb{E}\left[\boldsymbol{X}_1^{\mathrm{T}}\tau(Z)1_{\{U \leq u\}}\right]$ is bounded for all $u \in [0,1]$. Define $\boldsymbol{T}(u) = \left[T^{\mathrm{T}}(u), T^{\mathrm{T}}(1) - T^{\mathrm{T}}(u), 0_{k_2}^{\mathrm{T}}\right]^{\mathrm{T}}$ and the random processes,

$$\left\{\alpha_\infty^1(u)\right\}_{u \in [\epsilon, 1-\epsilon]} \stackrel{d}{=} \left\{(\boldsymbol{N}_\infty + \boldsymbol{T})^{\mathrm{T}}(u)\boldsymbol{M}^{-1}(u)R^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(u)R\boldsymbol{M}^{-1}(u)\,(\boldsymbol{N}_\infty + \boldsymbol{T})\,(u)\right\}_{u \in [\epsilon, 1-\epsilon]}.$$

In order to study the power of the test under $\bar{H}_{n1}$, we need the following extra assumption.

A6 $\mathbb{E} \|\boldsymbol{X}_1 \tau(Z)\| < \infty.$

Proposition 4: *Assume A1 − A4, and A6 for $\epsilon \in (0, (n − 2K)/ 2n]$, $K < n/2$. Under*
$\bar{H}_1$,

$$\hat{\varphi}_{n\epsilon} \rightarrow_p \infty, \tag{15}$$

and under $\bar{H}_{1n}$,

$$\hat{\varphi}_{n\epsilon} \rightarrow_d \sup_{u \in [\epsilon, 1−\epsilon]} \alpha^1_\infty(u), \tag{16}$$

Therefore, the test does not have trivial power in the direction of $\bar{H}_{n1}$ when $\sup_{u \in [\epsilon, 1−\epsilon]} \gamma(u) > 0$ with

$$\gamma(u) = \boldsymbol{T}^{\mathrm{T}}(u) \boldsymbol{M}^{-1}(u) R^{\mathrm{T}} \boldsymbol{\Sigma}_0^{-1}(u) R \boldsymbol{M}^{-1}(u) \boldsymbol{T}(u).$$

Under $A5$,

$$\gamma(u) = \frac{\boldsymbol{T}(u)^{\mathrm{T}} M_{11}^{-1}(1) \boldsymbol{T}(u)}{\sigma^2 \cdot u(1 − u)}.$$

This suggests choosing ϵ as small as possible in order to give more weight to the extreme values of Z.

The bootstrapped test statistic is

$$\hat{\varphi}^*_{n\epsilon} = n \sup_{K + \lfloor n\epsilon \rfloor \leq j \leq n − K − \lfloor n\epsilon \rfloor} \hat{\alpha}^*_n \left(\frac{j}{n} \right) \text{ for small } \epsilon \in \left(0, \frac{1}{2} − \frac{K}{n} \right], \ K < \frac{n}{2},$$

with

$$\hat{\alpha}^*_n(u) = \hat{\eta}^{*\mathrm{T}}_n(u) \hat{\boldsymbol{\Sigma}}^{-1}_n(u) \hat{\eta}^*_n(u),$$

and

$$\hat{\eta}^*_n(u) = R \hat{\boldsymbol{M}}^{-1}_n(u) \hat{\boldsymbol{N}}^*_n(u),$$

where $\hat{\boldsymbol{N}}^*_n(u) = \left(\hat{N}^{*\mathrm{T}}_{n1}(u), \hat{N}^{*\mathrm{T}}_{n1}(1) − \hat{N}^{*\mathrm{T}}_{n1}(u), \hat{N}^{*\mathrm{T}}_{n2}(u) \right)^{\mathrm{T}}$, $\hat{N}^*_{nj}(u) = n^{-1} \sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{j[i:n]} \hat{V}^*_{[i:n]}$, $j = 1, 2$, $\hat{V}^*_i = \hat{V}_i \xi_i$, $\{\xi_i\}_{i=1}^n$ are *i.i.d.* as ξ, which satisfies that,

A7 $\mathbb{E}(\xi) = 0$, $\mathbb{E}(\xi^2) = 1$ and $\xi \leq \kappa < \infty$ a.s

The bootstrap test, justified in next Proposition, is $\hat{\Phi}^*_{n\epsilon}(\alpha) = 1_{\{\hat{\varphi}_{n\epsilon} > \hat{c}^*_{\epsilon n}(\alpha)\}}$, where $\hat{c}^*_{\epsilon n}(\alpha) = \inf \{c : \mathbb{P}_\xi(\hat{\varphi}^*_{n\epsilon} \leq c) \geq 1 − \alpha\}$ and $\mathbb{P}_\xi$ is the induced probability of a random variable ξ.

Proposition 5: *Assume $A1 - A4$, and $A7$ for $\epsilon \in (0, (n - 2K)/2n]$, $K < n/2$. Under $\bar{H}_1$,*

$$\lim_{n \to \infty} \mathbb{P}_\xi \left(\hat{\varphi}_{n\epsilon}^* \leq c \right) = \mathbb{P} \left(\varphi_{\infty\epsilon} \leq c \right) \ \text{a.s.,}$$

and under $\bar{H}_1$ there exists a $C > 0$ such that,

$$\lim_{n \to \infty} \mathbb{P}_\xi \left(\hat{\varphi}_{n\epsilon}^* > C \right) = 1 \ a.s.$$

This implies that the asymptotic power function takes the value α under $\bar{H}_0$ and 1 under $\bar{H}_1$, i.e. $\lim_{n \to \infty} \mathbb{E} \left[\hat{\Phi}_{n\epsilon}^* (\alpha) \right] = \alpha$ under $\bar{H}_0$ and $\lim_{n \to \infty} \mathbb{E} \left[\hat{\Phi}_{n\epsilon}^* (\alpha) \right] = 1$ under $\bar{H}_1$. The test can also be based on the bootstrap $p - values$, $\hat{p}_\epsilon^* = \mathbb{P}_\xi \left(\hat{\varphi}_{n\epsilon}^* \geq \hat{\varphi}_{n\epsilon} \right)$, and we reject $\bar{H}_0$ at $\alpha - level$ of significance when $\hat{p}_\epsilon^* < \alpha$.

Since $\hat{c}_{\epsilon n}^*(\alpha)$ and $\hat{p}_\epsilon^*$ are difficult to compute in practice, they can be approximated by Monte Carlo as accurately as desired using the following algorithm.

1. Generate b sets of random numbers $\left\{ \xi_i^{(j)} \right\}_{i=1}^n$, $j = 1, ..., b$ i.i.d. as ξ, with b large.

2. Compute b test statistics $\hat{\varphi}_{n\epsilon j}^{(b)*}$, $j = 1, ..., b$, as $\hat{\varphi}_{n\epsilon}^*$, using the random numbers in 1.

Approximate the bootstrap critical values $\hat{c}_{\epsilon n}^*(\alpha)$ by

$$\hat{c}_{\epsilon n}^{(b)*}(\alpha) = \inf \left\{ c : \frac{1}{b} \sum_{j=1}^b 1_{\left\{ \hat{\varphi}_{n\epsilon j}^{(b)*} < c \right\}} \geq 1 - \alpha \right\},$$

and the corresponding $p - values$, $\hat{p}_\epsilon^*$, by

$$\hat{p}_\epsilon^{(b)*} = \frac{1}{b} \sum_{j=1}^b 1_{\left\{ \hat{\varphi}_{n\epsilon j}^{(b)*} \geq \hat{\varphi}_{n\epsilon} \right\}}.$$

The greater b, the better the bootstrap critical values and $p - values$ approximation. The same bootstrap approximations can be performed for tests based on test statistics $\tilde{\varphi}_n^{(j)}$, $j = 0, 1, 2$.

3. FINITE SAMPLE PROPERTIES

We generate samples $\{Y_i, Z_i, X_{11i}, ..., X_{1k_1i}, X_{21i}, X_{2k_2i}\}_{i=1}^n$ with

$$Y_i = \beta_{00}(Z_i) + \sum_{j=1}^{k_1} \beta_{0j}(Z_i) X_{1ji} + \sum_{j=1}^{k_2} \delta_{0j} X_{2ji} + U_i, \ i = 1, ..., n, \tag{17}$$

with $\{Z_i\}_{i=1}^n$ i.i.d. as uniform in $[0,1]$, $X_{\ell j i} = Z_i + e_{\ell j i}$, $e_{\ell j i}$ iid as uniform in $[0,1]$, $\ell = 1, 2$, $j = 1, ..., k_\ell$, and

$$U_i = \frac{\varepsilon_i \exp(\tau Z_i / 2)}{\sqrt{Var(\varepsilon_i \exp(\tau Z_i / 2))}},$$

with ε_i iid $N(0,1)$; that is, $Var(U_i) = 1$, and τ governs how severe the heteroskedasticity is. We generate the random coefficients as

$$\beta_{0j}(z) = 1 + \lambda \frac{f(z)}{\sqrt{Var(f(z))}},$$

for all $j = 0, 1, ..., k_1$, i.e. $Var(\beta_{0j}(Z)) = \lambda^2$, i.e. λ governs how serious is the departure from the null under the following models,

$$\text{a) } f(z) = z, \qquad \text{b) } f(z) = [1 + \exp(-\rho z)]^{-1},$$
$$\text{c) } f(z) = \sin(2\pi z), \quad \text{d) } f(z) = 1 + 2 \cdot 1_{\{z \le 0.4\}}.$$

Model a) is a simple linear model and b) is a nonlinear alternative, almost indistinguishable for $\rho = 1$ when $z \in [0,1]$, the lower ρ, the smaller the departure from linearity. We use model b) to check departures form linearity under different values of ρ. Model c) is harder to fit than a) or b) using smooth methods with moderate sample sizes, and d) is a jump model that cannot be estimated using smoothing methods. We only report results for the 0.4 quantile, but we have also tried other values and the results do not change substantially if the jump is not placed in extreme quantiles. Figure 1 represents η_0 for the different models and different λ values.

FIGURE 1 ABOUT HERE

The simulation study is implemented to provide evidence on the effect of ϵ's choice on $\hat{\Phi}_{n\epsilon}(\alpha)$, the accuracy of the bootstrap test, the relative performance of our test with respect to existing alternatives, and the performance of our test for model checking of interactive effects. The Monte Carlo study is based on 1.000 replications and the bootstrap replications are set to 1.000.

Figure 2 provides the percentage of rejections for different $\epsilon's$ for $\alpha = 0.05$. As expected, size accuracy is poor when ϵ is close to zero. For reasonable ϵ values, i.e. bigger that 0.1, the level is close to 5%, particularly for the larger sample sizes. On the other hand, under the alternatives, i.e., a), c) and d), the power converges to 1 as n diverges, independently of the value of ϵ. Of course , the power always increases with λ.

13

FIGURE 2 ABOUT HERE

In order to check the level accuracy of the bootstrap test, we compare the percentage of rejections using values of the asymptotic (Proposition 3) and bootstrap (Proposition 5) tests when Z is independent of X_1 and U using the test statistics $\tilde{\varphi}_n^{(j)}$, $j = 0, 1, 2$ in a pure varying coefficients model, i.e. with $\boldsymbol{\delta}_0 = \boldsymbol{0}$. Table 1 reports these results. The bootstrap tests exhibit very good size accuracy for the three test statistics. As expected, the asymptotic test based on $\tilde{\varphi}_n^{(0)}$ shows quite poor size properties, particularly for n small. However, the size accuracy of the asymptotic tests based on $\tilde{\varphi}_n^{(1)}$ and $\tilde{\varphi}_n^{(2)}$ is fairly good, but much worse than the corresponding bootstrap tests, as expected.

TABLE 1 ABOUT HERE

Now we perform the comparison with existing tests in the context of the partly linear model. We consider the omnibus specification test proposed by Stute (1997) for consistent testing of any nonparametric alternative, which is based on the CUSUM of residuals type process,

$$\hat{\psi}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i \prod_{j=1}^{k_1} 1_{\{X_{1j} \leq x_j\}} \prod_{m=1}^{k_2} 1_{\{X_{2m} \leq x_{k_1+m}\}}, \quad \boldsymbol{x} = (x_1, ..., x_{k_1+k_2})^{\mathrm{T}}.$$

The CUSUM test is designed for omnibus regression model checking i.e. it detects, in principle, any departure form linearity, including specifications different to the varying coefficient model. We consider the Kolmogorov-Smirnov type statistic,

$$\hat{\phi}_n = \sup_{\boldsymbol{x} \in \mathbb{R}^{k_1+k_2}} \sqrt{n} \left| \hat{\psi}_n(\boldsymbol{x}) \right|.$$

Our test is directional and is expected to be more powerful under H_1. We also consider the LR type bootstrap test of Cai et al. (2000) for testing $\bar{H}_0$ in the direction of $\bar{H}_1$, $\hat{T}_n = (RSS_0/RSS_1) - 1$ that compares restricted and unrestricted sum of squared residuals. LR type tests are asymptotically distribution free by the bandwidth converging to zero at a suitable rate as the sample size diverges (see Fan and Huang 2005 or Cai et al. 2017). However, tests based on critical values corresponding to the asymptotic distribution exhibit a poor size performance in finite samples. Cai et al. (2017) page 7 lines 15-19 argue that this is because the sensitivity of the test to bandwidth choice and

recommend approximating critical values with the assistance of bootstrap. This is why we only report the bootstrap version of Cai et al. (2000)'s test.

In the following set of simulations we consider different X_2 dimensions, $k_2 = 1, 2, 3$, $\lambda = 0.25$ and $\tau = 1$. Table 2 provides the percentage of rejections in this simulation study. It shows that, under $\bar{H}_1$, our directional test works better than the omnibus CUSUM as k_2 increases because of the curse of dimensionality. For instance, when $k_2 = 3$ and under model d), our test rejects more than twice than the CUSUM test. The smoothing based test has similar power than ours in all models but the jump model d), due to the poor performance of the Nadaraya-Watson estimator for estimating discontinuous regressions.

<div align="center">TABLE 2 ABOUT HERE</div>

Table 3 reports the percentage of rejections for different X_1 dimensions, $k_1 = 1, 2, 3$, $\lambda = 0.25$ and $\tau = 1$. Note that, again, our directional test works better than the omnibus CUSUM as k_1 increases. For instance, when $k_1 = 3$ and under model d), the power of our test is almost twice the CUSUM test. The test using $\hat{T}_n$ works similarly to ours in general, but our test performs better when $k_1 = 3$. The smooth test also suffers of the curse of dimensionality; the power decreases as k_1 increases. Also, the LM test detects departures from the null in the direction of jump model d) much less than the other tests, which do not require to estimate the model under the alternative using smoothing.

Under model d) our test also works much better than the LM smoothing based test because of the curse of dimensionality of the Nadaraya-Watson estimator needed to compute $\hat{T}_n$.

<div align="center">TABLE 3 ABOUT HERE</div>

In the next set of simulations we apply the test as a regression model check of the linearity hypothesis when $k_1 = 0, k_2 = 1$ and $X_2 = Z$. That is, $\bar{H}_0$ is equivalent to omnibus specification testing of the simple regression model $\mathbb{E}(Y \mid Z) = \bar{\beta}_{00} + Z\delta_{00}$ $a.s.$ The resulting test competes with the CUSUM test based on $\hat{\phi}_n$. Since β_{00} is not identifiable, tests based on comparing fits under the null and the alternative, like the LR test using $\hat{T}_n$ as test statistic, cannot be implemented. We compare our test with the omnibus specification test, designed to detect more general non-linear alternatives. We consider model $b)$ with different ρ values in order to check the performance of the test

under small departures from the linearity hypothesis. Table 4 shows that our test rejects almost double than the CUSUM test for all ρ values.

$$\boxed{\text{TABLE 4 ABOUT HERE}}$$

We also consider the test for model checking of non-linear regression models. We consider testing that $\mathbb{E}\left(Y\,|\,Z\right) = \bar{\beta}_{00} + \sum_{\ell=1}^{L} Z^{\ell}\delta_{0\ell-1}$ $a.s.$ in the direction

$$\mathbb{E}\left(Y\,|\,Z\right) = \beta_{00}(Z) + \sum_{\ell=1}^{L} Z^{\ell}\delta_{0\ell-1}a.s. \text{ with } Var\left(\beta_{00}(Z)\right) \geq 0 \ a.s.$$

and β_{00} unknown. Our test is omnibus for the nonlinear specification hypothesis, since the direction of interest nests any possible departure from the null. This corresponds to applying our test to model (5) with $g_j(z) = z^j$, $j = 1, ..., L$. Table 5 reports rejections for model b) with $\rho = 15$, which produces a sensitive departure from linearity, for different L values

$$\boxed{\text{TABLE 5 ABOUT HERE}}$$

Next, we consider the performance of the test as a specification test of interactive effects in the context of model (5) with $k_1 > 0$, $L = 1$, $g_0(z) = 1$ and $g_1(z) = z$. That is, our test is implemented for testing the hypothesis

$$\mathbb{E}\left(Y\,|\,X_1, Z\right) = \beta_{00}(Z) + \delta_{00}Z + \sum_{j=1}^{k_1} \left(\beta_{0j}(Z)X_{1j} + \delta_{0j}X_{1j}Z\right) \ a.s.$$

in the direction

$$\mathbb{E}\left(Y\,|\,X_1, Z\right) = \bar{\beta}_{00} + \delta_{00}Z + \sum_{j=1}^{k_1} \left(\bar{\beta}_{0j}X_{1j} + \delta_{0j}X_{1j}Z\right) \ a.s..$$

Table 6 reports the percentage of rejections for ours and CUSUM test in model b) with $\lambda = 0.5$, different ρ values and $k_1 = 1, 2, 3$. Our test performs better than CUSUM in most cases.

$$\boxed{\text{TABLE 6 ABOUT HERE}}$$

Now, we consider testing non-linear specification of interactive effects in the context of model (5) with $k_1 > 0$, $L = 1, 2, 3, 4$, $g_0(z) = 1$ and $g_j(z) = z^j$. Our test is implemented for testing the hypothesis

$$\mathbb{E}\left(Y\,|\,X_1, Z\right) = \beta_{00}(Z) + \sum_{\ell=1}^{L} Z^{\ell}\delta_{0\ell-1} + \sum_{j=1}^{k_1} \left(\beta_{0j}(Z)X_{1j} + X_{1j}\sum_{\ell=1}^{L} Z^{\ell}\delta_{0j+L+\ell-1}\right) a.s$$

16

in the direction

$$\mathbb{E}\left(Y \mid X_1, Z\right) = \bar{\beta}_{00} + \sum_{\ell=1}^{L} Z^\ell \delta_{0\ell-1} + \sum_{j=1}^{k_1} \left(\bar{\beta}_{0j} X_{1j} + X_{1j} \sum_{\ell=1}^{L} Z^\ell \delta_{0j+L+\ell-1} \right) \; a.s.$$

Table 7 reports the percentage of rejections for both ours and CUSUM tests under model b) with $\lambda = 0.5$, $\rho = 15$, $k_1 = 2$ and different L values. Our test performs better in general.

$$\boxed{\text{TABLE 7 ABOUT HERE}}$$

4. AN APPLICATION TO MODELING EDUCATION RETURNS

We complement the previous Monte Carlo study with an application to using IQ as control, or proxy, variable of "ability" in a returns of education model. This is based on Blackburn and Neumark (1992) work, which is used in Wooldridge (2009) textbook (example 9.3). The data consists of 663 observations from the Young Men's Cohort National Longitudinal Survey. The main objective consists of estimating the marginal effect of education on wages, controlling for relevant covariates, which include unobserved "ability". A reasonable model using IQ as proxy variable (Wooldridge, 2009, example 9.3) is

$$Log\left(WAGE\right) = \bar{\beta}_{00} + \bar{\beta}_{01} \cdot EDUC + \bar{\beta}_{02} \cdot IQ + \mathbf{X}_2^{\mathrm{T}} \boldsymbol{\delta}_{01} + U, \tag{18}$$

where $WAGE$ are USD monthly earnings, $EDUC$ is years of education, IQ is intelligence quotient (proxy of ability), and $\mathbf{X}_2^{\mathrm{T}} = (EXPER, TENURE, MARRIED, SOUTH, URBAN, BLACK)^{\mathrm{T}}$, $EXPER$ are years of work experience, $TENURE$ years with current employer, $MARRIED$ a dummy (1 if married), $BLACK$ dummy (1 if black), $SOUTH$ dummy (1 if live in south), $URBAN$ dummy (1 if live in urban area SMSA), and $\boldsymbol{\delta}_{01} = (\delta_{01}, ..., \delta_{06})^{\mathrm{T}}$. The OLS estimators of $\bar{\beta}_{01}$ and $\bar{\beta}_{02}$ in this model (heteroskedasticity robust SE in parenthesis) are $0.054\,(0.006)$ and $0.0036\,(0.001)$, respectively. The OLS estimator of the marginal effect of $EDUC$ $(\bar{\beta}_{01})$ is inconsistent when $\mathbb{E}\left(U \mid EDUC, IQ, \mathbf{X}_2\right)$ depends on $EDUC$, i.e. IQ is not a good proxy for ability, but also when it only depends on IQ in a nonlinear form. A reasonable alternative to (18) is the varying coefficients model

$$Log\left(WAGE\right) = \beta_{00}(IQ) + \beta_{01}(IQ) \cdot EDUC + \mathbf{X}_2^{\mathrm{T}} \boldsymbol{\delta}_0 + U, \tag{19}$$

which allows $EDUC$ partial effects to be an unknown function of IQ. Figure 3 provides estimates of β_{00} and β_{01} varying coefficients using Cai et al. (2000) procedure, which

17

uses a modified manifold cross-validation criterion for choosing the bandwidth. We also provide OLS estimates of the parametric specification $\beta_{0j}(IQ) = \bar{\beta}_{0j}^{(1)} + \bar{\beta}_{0j}^{(2)} IQ + \bar{\beta}_{0j}^{(3)} IQ^2$, $j = 0, 1$.

$$\boxed{\text{FIGURE 3 ABOUT HERE}}$$

The $p - values$ for testing $H_0 : Var(\beta_{0j}(IQ)) = 0$, $j = 1, 2$ versus $H_1 : Var(\beta_{0j}(IQ)) > 0$ some $j = 1, 2$, or $H_2 : Var(\beta_{00}(IQ)) = 0$ and $Var(\beta_{01}(IQ)) > 0$ are reported in Table 8, where we provide the $p - values$.

$$\boxed{\text{TABLE 8 ABOUT HERE}}$$

We also report the smoothing LR test of Cai et al. (2000). Here the CUSUM test is unable to reject the null hypothesis, but the directional tests reject H_0 in the two directions considered. The $p - value$ of our test is the smallest when testing in the direction H_1, but the corresponding $p - value$ for the smoothing LR test based on $\hat{T}_n$ is the smallest in the direction H_2.

Next, we apply our test as a model check of the interactive effect of $EDUC$. The maintained specification is

$$Log\,(WAGE) = (\beta_{00}(IQ) + \delta_{07}IQ) + (\beta_{01}(IQ) + \delta_{08}) \cdot EDUC + \mathbf{X}_2^\mathsf{T}\delta_0 + U, \qquad (20)$$

which is model (19) augmented with the explanatory variables $(IQ, EDUC)$ in the constant coefficients terms, $i.e.$ $\mathbf{X}_2^\mathsf{T}$ in (19) is substituted by $(\mathbf{X}_2^\mathsf{T}, IQ, EDUC)$ in (20). Then H_0 is in fact a specification test of the functional form of the varying coefficients in (19).

$$\boxed{\text{TABLE 9 ABOUT HERE}}$$

In this case, see table 9, we are unable to reject the specification of the interactive effect either with the CUSUM or with our test. We conclude that the specification including IQ and a simple interactive effect $EDUC$ with IQ cannot be rejected.

5. CONCLUSIONS.

We have proposed a test for constancy of coefficients in semi-varying coefficients models, where the variable responsible for the coefficient varying may depend on the rest of explanatory variables in an unknown form. The test, implemented using bootstrap, is based on comparing the OLS coefficients of subsamples of concomitants to the explanatory variable in the varying coefficients. The test is justified under fairly weak regularity

18

conditions, which allow discontinuous random coefficients under the alternative hypothesis. Our test forms a basis for specification testing of parametric varying coefficients and, in particular, for testing the functional form of interactive effects. Simulation results have provided evidence of the good performance of our test in finite samples compared with a CUSUM-type test, designed to omnibus specification testing of linear regression models, and a smooth LM test, designed to test varying coefficients constancy in the direction of smooth alternatives. The CUSUM test, like ours, does not require estimating the model on the alternative, but the LM-type test compares the restricted and unrestricted sum of squared residuals and, hence, requires estimating the nonparametric smooth varying coefficients. Simulations show that, unlike our test, the two competitors suffer of the curse of dimensionality. These also show that the LM smooth test exhibit a lack of power, compared with the two competitors, under alternatives with discontinuous varying coefficients. We have also included a real data application to model interactive effects of IQ in a returns of education model.

The proposed methodology is applicable to testing constancy of a subset of varying coefficients or a linear combination of them. However, since the model under the null must be estimated, smooth estimation of the unrestricted varying coefficients is necessary. A formal justification of the resulting test is technically demanding, but it seems possible to take advantage of existing asymptotic inference results for varying coefficient models.

A relevant extension consists of allowing endogenous explanatory variables using an instrumental variables approach, see e.g. Cai et al. (2017). Extensions to nonlinear and multiple equations estructural systems seems also feasible.

APPENDIX

Proof of Proposition 1. (8) follows from Davidov and Ergorov (2000) Theorem 1. Recall that Z is distributed as an $U(0,1)$. A typical uniformity argument shows that $\sup_{u \in [0,1]} \left\| \left(\tilde{M}_{jn} - M_j \right)(u) \right\| = o_p(1)$ $a.s.$ with

$$\tilde{M}_{jn}(u) = n^{-1} \sum_{i=1}^{n} \boldsymbol{X}_{ji} \boldsymbol{X}_{ji}^{\mathrm{T}} 1_{\{Z_i \leq u\}}, j = 1, 2.$$

Then (9) follows by noticing that $\hat{M}_{jn}(u) = \tilde{M}_{jn}(Z_{n:\lfloor nu \rfloor})$ and that $\sup_{u \in [0,1]} \left| Z_{n:\lfloor nu \rfloor} - u \right| = o(1)$ $a.s.$ by Glivenko-Cantelli theorem. ∎

Proof of Proposition 2. Define $\boldsymbol{X}_i(u) = [\boldsymbol{X}_{1i}^{\mathrm{T}}(u), \boldsymbol{X}_{1i}^{\mathrm{T}}(1) - \boldsymbol{X}_{1i}^{\mathrm{T}}(u), \boldsymbol{X}_{2i}^{\mathrm{T}}]^{\mathrm{T}}$, with $\boldsymbol{X}_{1i}(u) = \boldsymbol{X}_{1i}(1)1_{\{Z_i \leq u\}}$, and $\bar{\boldsymbol{\theta}}_0 = \left(\bar{\boldsymbol{\beta}}_0^{\mathrm{T}}, \bar{\boldsymbol{\beta}}_0^{\mathrm{T}}, \bar{\boldsymbol{\delta}}_0^{\mathrm{T}}\right)^{\mathrm{T}}$. The result is immediate from Proposition 1 assuming $A3$ and $A4$, after showing that

$$\sup_{\frac{K}{n}+\epsilon \leq u \leq \frac{K}{n}-\epsilon} \left\| \left(\hat{\boldsymbol{\Omega}}_n - \hat{\boldsymbol{\Omega}}_n^0\right)(u) \right\| = o_p(1),$$

with

$$\hat{\boldsymbol{\Omega}}_n^0(u) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i(u)\boldsymbol{X}_i^{\mathrm{T}}(u)V_i^2.$$

Under $\bar{H}_0$, $\hat{\boldsymbol{\theta}}_n(1) = \bar{\boldsymbol{\theta}}_0 + O_p(n^{-1/2})$ and

$$\hat{\boldsymbol{\Omega}}_n(u) - \hat{\boldsymbol{\Omega}}_n^0(u) = \frac{1}{n}\sum_{i=1}^{\lfloor nu \rfloor} \left(\boldsymbol{X}_{[i:n]}^{\mathrm{T}}(u)\left(\hat{\boldsymbol{\theta}}_n(1) - \bar{\boldsymbol{\theta}}_0\right)\right)^2 \boldsymbol{X}_{[i:n]}\boldsymbol{X}_{[i:n]}^{\mathrm{T}}$$

$$-2\left(\hat{\boldsymbol{\theta}}_n(1) - \bar{\boldsymbol{\theta}}_0\right)^{\mathrm{T}} \frac{1}{n}\sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{[i:n]}\boldsymbol{X}_{[i:n]}\boldsymbol{X}_{[i:n]}^{\mathrm{T}}V_{[i:n]}.$$

By Proposition 1, $\left\| \hat{\boldsymbol{\theta}}_n(1) - \bar{\boldsymbol{\theta}}_0 \right\| = O_p\left(n^{-1/2}\right)$. Then, by $A3$, $A4$, after applying Hölder's inequality, $\sup_{\frac{K}{n}+\epsilon \leq u \leq \frac{K}{n}-\epsilon} \left\| \left(\hat{\boldsymbol{\Omega}}_n - \hat{\boldsymbol{\Omega}}_n^0\right)(u) \right\| = O_p\left(n^{-1}\right) + O_p\left(n^{-1/2}\right)$. $\blacksquare$

Proof of Proposition 3. First notice that because Z is independent of $(U, \boldsymbol{X})$, the concomitants $\left\{U_{[i:n]}, \boldsymbol{X}_{[i:n]}\right\}_{i=1}^n$ are i.i.d. as $(U, \boldsymbol{X})$. Define

$$\tilde{\varphi}_n^\dagger = n \max_{K \leq \ell \leq n-K} \tilde{\alpha}_n^\dagger\left(\frac{\ell}{n}\right)$$

with

$$\hat{\eta}_n^\dagger(u) = M_{11}^{-1}(1)\frac{\hat{N}_{1n}(u) - u\hat{N}_{1n}(1)}{u(1-u)},$$

$$\tilde{\alpha}_n^\dagger(u) = n \cdot \hat{\eta}_n^{\dagger\mathrm{T}}(u)\frac{M_{11}(1)u(1-u)}{\sigma^2}\hat{\eta}_n^\dagger(u)$$

Applying the extension of Darling-Erdős' theorem (Darling and Erdős, 1956) to the vector case, as in Horváth (1993)

$$a(\log n)\sqrt{\tilde{\varphi}_n^\dagger} - b_{1+k_1}(\log n) \to_d E.$$

Therefore, it suffices to prove that

$$\sup_{K \leq \ell \leq n-K} \left| \left(\tilde{\varphi}_n^{(0)} - \tilde{\varphi}_n^\dagger\right)\left(\frac{\ell}{n}\right) \right| = o_p(1). \tag{21}$$

To this end, first, apply the Marcinkiewicz-Zygmund strong law of large numbers (Chow and Teicher, 1988, pp 125), to establish that, for the $\delta > 0$ in $A3'$,

$$\left[n\hat{M}_{11}\left(\frac{\ell}{n}\right) - \ell M_{11}(1)\right] = o\left(\ell^{2/(2+\delta)}\right) \text{ a.s. as } \ell \to \infty.$$

$$\left[n \left(\hat{M}_{11}(1) - \hat{M}_{11}\left(\frac{\ell}{n}\right) \right) - (n-\ell) M_{11}(1) \right] \;=\; o\left((n-\ell)^{2/(2+\delta)} \right) \text{ a.s. as } \ell \to \infty,$$

$$\sum_{i=1}^{n} V_i^2 - n\sigma^2 \;=\; o\left(n^{2/(2+\delta)} \right) \text{ a.s. as } n \to \infty. \qquad (22)$$

Hence,

$$\max_{K \le \ell \le n} \left\| \frac{\ell}{n} M_{11}(1) - \hat{M}_{11n}\left(\frac{\ell}{n}\right) \right\| \;=\; o_p\left(n^{-\frac{\delta}{(2+\delta)}} \right), \qquad (23)$$

$$\max_{1 \le \ell \le n-K} \left\| \frac{n-\ell}{n} M_{11}(1) - \left(\hat{M}_{11n}(1) - \hat{M}_{11n}\left(\frac{\ell}{n}\right) \right) \right\| \;=\; o_p\left(n^{-\frac{\delta}{(2+\delta)}} \right). \qquad (24)$$

By $A.2.$,

$$\max_{K \le \ell \le 1} \left\| \frac{\ell}{n} \hat{M}_{11n}^{-1}\left(\frac{\ell}{n}\right) \right\| \;=\; O_p(1), \qquad (25)$$

$$\max_{1 \le \ell \le n-K} \left\| \frac{n-\ell}{n} \left[\hat{M}_{11n}(1) - \hat{M}_{11n}\left(\frac{\ell}{n}\right) \right]^{-1} \right\| \;=\; O_p(1). \qquad (26)$$

Also, by the law of the iterated logarithm for partial sums,

$$\max_{1 \le \ell \le n} \left\| \frac{n}{\sqrt{\ell}} \hat{N}_{1n}\left(\frac{\ell}{n}\right) \right\| = O_p\left(\sqrt{\log\log n} \right) \qquad (27)$$

and by (22), (25) and (27)

$$\hat{\sigma}_n^2 - \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \left(V_i^2 - \sigma^2 \right) - \hat{S}_n^{\mathrm{T}}(1) \hat{M}_{11n}^{-1}(1) \hat{S}_n(1) = o_p\left(n^{-\delta/(2+\delta)} \right) + O_p\left(\frac{\log\log n}{n} \right).$$

Notice that,

$$n \cdot \left(\tilde{\alpha}_n - \tilde{\alpha}_n^{\dagger} \right)\left(\frac{\ell}{n}\right) \;=\; \left[\sqrt{\frac{\ell(n-\ell)}{n}} \left(\hat{\eta}_n - \hat{\eta}_n^{\dagger} \right)\left(\frac{\ell}{n}\right) \right]^{\mathrm{T}} \left(\frac{M_{11}}{\sigma^2 + o_p\left(n^{-\frac{\delta}{2+\delta}} \right)} \right)$$
$$\times \left[\sqrt{\frac{\ell(n-\ell)}{n}} \left(\hat{\eta}_n - \hat{\eta}_n^{\dagger} \right)\left(\frac{\ell}{n}\right) \right],$$

where

$$\sqrt{\frac{\ell(n-\ell)}{n}} \left(\hat{\eta}_n - \hat{\eta}_n^{\dagger} \right)\left(\frac{\ell}{n}\right) = \sqrt{\frac{\ell(n-\ell)}{n}} \left\{ \left[\hat{M}_{11n}^{-1}\left(\frac{\ell}{n}\right) - \frac{n}{\ell} M_{11}^{-1}(1) \right] \hat{N}_{1n}\left(\frac{\ell}{n}\right) \right.$$
$$- \left[\left(\hat{M}_{11n}(1) - \hat{M}_{11n}\left(\frac{\ell}{n}\right) \right)^{-1} - \frac{n}{n-\ell} M_{11}^{-1}(1) \right]$$
$$\left. \times \left[\hat{N}_{1n}(1) - \hat{N}_{1n}\left(\frac{\ell}{n}\right) \right] \right\}$$

$$= \sqrt{\frac{(n-\ell)\,\ell^2}{n^3}} \left\{ \frac{n}{\ell} \hat{M}_{11n}^{-1} \left(\frac{\ell}{n} \right) \left[\frac{\ell}{n} M_{11}(1) - \hat{M}_{11n} \left(\frac{\ell}{n} \right) \right] M_{11}^{-1}(1) \frac{n}{\sqrt{\ell}} \hat{N}_{1n} \left(\frac{\ell}{n} \right) \right\}$$

$$- \sqrt{\frac{\ell\,(n-\ell)^2}{n^3}} \left\{ \frac{n}{n-\ell} \left(\hat{M}_{11n}(1) - \hat{M}_{11n} \left(\frac{\ell}{n} \right) \right)^{-1} \right.$$

$$\times \left[\frac{n-\ell}{n} M_{11}(1) - \left(\hat{M}_{11n}(1) - \hat{M}_{11n} \left(\frac{\ell}{n} \right) \right) \right]$$

$$\left. \times M_{11}^{-1}(1) \frac{n}{\sqrt{n-\ell}} \left[\hat{N}_{1n}(1) - \hat{N}_{1n} \left(\frac{\ell}{n} \right) \right] \right\}$$

Therefore, by (23)-(27),

$$\sqrt{\frac{\ell\,(n-\ell)}{n}} \max_{K \leq \ell \leq n-K} \left\| (\hat{\eta}_n - \hat{\eta}_n^\dagger) \left(\frac{\ell}{n} \right) \right\| \leq O_p(1) \times o_p \left(\frac{1}{n^{\frac{\delta}{(2+\delta)}}} \right) \times O_p \left(\sqrt{\log\log n} \right)$$

$$= o_p(1), \tag{28}$$

which proves (21).

Finally, (13) and (14) follow by (28) and

$$\left\{ u(1-u) M^{1/2}(1) \hat{\eta}_n^\dagger(u) / \sigma \right\}_{u \in [0,1]} \to_d \left\{ W_0(u) - u W_0(1) \right\}_{u \in [0,1]} \ in \ D\,[0,1]\,,$$

by Proposition 1. ∎

Proof of Proposition 4. First notice that, by Proposition 1 and 2, uniformly in $u \in [\epsilon, 1-\epsilon]$,

$$\frac{\tilde{\varphi}_n}{n} \to_p = \sup_{\epsilon \leq u \leq 1-\epsilon} \eta_0^{\mathrm{T}}(u) \boldsymbol{\Sigma}_0^{-1}(u) \eta_0(u),$$

which proves (15). In order to prove, (16), notice that under $\bar{H}_{1n}$

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}_0 \right)(u) = \hat{\boldsymbol{M}}_n^{-1}(u) \left[\frac{1}{n} \sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{[i:n]} \tau(Z_{i:n}) + \sqrt{n} \hat{\boldsymbol{N}}_n(u) \right],$$

where, under $A6$, $\sup_{\epsilon \leq u \leq 1-\epsilon} \left\| n^{-1} \sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{[i:n]} \tau(Z_{i:n}) - \boldsymbol{T}(u) \right\| = o(1)$ a.s. using same arguments to prove (8) in Proposition 1. Then, apply (8), (9) and the continuous mapping Theorem to complete the proof. ∎

Proof of Proposition 5. Let $\mathbb{P}_\xi$ be the induced probability distribution of ξ. It suffices to show that for any $c > 0$,

$$\mathbb{P}_\xi \left(\hat{\varphi}_{n\epsilon}^* \leq c \right) \to \mathbb{P} \left(\varphi_{\infty\epsilon} \leq c \right) + o(1) \ a.s., \tag{29}$$

Notice that uniformly in $u \in [0,1]$,

$$\hat{\eta}_n^*(u) = R \left[\boldsymbol{M}(u) + o(1) \right]^{-1} \hat{\boldsymbol{N}}_n^*(u) \ a.s.,$$

22

Mimicking the strategy of proof in Stute al. (1998) (SGQ) for a similar bootstrap statistics, (29) follows by showing that conditional to the sample, $\sqrt{n}\hat{\boldsymbol{N}}_n^*$ converges in distribution to $\boldsymbol{N}_\infty$ a.s., i.e. for almost all sample $\{Y_i, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, Z_i\}_{i=1}^n$, by showing the convergence of the finite dimensional distributions (fidis) and tightness. Henceforth, $\mathbb{E}_\xi$ is the expectation operator corresponding to $\mathbb{P}_\xi$. For fidis convergence, first notice that for $u_1, u_2 \in [0,1]$,

$$
\begin{aligned}
n\mathbb{E}_\xi\left[\hat{\boldsymbol{N}}_n^*(u_1)\hat{\boldsymbol{N}}_n^{*\mathrm{T}}(u_2)\right] &= \frac{1}{n}\sum_{i=1}^{\lfloor n(u_1 \wedge u_2)\rfloor} \boldsymbol{X}_{[i:n]}\boldsymbol{X}_{[i:n]}^{\mathrm{T}}\hat{V}_{[i:n]}^2 \\
&= \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}V_i^2 \mathbf{1}_{\left\{Z_i \leq Z_{\lfloor n(u_1 \wedge u_2)\rfloor:n}\right\}} \\
&\quad + \frac{1}{n}\sum_{i=1}^n \left[\boldsymbol{X}_i^{\mathrm{T}}\left(\hat{\boldsymbol{\theta}}_n(1)-\bar{\boldsymbol{\theta}}_0\right)\right]^2 \boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}}V_i^2 \mathbf{1}_{\left\{Z_i \leq Z_{\lfloor n(u_1 \wedge u_2)\rfloor:n}\right\}} \\
&= \boldsymbol{\Omega}_0\left(u_1 \wedge u_2\right) + o(1) \ a.s. \tag{30}
\end{aligned}
$$

since $\hat{\boldsymbol{\theta}}_n(1) = \bar{\boldsymbol{\theta}}_n + o(1)$ a.s., and applying the arguments for proving (9) using A3 and A4. Then, fixing $u_1, ..., u_q$, by the Cramér-Wold device, it suffices to show that for any $c < \infty$,

$$
\mathbb{P}_\xi\left\{\sqrt{n}\sum_{j=1}^q a_j\boldsymbol{b}^{\mathrm{T}}\hat{\boldsymbol{N}}_n^*(u_j) \leq c\right\} \to_p \mathbb{P}\left\{\sum_{j=1}^q a_j\boldsymbol{b}^{\mathrm{T}}\boldsymbol{N}_\infty(u_j) \leq c\right\}, \tag{31}
$$

for any $\boldsymbol{b} = (b_1, ..., b_{k+1})^{\mathrm{T}}$ and $\boldsymbol{a} = (a_1, ..., a_p)^{\mathrm{T}}$. Write $W_i = \sum_{j=1}^q a_j\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_i\mathbf{1}_{\left\{Z_i \leq Z_{\lfloor nu_j\rfloor:n}\right\}}$

$$
\begin{aligned}
\sqrt{n}\sum_{j=1}^q a_j\boldsymbol{b}^{\mathrm{T}}\hat{\boldsymbol{N}}_n^*(u_j) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\hat{V}_i\xi_i \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i\hat{V}_i\xi_i.
\end{aligned}
$$

Then (31) follows by checking the Linderberg's condition

$$
L_n(\delta) = \frac{1}{n}\sum_{i=1}^n W_i^2\int_{\left\{|W_i\hat{V}_i\xi_i| \geq \delta\sqrt{n}\right\}} W_i^2\hat{V}_i^2\xi_i^2 d\mathbb{P}_\xi \to 0 \ a.s.
$$

Define $\bar{W}_i = \sum_{j=1}^q a_j\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_i$. Since $|\xi| \leq \tau$,

$$
\begin{aligned}
L_n(\delta) &\leq \frac{\kappa^2}{n}\sum_{i=1}^n \mathbf{1}_{\left\{|\bar{W}_i\hat{V}_i| \geq \frac{\delta\sqrt{n}}{\tau}\right\}}\bar{W}_i^2\hat{V}_i^2 \\
&= \frac{\kappa^2}{n}\sum_{i=1}^n \mathbf{1}_{\left\{|\bar{W}_iV_i| \geq \frac{\delta\sqrt{n}}{\tau}\right\}}\bar{W}_i^2V_i^2 + o(1) \ a.s. \\
&= o(1) \ a.s.
\end{aligned}
$$

23

using arguments in (30). In order to show tightness, it suffices to check Billingsley (1968) Theorem 15.7 as in SGQ Lemma A3. Define

$$\hat{\alpha}_{n\boldsymbol{b}}^*(u) = \sqrt{n}\,\boldsymbol{b}^{\mathrm{T}}\hat{\boldsymbol{N}}_n^*(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_i 1_{\left\{Z_i \le Z_{\lfloor nu \rfloor:n}\right\}}\hat{V}_i\xi_i.$$

We must show, as in SGQ Lemma 3, that for any $\boldsymbol{b} \in \mathbb{R}^{1+k}$ and $0 \le u_0 \le u_1 \le u_2 \le 1$,

$$\mathbb{E}_\xi\left[\hat{\alpha}_{n\boldsymbol{b}}^*(u_1) - \hat{\alpha}_{n\boldsymbol{b}}^*(u_0)\right]^2\left[\hat{\alpha}_{n\boldsymbol{b}}^*(u_2) - \hat{\alpha}_{n\boldsymbol{b}}^*(u_1)\right]^2 \le C\left[J_n(u_2) - J_n(u_0)\right]^2, \qquad (32)$$

where $C < \infty$ is a generic constant, J_n monotone $a.s.$, and $J_n \to J$ a.s. Then, applying Lemma 5.1 of Stute (1997),

$$LHS(32) \le \frac{3}{n^2}\sum\sum_{i\ne j}\mathbb{E}_\xi\lambda_i^2\mathbb{E}_\xi\gamma_i^2,$$

$\lambda_i = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_i\hat{V}_i\xi_i 1_{\left\{Z_{\lfloor nu_0\rfloor:n}\le Z_i\le Z_{\lfloor nu_1\rfloor:n}\right\}}$ and $\gamma_i = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_i\hat{V}_i\xi_i 1_{\left\{Z_{\lfloor nu_2\rfloor:n}\le Z_i\le Z_{\lfloor nu_1\rfloor:n}\right\}}$. Then,

$$
\begin{aligned}
LHS(32) &\le \frac{3}{n^2}\sum\sum_{i\ne j}\left(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_i\right)^2\left(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_j\right)^2\hat{V}_i^2\hat{V}_j^2 1_{\left\{Z_{\lfloor nu_0\rfloor:n}\le Z_i\le Z_{\lfloor nu_1\rfloor:n}\right\}}1_{\left\{Z_{\lfloor nu_1\rfloor:n}\le Z_i\le Z_{\lfloor nu_2\rfloor:n}\right\}}\\
&\le 3\left[J_n(u_2) - J_n(u_0)\right]^2,
\end{aligned}
$$

where

$$J_n(u) = \frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}_i\right)^2\hat{V}_i^2 1_{\left\{Z_i\le Z_{\lfloor nu\rfloor:n}\right\}}$$

is monotone and $J_n(u) \to J(u) = \mathbb{E}\left(\left(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X}\right)^2 V^2 1_{\{Z\le u\}}\right)$ $a.s.$ uniformly in u. ∎

REFERENCES

Ahmad, I., Leelahanon, S., Li, Q., 2005. Efficient estimation of a semiparametric partially linear varying coefficient model. Ann. Stat. 33, 258–283.

Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 65, 1097–1128.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Aue, A., Horváth, L., Kokoszka, P., Steinebach, J., 2008. Monitoring shifts in mean: Asymptotic normality of stopping times. Test 17, 515–530.

Bhattacharya, P.K., 1974. Convergence of sample paths of normalized sums of induced order statistics. Ann. Stat. 2, 1034–1039.

Bhattacharya, P.K., 1976. An invariance principle in regression analysis. Ann. Stat. 4, 621–624.

Bhattacharyya, G.K., Johnson, R.A., 1968. Nonparametric tests for shift at an unknown time point. Ann. Math. Stat. 39, 1731–1743.

Billingsley, P., 1968. Convergence of probability measures. John Wiley & Sons.

Blackburn, M.L., Neumark, D., 1995. Are OLS estimates of the return to schooling biased downward? Another look. Rev. Econ. Stat. 77, 217–230.

Brown, R.L., Durbin, J., Evans, J.M., 1975. Techniques for testing the constancy of regression relationships over time. J. R. Stat. Soc. Ser. B 37, 149–192.

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. J. Am. Stat. Assoc. 95, 941–956.

Cai, Z., Fan, J., Lin, M. and Su, J. (2017). Inference for a partially varying coefficient model with endogenous regressors, J. Bus. Econ. Stat. In Press.

Chan, K.S., Tsay, R.S., 1998. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. Biometrika 85, 413–426.

Chen, H., 1988. Convergence rates for parametric components in a partly linear model. Ann. Stat. 16, 136–146.

Chernoff, H., Zacks, S., 1964. Estimating the current mean of a normal distribution which is subjected to changes in time. Ann. Math. Stat. 35, 999–1018.

Chiang, C.-T., Rice, J.A., Wu, C.O., 2001. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. J. Am. Stat. Assoc. 96, 605–619.

Chibisov, D.M., 1964. Some theorems on the limiting behaviour of an empirical distribution function. Tr. Mat. Instituta Im. VA Steklova 71, 104–112.

Chow, Y.S., Teicher, H., 1988. Probabilty theory.

Cleveland, W.S., Grosse, E., Shyu, W.M., 1991. Local regression models, in: Chambers, J.M., Hastie, T.J. (Eds.), Statistical Models in S. Pacific Grove: Wadsworth & Brooks, pp. 309–376.

Csörgő, M., Horváth, L., 1997. Limit theorems in change-point analysis, Wiley series in probability and statistics. Wiley.

Csörgő, M., Horváth, L., 1988. 20 Nonparametric methods for changepoint problems. Handb. Stat. 7, 403–425.

Darling, D.A., Erdős, P., others, 1956. A limit theorem for the maximum of normalized sums of independent random variables. Duke Math. J. 23, 143–155.

Davydov, Y., Egorov, V., 2000. Functional limit theorems for induced order statistics. Math. Methods Stat. 9, 297–313.

Einmahl, U., 1989. The Darling-Erdős theorem for sums of iid random variables. Probab. Theory Relat. Fields 82, 241–257.

Engle, R.F., Granger, C.W.J., Rice, J., Weiss, A., 1986. Semiparametric estimates of the relation between weather and electricity sales. J. Am. Stat. Assoc. 81, 310–320.

Fan, J., Huang, T., 2005. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11, 1031–1057.

Fan, J., Zhang, C., Zhang, J., 2001. Generalized likelihood ratio statistics and wilks phenomenon. Ann. Stat. 29, 153–193.

Fan, J., Zhang, W., 2000. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. Scand. J. Stat. 27, 715–731.

Fan, J., Zhang, W., 1999. Statistical estimation in varying coefficient models. Ann. Stat. 27, 1491–1518.

Feng, G., Gao, J., Peng, B., Zhang, X., 2017. A varying-coefficient panel data model with fixed effects: Theory and an application to US commercial banks. J. Econom. 196, 68–82.

Gaenssler, P., Stute, W., 1979. Empirical processes: A survey of results for independent and identically distributed random variables. Ann. Probab. 7, 193–243.

Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. R. Stat. Soc. Ser. B 55, 757–796.

Hawkins, D.L., 1989. A u-i approach to retrospective testing for shifting parameters in a linear model. Commun. Stat. Methods 18, 3117–3134.

Hawkins, D.M., 1977. Testing a sequence of observations for a shift in location. J. Am. Stat. Assoc. 72, 180–186.

Heckman, N.E., 1986. Spline smoothing in a partly linear model. J. R. Stat. Soc. Ser. B 48, 244–248.

Hinkley, D. V, 1970. Inference about the change-point in a sequence of random variables. Biometrika 57, 1–17.

Hoderlein, S., Sherman, R., 2015. Identification and estimation in a correlated random coefficients binary response model. J. Econom. 188, 135–149.

Hoover, D.R., Rice, J.A., Wu, C.O., Yang, L.-P., 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85, 809–822.

Horváth, L., 1993. The maximum likelihood method for testing changes in the parameters of normal observations. Ann. Stat. 21, 671-680.

Horváth, L., Shao, Q.-M., 1995. Limit theorems for the union-intersection test. J. Stat. Plan. Inference 44, 133–148.

Hu, T., Xia, Y., 2012. Adaptive semi-varying coefficient model selection. Stat. Sin. 22, 575–599.

James, B., James, K.L., Siegmund, D., 1987. Tests for a change-point. Biometrika 74, 71–83.

Kauermann, G., Tutz, G., 1999. On model diagnostics using varying coefficient models. Biometrika 86, 119–128.

Kim, M.-O., 2007. Quantile regression with varying coefficients. Ann. Stat. 35, 92–108.

Li, D., Chen, J., Lin, Z., 2011. Statistical inference in partially time-varying coefficient models. J. Stat. Plan. Inference 141, 995–1013.

Li, G., Xue, L., Lian, H., 2011. Semi-varying coefficient models with a diverging number of components. J. Multivar. Anal. 102, 1166–1174.

Li, K., Li, D., Liang, Z., Hsiao, C., 2017. Estimation of semi-varying coefficient models with nonstationary regressors. Econom. Rev. 36, 354–369.

McCabe, B.P.M., Tremayne, A.R., 1995. Testing a time series for difference stationarity. Ann. Stat. 23, 1015–1028.

Nyblom, J., 1989. Testing for the constancy of parameters over time. J. Am. Stat. Assoc. 84, 223–230.

Qu, A., Li, R., 2006. Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics 62, 379–391.

Quandt, R.E., 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. J. Am. Stat. Assoc. 55, 324–330.

Quandt, R.E., 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. J. Am. Stat. Assoc. 53, 873–880.

Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.

Schick, A., 1986. On asymptotically efficient estimation in semiparametric models. Ann. Stat. 14, 1139–1151.

Sen, A., Srivastava, M.S., 1975. On tests for detecting change in mean. Ann. Stat. 3, 98–108.

Sen, P.K., 1976. A note on invariance principles for induced order statistics. Ann. Probab. 4, 474–479.

Shiller, R.J., 1984. Smoothness priors and nonlinear regression. J. Am. Stat. Assoc. 79, 609–615.

Shorack, G.R., 1979. The weighted empirical process of row independent random variables with arbitrary distribution functions. Stat. Neerl. 33, 169–189.

Speckman, P., 1988. Kernel smoothing in partial linear models. J. R. Stat. Soc. Ser. B 50, 413–436.

Stute, W., Manteiga, W.G., Quindimil, M.P., 1998. Bootstrap approximations in model checks for regression. J. Am. Stat. Assoc. 93, 141–149.

Stute, W., 1993. U-functions of concomitants of order statistics. Probab. Math. Stat. 14, 143–155.

Stute, W., 1997. Nonparametric model checks for regression. Ann. Stat. 25, 613–641.

Wahba, G., 1985. A comparison of GCV and GML for choosing the fmoothing parameter in the generalized spline smoothing problem. Ann. Stat. 13, 1378–1402.

Wooldridge, J.M., 2009. Introductory econometrics: A modern approach. Nelson Education.

Wu, C.O., Chiang, C.-T., Hoover, D.R., 1998. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. J. Am. Stat. Assoc. 93, 1388–1402.

Xia, Y., Zhang, W., Tong, H., 2004. Efficient estimation for semivarying-coefficient models. Biometrika 91, 661–681.

Zhang, W., Lee, S.-Y., Song, X., 2002. Local polynomial fitting in semivarying coefficient model. J. Multivar. Anal. 82, 166–188.
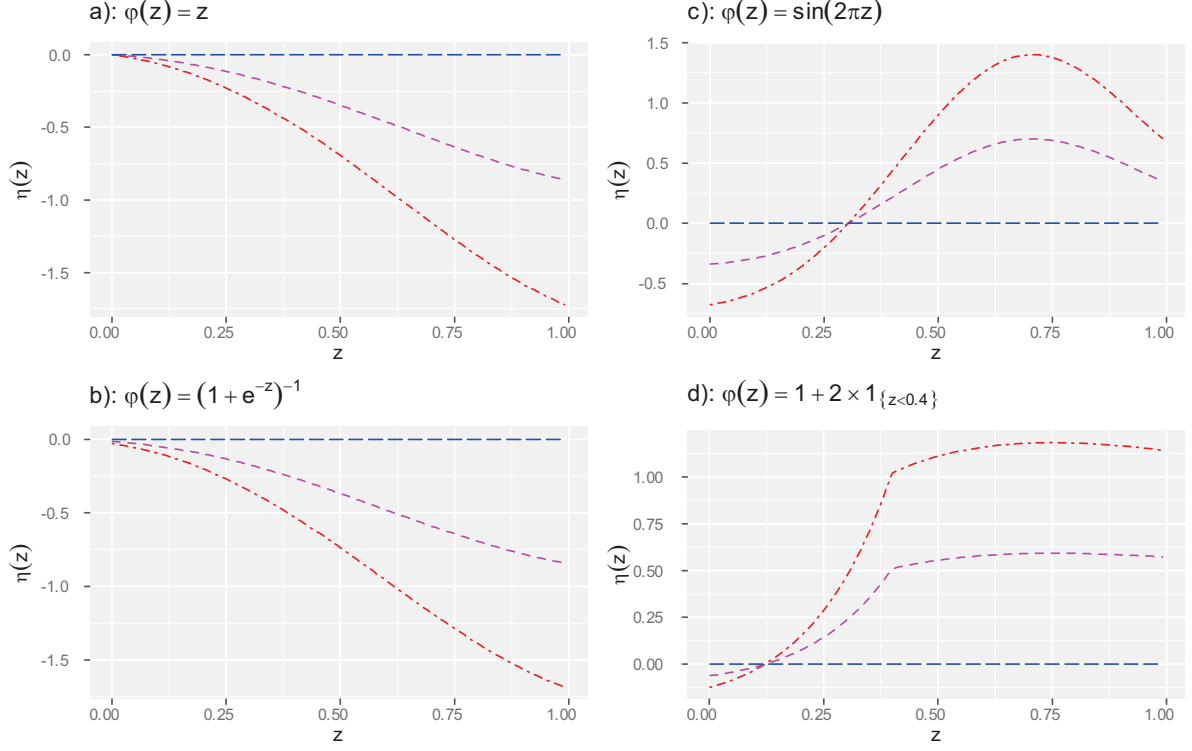
Figure 1: *Representation of η_0 for different models when $\lambda = 0$ (blue curve), $\lambda = 0.25$ (purple curve), and $\lambda = 0.5$ (red curve).*
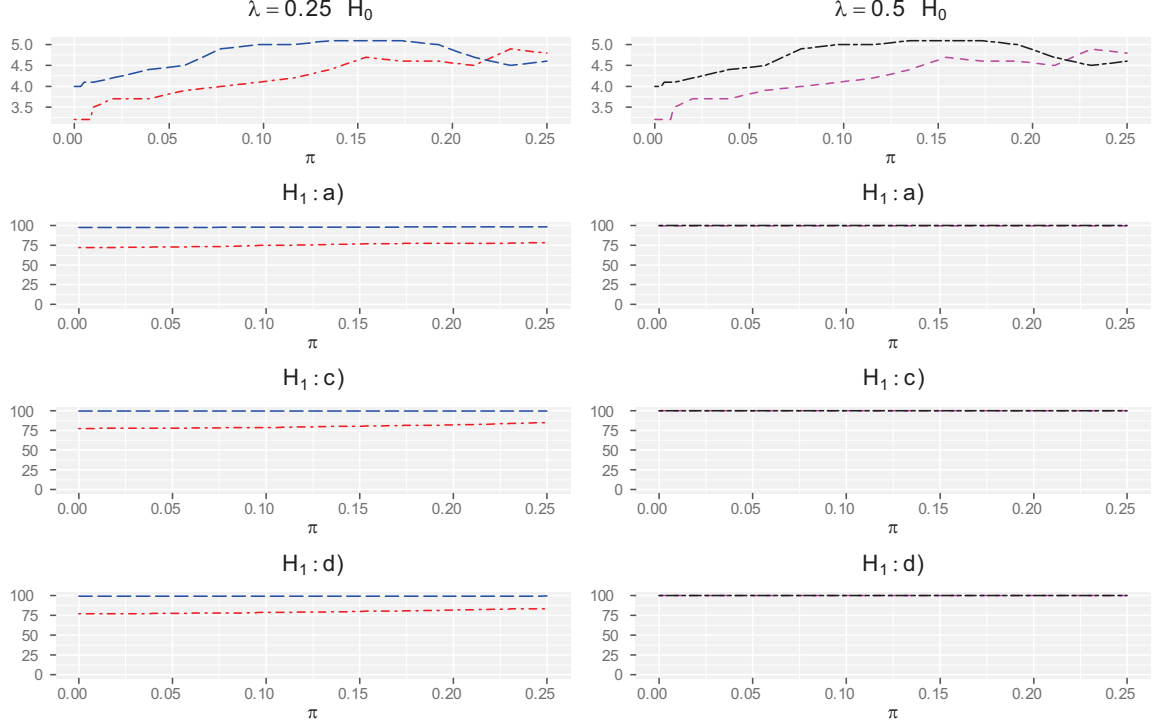

Figure 2: *Representation of $\hat{\Phi}_{n\epsilon}(\alpha)$ for different models when $\lambda = 0.25$ (red curve) and $\lambda = 0.5$ (blue curve).*
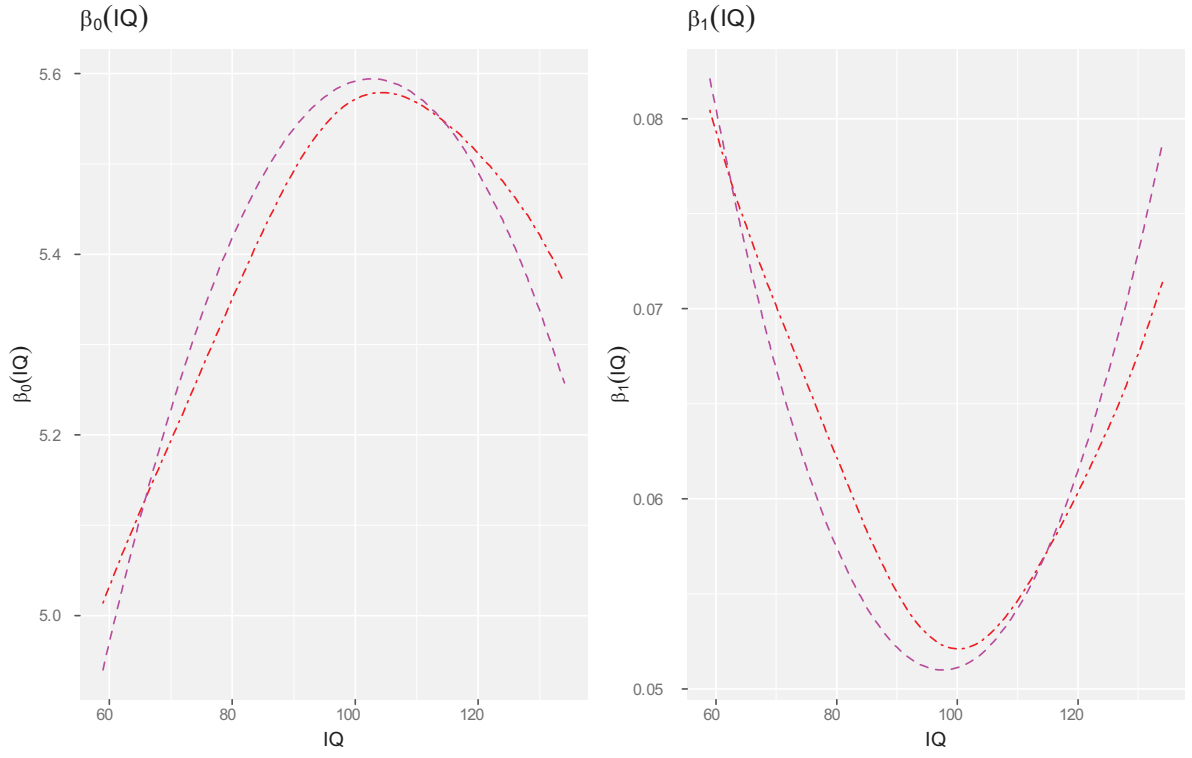
Figure 3: *Representation of $\beta_{00}(IQ)$ and $\beta_{01}(IQ)$ for the estimates of the varying coefficients using kernels with a plug-in bandwidth (red curve), and OLS estimates of the parametrization (purple curve).*

α	1%				5%				10%			
k_1	0	1	2	3	0	1	2	3	0	1	2	3
$\tilde{\varphi}_n^{(0)}$ (bootstrap)												
50	0.2	0.3	0.2	0.5	2.5	2.5	2.8	2.4	5.7	6.1	6.4	5.9
100	0.5	0.5	0.7	0.6	3.5	3.2	3.1	2.4	8.5	6.6	6.0	5.8
200	1.2	1.1	0.5	0.4	4.4	4.2	3.6	2.1	8.4	7.9	6.3	4.4
500	0.7	0.7	0.7	1.1	4.1	3.8	3.6	4.5	9.1	7.7	8.4	8.4
$\tilde{\varphi}_n^{(1)}$ (bootstrap)												
50	0.6	0.7	0.1	0.1	4.1	4.1	3.2	3.4	8.7	9.6	8.0	8.3
100	1.2	1.0	0.7	0.5	4.6	4.5	3.9	3.6	9.5	9.4	9.1	8.5
200	1.2	1.2	0.7	0.7	5.3	4.3	5.2	3.9	9.7	10.5	9.2	7.8
500	1.0	0.9	0.6	1.2	4.7	4.1	4.5	5.1	11.1	9.0	8.7	9.5
$\tilde{\varphi}_n^{(2)}$ (bootstrap)												
50	0.7	1.0	0.3	1.3	4.5	5.1	5.2	4.6	9.2	11.6	9.9	11.3
100	1.0	1.0	0.7	0.6	5.1	4.5	4.5	4.4	11.0	9.4	9.2	9.6
200	1.2	1.0	0.7	0.7	5.2	5.1	4.9	2.6	10.3	10.8	9.2	8.4
500	1.2	1.0	1.2	1.5	4.7	5.4	5.7	5.9	9.7	10.0	9.4	10.2
$\tilde{\varphi}_n^{(0)}$ (asymptotic)												
50	0.0	0.1	0.5	3.6	1.7	2.4	6.7	23.3	5.9	8.2	18.1	43.9
100	0.0	0.0	0.1	1.0	1.3	1.1	3.9	9.1	5.3	5.9	10.3	21.8
200	0.0	0.0	0.0	0.4	1.5	1.4	2.5	4.6	4.4	4.6	5.9	13.4
500	0.0	0.0	0.0	0.0	1.5	1.0	2.4	3.3	4.3	3.9	6.2	10.3
$\tilde{\varphi}_n^{(1)}$ (asymptotic)												
50	0.5	0.1	0.0	0.0	2.9	2.7	1.9	1.0	6.8	6.4	5.1	3.3
100	1.0	0.7	0.6	0.1	4.9	3.8	3.7	2.4	10.8	8.5	7.9	6.7
200	1.3	1.2	0.5	0.6	4.6	5.3	4.0	4.1	8.5	10.3	8.8	7.4
500	0.8	1.1	0.8	0.4	4.9	4.5	4.9	4.3	9.5	9.4	9.5	8.7
$\tilde{\varphi}_n^{(2)}$ (asymptotic)												
50	0.2	0.1	0.1	0.0	2.0	1.5	1.2	1.4	4.9	4.0	4.1	3.7
100	0.3	0.2	0.4	0.1	3.3	2.5	2.6	4.6	7.8	5.5	5.0	4.6
200	0.7	0.7	0.4	0.3	4.1	3.5	3.2	1.6	8.2	7.2	6.4	4.1
500	0.7	0.7	0.7	0.8	4.4	3.9	4.0	4.7	8.1	8.3	7.8	8.1

Table 1. *Percentage of times H_0 was rejected ($k_2 = 0$ and $\tau = 0$)*

Model	H_0			$H_1 : a$			$H_1 : c$			$H_1 : d$		
k_2	1	2	3	1	2	3	1	2	3	1	2	3
$\hat{\varphi}_{n0.02}$												
50	3.3	4.4	4.6	11.5	9.5	7.1	13.3	12.5	11.4	15.0	11.2	10.3
100	4.0	5.0	4.6	26.6	15.9	12.4	25.9	23.0	21.2	30.0	20.5	19.8
200	4.5	4.3	3.6	49.1	31.4	22.4	56.0	45.6	40.6	60.9	45.4	38.4
$\hat{\phi}_n$												
50	4.5	4.4	4.6	12.7	9.4	4.8	14.0	8.1	6.4	14.4	7.9	6.3
100	4.6	5.0	5.4	26.8	10.9	7.8	27.8	16.5	9.9	28.4	14.4	8.5
200	4.4	4.7	4.1	48.1	20.9	11.7	57.0	34.6	18.2	56.9	30.5	15.0
$\hat{T}_n$												
50	4.7	4.9	6.6	15.8	9.0	7.6	15.0	13.7	7.4	13.2	10.3	9.3
100	3.8	4.0	6.2	32.1	21.6	12.1	31.9	29.0	18.7	29.5	21.4	18.8
200	4.9	5.1	4.2	57.7	40.4	28.3	62.4	55.3	45.5	56.8	41.5	33.6

Table 2. *Percentage of times H_0 was rejected, 5% of significance ($k_1 = 0$, $\lambda = 0.25$ and $\tau = 1$)*

Model	H_0			$H_1 : a$			$H_1 : c$			$H_1 : d$		
k_1	1	2	3	1	2	3	1	2	3	1	2	3
$\hat{\varphi}_{n0.02}$												
50	2.7	3.4	2.3	18.3	20.5	20.5	26.8	41.6	54.1	25.2	30.6	34.1
100	3.8	4.1	3.1	47.2	59.2	69.7	66.9	92.7	98.5	63.6	85.9	94.6
200	3.9	3.2	4.0	84.1	96.3	98.9	97.2	100	100	97.1	100	100
$\hat{\phi}_n$												
50	4.4	4.6	5.2	21.3	17.7	16.4	22.9	23.4	22.9	18.8	18.4	16.1
100	5.0	5.4	4.3	41.6	40.5	39.6	55.4	61.6	56.7	45.8	42.3	35.8
200	4.7	4.1	5.9	76.3	83.2	81.4	93.8	96.2	94.7	86.2	84.2	76.7
$\hat{T}_n$												
50	4.5	4.8	5.7	18.2	20.2	22.7	22.0	48.4	27.2	15.8	42.7	19.8
100	4.2	4.9	4.7	44.8	55.3	36.5	67.0	61.5	42.8	48.8	54.8	39.6
200	4.9	4.8	4.5	71.1	94.0	53.5	97.2	97.7	53.6	89.0	89.8	52.2

Table 3. *Percentage of times H_0 was rejected, 5% of significance ($k_1 = 1$, $\lambda = 0.25$ and $\tau = 1$)*

λ	0.25						0.5					
ρ	1	2	3	4	5	15	1	2	3	4	5	15
$\hat{\varphi}_{n0.02}$												
50	3.8	4.4	5.3	6.6	7.5	11.8	4.1	5.4	8.3	11.6	16.6	35.5
100	4.0	4.7	6.1	7.8	10.6	25.2	3.9	6.2	11.7	21.5	33.6	76.2
200	3.9	4.3	6.4	11.2	18.3	56.3	4.1	6.5	19.0	39.7	61.5	98.7
$\hat{\phi}_n$												
50	5.6	5.4	5.8	6.0	6.5	7.7	5.6	5.7	6.6	8.6	10.9	19.2
100	4.9	5.6	6.7	7.7	9.0	13.7	5.2	7.2	10.0	14.4	20.8	49.7
200	4.3	4.7	6.7	8.6	12.0	24.8	4.6	6.8	13.9	25.4	40.0	87.1

Table 4. *Percentage of times H_0 was rejected, 5% of significance ($k_1 = 0$, $k_2 = 1$ and $\tau = 1$)*

λ	0.25				0.5			
L	1	2	3	4	1	2	3	4
$\hat{\varphi}_{n0.02}$								
50	11.8	7.2	4.4	2.9	35.5	16.4	6.1	2.9
100	25.2	11.8	6.3	4.7	76.2	38.4	11.7	4.9
200	56.3	24.9	7.3	3.9	98.7	77.9	19.6	6.2
$\hat{\phi}_n$								
50	7.7	5.3	6.2	6.1	19.2	8.2	6.0	5.9
100	13.7	6.0	5.6	6.2	49.7	10.5	5.7	6.1
200	24.8	6.3	4.3	4.2	87.1	18.7	5.4	4.6

Table 5. *Percentage of times H_0 was rejected, 5% of significance ($k_1 = 0$, $k_2 = 1$, $\rho = 15$ and $\tau = 1$)*

ρ	1			2			3			15		
k_1	1	2	3	1	2	3	1	2	3	1	2	3
$\hat{\varphi}_{n0.02}$												
50	3.6	3.3	2.8	4.2	3.7	3.8	4.8	4.4	5.4	8.5	13.8	16.7
100	3.7	5.4	3.0	4.8	5.8	5.4	6.1	8.8	11.6	19.8	38.8	51.6
200	3.8	3.8	4.9	4.9	6.7	10.4	8.1	16.6	26.7	48.3	84.1	94.0
$\hat{\phi}_n$												
50	4.5	6.1	7.0	4.6	6.5	6.6	4.9	7.0	7.8	7.8	9.4	10.3
100	5.3	7.1	4.8	6.4	6.9	5.6	7.1	8.7	7.3	12.1	15.5	11.2
200	4.5	5.9	5.1	4.9	7.1	6.8	8.6	9.0	10.6	21.2	34.1	27.5

Table 6. *Percentage of times H_0 was rejected, 5% of significance ($k_2 = 0$, $\lambda = 0.5$ and $\tau = 1$)*

L	1	2	3
$\hat{\varphi}_{n0.02}$			
50	13.8	7.0	4.6
100	38.8	17.3	6.9
200	84.1	47.4	10.4
$\hat{\phi}_n$			
50	9.4	7.7	8.7
100	15.5	8.5	6.3
200	34.1	14.5	7.5

Table 7. *Percentage of times H_0 was rejected, 5% of significance ($k_1 = 2$, $k_2 = 0$, $\lambda = 0.5$ $\rho = 15$ and $\tau = 1$)*

Test	$H_1 : Var(\beta_{00}(IQ)) > 0$ or $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) = 0$ and $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) > 0$ and $Var(\beta_{01}(IQ)) = 0$
$\hat{\varphi}_{n0.003}$	0.012	0.017	0.08
$\hat{\phi}_n$		0.734	
$\hat{T}_n$	0.041	0.009	0.009

Table 8. *p-value of testing H_0 versus H_1 and H_2*

Test	$H_1 : Var(\beta_{00}(IQ)) > 0$ or $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) = 0$ and $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) > 0$ and $Var(\beta_{01}(IQ)) = 0$
$\hat{\varphi}_{n0.003}$	0.6489	0.405	0.484
$\hat{\phi}_n$	0.491	0.653	0.543

Table 9. *p-value of testing H_0 versus H_1 and H_2*