Testing Constancy in Varying Coefficient Models*

Miguel A. Delgado† & Luis A. Arteaga-Molina
Universidad Carlos III de Madrid

Abstract

This article proposes a coefficient constancy test in semi-varying coefficient models, which only needs to estimate the restricted coefficients under the null hypothesis. The test statistic resembles the union-intersection test after ordering the data according to the varying coefficients' explanatory variable. This statistic depends on a trimming parameter that can be chosen by the data-driven calibration method we propose. A bootstrap test is justified under fairly general regularity conditions. Under more restrictive assumptions, the critical values can be tabulated, and trimming is unnecessary. The proposed test can be applied to specification testing of partial effects in the direction of non(semi)-parametric alternatives. The finite sample performance is studied by means of Monte Carlo experiments, and a real data application for modelling education returns.

Keywords: Varying coefficient models; Model checks; UI tests; Concomitants; Partial effects model checks; Wild bootstrap; Trimming data-driven calibration.

JEL Codes: C12, C14, C52.

*We thank the Editor, Associate Editor and two referees for valuable comments and suggestions that improved the quality of the article. Research funded by the Ministerio de Economía y Competitividad (Spain), ECO2017-86675-P & MDM 2014-0431, and by the Comunidad de Madrid (Spain), MadEco-CM S2015/HUM-3444.

†Corresponding author: e-mail: delgado@est-econ.uc3m.es.

1. INTRODUCTION

This article proposes a coefficient constancy test for semi-varying coefficient (SVC) models, where some partial effects are constant and others are non(semi)-parametric functions of an explanatory variable. Existing tests are based on the discrepancy between the restricted and unrestricted sum of squared residuals using smooth estimates of the varying coefficients. See Kauermann and Tutz (1999), Cai et al. (2000), Fan and Zhang (2000), Fan et al. (2001), Li et al. (2002), Fan and Huang (2005), and Cai et al. (2017) among others. In this paper, we propose using a union-intersection (UI) test statistic based on the concomitants (induced order statistics) of the varying coefficients' expanatory variable. Therefore, there is no need to estimate the possibly discontinuous unrestricted varying coefficients under the alternative hypothesis.

The classical CUSUM of residuals in time series models forms a basis for testing the stationarity of the errors, while the CUSUM of concomitant residuals forms a basis for testing that regressors and regression errors are independent in mean. Likewise, the UI tests were proposed for parameter stability testing in time series, and we propose using the UI tests for coefficient constancy testing in varying coefficient models. While the UI tests in time series are based on partial sums of sequential observations, our UI test is based on partial sums of concomitants.

The test can be carried out with parametric, semiparametric, or non-parametric alternatives in mind. In particular, it can be applied to specification testing of partial effects in the direction of semiparametric alternatives, which complements existing specification tests for the regression model. Our test is omnibus, for the constant coefficients' hypothesis, when the varying coefficients' variable is independent in mean of the explanatory variables' cross-products, and also in pure varying coefficient models with no constant coefficients. A bootstrap-assisted test is justified under fairly general conditions. The test statistic depends on a trimming parameter, like other UI tests, to avoid observations close to the boundary of the varying coefficients' support. Such parameter should be chosen small in order to detect as many alternatives as possible. We provide a data-driven trimming calibration method for choosing the smallest amount of trimming that minimizes the error level of the test. Under restrictive assumptions, trimming can be avoided, and the critical values can be tabulated. Under these restrictions, we propose a Neyman-type test and a functional likelihood ratio (LR) test, optimal under local alternatives.

There is a large body of literature that supports using SVC models in economics. The partly linear regression (PLR) model has been proven useful for identifying partial effects in models with unobserved explanatory variables by means of proxy variables. For instance, Olley and Pakes (1996) applied the PLR model to identify output elasticities in a Cobb-Douglas production function specification, where investment is used as a proxy variable of unobserved productivity. See, for instance, Levinsohn and Petrin (2003), Wooldridge (2009b) or Lee et al. (2019) for further developments. Frölich (2008) provides a detailed discussion on using the PLR model to overcome problems with endogenous variables. The test we propose can be used to test a linear regression specification in the direction of a PLR model. When partial effects are expected to vary according to some control variable, the SVC model provides a flexible way of modelling partial effects. For instance, production function models with output elasticities depending on intermediate production and management expenses, e.g. Li et al., 2002. Wang and Xia (2009) and Fan and Huang (2005) use US district data on the Boston area to study the relation between house prices and different explanatory variables, with varying coefficients depending on population lower income status. Chou et al. (2004) proposed a model where the varying coefficients depend on age in a model for health insurance and savings over the life cycle. In all these applications, a significance test for the varying coefficient control variable is well motivated. In particular, the test we propose can be used to test the parametric specification of partial effects in the direction of non(semi)-parametric alternatives.

The rest of the article is organized as follows. The next section presents the testing problem. Section 3 introduces the test statistic, justifies the validity of the test under regularity conditions, and discusses the data-driven calibration algorithm for trimming choice. Section 4 investigates the finite sample properties of the test by means of Monte Carlo experiments. Section 5 reports an application of our proposal for modelling education returns controlling for unobserved individual ability using IQ as proxy variable. Conclusions and final remarks are in Section 6. Mathematical proofs are gathered in an appendix at the end of the article.

2. TESTING PROBLEM

Assume that the random variable Y and the $\mathbb{R}^{1+k_1+k_2}-$valued random vector of explanatory variables $\boldsymbol{W}=(Z,\boldsymbol{X}_1^{\mathrm{T}},\boldsymbol{X}_2^{\mathrm{T}})^{\mathrm{T}}$ are related according to the SVC model

$$Y=\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\beta}_0(Z)+\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{\delta}_0+U, \tag{1}$$

where "T" means transpose, U is an unobserved error term such that $\mathbb{E}(U|\,\boldsymbol{W})=0$ a.s., $\boldsymbol{X}_j=\left(X_{j1},....,X_{jk_j}\right)^{\mathrm{T}}$ is a $k_j\times 1$ random vector, $j=1,2$, with either $X_{11}=1$ or $X_{21}=1$ to allow for a varying or constant intercept term. The varying coefficient vector $\boldsymbol{\beta}_0=\left(\beta_{01},...,\beta_{0k_1}\right)^{\mathrm{T}}:\mathbb{R}\rightarrow\mathbb{R}^{k_1}$ consists of possibly non-smooth functions, and $\boldsymbol{\delta}_0=(\delta_{01},...,\delta_{0k_2})^{\mathrm{T}}$ is a $k_2\times 1$ vector of unknown parameters. The null hypothesis is

$$H_0:\boldsymbol{\beta}_0(Z)=\bar{\boldsymbol{\beta}}_0 \text{ a.s.},$$

where $\bar{\boldsymbol{\beta}}_0:=\mathbb{E}(\boldsymbol{\beta}_0(Z))=\left(\bar{\beta}_{01},...,\bar{\beta}_{0k_1}\right)^{\mathrm{T}}$ a.s., which can be equivalently expressed as $H_0:Var\left(\beta_{0j}(Z)\right)=\mathbb{E}\left(\beta_{0j}(Z)-\bar{\beta}_{0j}\right)^2=0$ for all $j=1,...,k_1$.

This hypothesis nests the case $k_2=0$, $k_1=1$ with $X_{11}=1$, i.e. when $Y=\boldsymbol{\beta}_0(Z)+U$, a pure non-parametric model. In this case, Y and Z are independent in mean under H_0, which can be tested using the Bhattacharya's (1974) CUSUM of residual concomitants test, related to our proposal.

Model (1) also nests the model with

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

where $\delta_{0j}\in\mathbb{R}^{m_j}$ are unknown parameter vectors, and $g_j:\mathbb{R}\rightarrow\mathbb{R}^{m_j}$ is a known vector of functions, $j=1,...,k_1$. In this case, (1) can be expressed as

$$\mathbb{E}(Y|\,Z,\boldsymbol{X}_1)=\boldsymbol{X}_1^{\mathrm{T}}\left[\boldsymbol{\beta}_0(Z)+\boldsymbol{r}_{\boldsymbol{\delta}_0}(Z)\right] \text{ a.s.} \tag{2}$$

with non-parametric $\boldsymbol{\beta}_0$ and parametric $\boldsymbol{r}_{\boldsymbol{\delta}_0}(\cdot)=\left(g_1^{\mathrm{T}}(\cdot)\delta_{01},...,g_{k_1}^{\mathrm{T}}(\cdot)\delta_{0k_1}\right)^{\mathrm{T}}$, for some $\boldsymbol{\delta}_0=\left(\delta_{01}^{\mathrm{T}},...,\delta_{0k_1}^{\mathrm{T}}\right)^{\mathrm{T}}\in\mathbb{R}^{k_1}$. Therefore, assuming (2), H_0 is equivalent to checking that $\mathbb{E}(Y|\,Z,\boldsymbol{X}_1)=\boldsymbol{X}_1^{\mathrm{T}}\left[\bar{\boldsymbol{\beta}}_0+\boldsymbol{r}_{\boldsymbol{\delta}_0}(Z)\right]$ a.s. for some $\left(\bar{\boldsymbol{\beta}}_0^{\mathrm{T}},\boldsymbol{\delta}_0^{\mathrm{T}}\right)^{\mathrm{T}}\in\mathbb{R}^{2k_1}$, in the direction (2) for non-parametric $\boldsymbol{\beta}_0$, where Z can be some component of $\boldsymbol{X}_1$. When Z is not a component of $\boldsymbol{X}_1$, H_0 specifies a parametric model for the partial effects of $\boldsymbol{X}_1$, $\bar{\boldsymbol{\beta}}_0+\boldsymbol{r}_{\boldsymbol{\delta}_0}(Z)$.

Define $\boldsymbol{S}(u) = \left(S_1^{\mathrm{T}}(u), \left[S_1(1) - S_1(u)\right]^{\mathrm{T}}, S_2^{\mathrm{T}}(1)\right)^{\mathrm{T}}$

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

where $S_j(u) = \mathbb{E}\left(\boldsymbol{X}_j Y 1_{\{F_Z(Z) \leq u\}}\right)$, $M_{\ell j}(u) = \mathbb{E}\left(\boldsymbol{X}_\ell \boldsymbol{X}_j^{\mathrm{T}} 1_{\{F_Z(Z) \leq u\}}\right)$, $j, \ell = 1, 2$, and F_Z is the cumulative distribution function (CDF) of Z. Henceforth, $\mathbf{0}$ is a matrix of zeroes of a dimension given by the context. Assume,

A1: F_Z is continuous.

A2: $\mathit{Rank}\,(\boldsymbol{M}(u)) = 2k_1 + k_2$ for each $u \in (0, 1)$.

Our test is based on comparing the vector of functions $\boldsymbol{b}_0^+$ and $\boldsymbol{b}_0^-$, where $\boldsymbol{\theta}_0(u) = \left(\boldsymbol{b}_0^{-\mathrm{T}}(u), \boldsymbol{b}_0^{+\mathrm{T}}(u), \boldsymbol{d}_0^{\mathrm{T}}(u)\right)^{\mathrm{T}}$, and

$$\begin{aligned} \boldsymbol{\theta}_0(u) = \underset{\boldsymbol{b}^-, \boldsymbol{b}^+, \boldsymbol{d}}{\arg\min} \Bigg\{ & \mathbb{E}\left[\left(Y - \boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{b}^- - \boldsymbol{X}_2^{\mathrm{T}} \boldsymbol{d}\right)^2 1_{\{F_Z(Z) \leq u\}}\right] \\ & + \mathbb{E}\left[\left(Y - \boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{b}^+ - \boldsymbol{X}_2^{\mathrm{T}} \boldsymbol{d}\right)^2 1_{\{F_Z(Z) > u\}}\right] \Bigg\} \\ = \boldsymbol{M}^{-1}(u) \boldsymbol{S}(u), \ u \in (0, 1). \end{aligned} \quad (3)$$

The test statistic is a functional of the sample version of

$$\boldsymbol{\eta}_0(u) = \left(\boldsymbol{b}_0^- - \boldsymbol{b}_0^+\right)(u) = \boldsymbol{R}\boldsymbol{M}^{-1}(u)\boldsymbol{S}(u),$$

with $\boldsymbol{R} = \left[\boldsymbol{I}_{k_1} \vdots -\boldsymbol{I}_{k_1} \vdots \, \mathbf{0}\right]$, and $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, which detects any alternative to H_0 of the form,

$$H_{1\eta} : \boldsymbol{\eta}_0(u) \neq \mathbf{0} \text{ for some } u \in (0, 1).$$

We can express $\boldsymbol{b}_0^{\pm}$ in terms of $\boldsymbol{\beta}_0$, as

$$\boldsymbol{\theta}_0(u) = \boldsymbol{M}^{-1}(u) \cdot \mathbb{E}\left(\boldsymbol{m}(u)\left(\boldsymbol{\beta}_0^{\mathrm{T}}(Z), \boldsymbol{\beta}_0^{\mathrm{T}}(Z), \boldsymbol{\delta}_0^{\mathrm{T}}\right)^{\mathrm{T}}\right), \quad (4)$$

with $\boldsymbol{M}(u)=\mathbb{E}\left(\boldsymbol{m}\left(u\right)\right)$. Under H_0,

$$\begin{aligned}
&\mathbb{E}\left[\left(Y-\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{b}^{-}-\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{d}\right)^2 1_{\{F_Z(Z)\le u\}}\right]+\mathbb{E}\left[\left(Y-\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{b}^{+}-\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{d}\right)^2 1_{\{F_Z(Z)>u\}}\right] \qquad (5)\\
&=\mathbb{E}\left(U^2\right)+\ \mathbb{E}\left(\left[\left(\boldsymbol{X}_1^{\mathrm{T}}\left(\bar{\boldsymbol{\beta}}_0-\boldsymbol{b}^{-}\right)+\boldsymbol{X}_2^{\mathrm{T}}\left(\boldsymbol{\delta}_0-\boldsymbol{d}\right)\right)1_{\{F_Z(Z)\le u\}}\right]^2\right)\\
&\quad+\mathbb{E}\left(\left[\left(\boldsymbol{X}_1^{\mathrm{T}}\left(\bar{\boldsymbol{\beta}}_0-\boldsymbol{b}^{+}\right)+\boldsymbol{X}_2^{\mathrm{T}}\left(\boldsymbol{\delta}_0-\boldsymbol{d}\right)\right)1_{\{F_Z(Z)>u\}}\right]^2\right)\\
&\ge\mathbb{E}\left(U^2\right)\ \text{ for all }\ \left(\boldsymbol{b}^{-\mathrm{T}},\boldsymbol{b}^{+\mathrm{T}},\boldsymbol{d}^{\mathrm{T}}\right)\in\mathbb{R}^{2k_1+k_2}\ \text{and all}\ u\in(0,1),
\end{aligned}$$

i.e. under H_0, $\boldsymbol{b}_0^{\pm}(u)=\bar{\boldsymbol{\beta}}_0$ for all $u\in(0,1)$. Hence, from either (4) or (5),

$$H_{0\boldsymbol{\eta}}:\boldsymbol{\eta}_0(u)=\mathbf{0}\ \text{for all}\ u\in(0,1),$$

is a necessary condition for H_0. But $H_{0\boldsymbol{\eta}}$ is also sufficient in many situations, as we show in the remarks below.

Remark 1 *Suppose $M_{1j}(u)=uM_{1j}(1)$ for all $u\in(0,1)$, $j=1,2$. This is equivalent to assume that $\mathbb{E}\left(\boldsymbol{X}_1\boldsymbol{X}_j^{\mathrm{T}}\middle|\,Z\right)=M_{11}(1)$ a.s., $j=1,2$. Therefore,*

$$S_1(u)=uM_{11}(1)\,\mathbb{E}\left(\boldsymbol{\beta}_0(Z)\,1_{\{F_Z(Z)\le u\}}\right)+uM_{12}(1)\boldsymbol{\delta}_0.$$

Reasoning as in Andrews (1993) Lemma A.5, define $\boldsymbol{\upsilon}=(\upsilon_1^{\mathrm{T}},\upsilon_2^{\mathrm{T}},\upsilon_3^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{M}^{-1}(u)\,\boldsymbol{S}(u)$. Then, $\boldsymbol{M}(u)\,\boldsymbol{\upsilon}=\boldsymbol{S}(u)$, and

$$\begin{bmatrix} uM_{11}(1) & 0\\ 0 & (1-u)\,M_{11}(1)\end{bmatrix}\begin{pmatrix}\upsilon_1\\ \upsilon_2\end{pmatrix}=\begin{bmatrix} S_1(u)\\ S_1(1)-S_1(u)\end{bmatrix}-\begin{bmatrix} uM_{12}(1)\upsilon_3\\ (1-u)\,M_{12}(1)\upsilon_3\end{bmatrix}.$$

Therefore,

$$\begin{pmatrix}\upsilon_1\\ \upsilon_2\end{pmatrix}=\begin{bmatrix} M_{11}^{-1}(1)S_1(u)/\,u\\ M_{11}^{-1}(1)\left[S_1(1)-S_1(u)\right]/\,(1-u)\end{bmatrix}-\begin{bmatrix} M_{11}^{-1}(1)M_{12}(1)\upsilon_3\\ M_{11}^{-1}(1)M_{12}(1)\upsilon_3\end{bmatrix},$$

and

$$\begin{aligned}
\boldsymbol{\eta}_0(u) &= \boldsymbol{R}\boldsymbol{M}^{-1}(u)\,\boldsymbol{S}(u)\\
&= \upsilon_1-\upsilon_2\\
&= M_{11}^{-1}(1)\frac{S_1(u)-uS_1(1)}{u(1-u)}\\
&= \frac{1}{u(1-u)}\int_{-\infty}^{F_Z^{-1}(u)}\left(\boldsymbol{\beta}_0(z)-\bar{\boldsymbol{\beta}}_0\right)F_z(dz).
\end{aligned}$$

Hence, $\boldsymbol{\eta}_0(u)=0$ for all $u\in(0,1)$ iff $\boldsymbol{\beta}_0(Z)=\bar{\boldsymbol{\beta}}_0$ a.s.

Remark 2 *Suppose $\boldsymbol{\delta}_0 = \mathbf{0}$, i.e., (1) is a pure varying coefficient model. Then, for all $u \in (0,1)$,*

$$\begin{aligned}\boldsymbol{\eta}_0(u) &= M_{11}^{-1}(u)S_1(u) - [M_{11}(1) - M_{11}(u)]^{-1}[S_1(1) - S_1(u)] \\ &= [M_{11}(1) - M_{11}(u)]^{-1} M_{11}(1)M_{11}^{-1}(u)\left[S_1(u) - M_{11}(u)M_{11}^{-1}(1)S_1(1)\right].\end{aligned}$$

Therefore, for all $u \in (0,1)$

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

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

$$\begin{aligned}\eta_0(u) = 0 \text{ all } u \in (0,1) \quad &\Leftrightarrow \quad \boldsymbol{\beta}_0(Z) = \mathbb{E}[J(Z)]^{-1}\mathbb{E}[J(Z)\boldsymbol{\beta}_0(Z)] \text{ a.s.} \\ &\Leftrightarrow \quad \boldsymbol{\beta}_0(Z) = \bar{\boldsymbol{\beta}}_0 \text{ a.s.}\end{aligned}$$

Therefore, when either $\boldsymbol{X}_1\boldsymbol{X}_\ell^{\mathrm{T}}$, $\ell = 1,2$, and Z are independent in mean, or when all parameters are varying ($k_2 = 0$), H_0 and $H_{0\boldsymbol{\eta}}$ are equivalent. That is, under these conditions our test is omnibus for H_0, i.e. able to detect any alternative of the form

$$H_1 : Var\left(\beta_{0j}(Z)\right) > 0 \text{ for some } j = 1, ..., k_1.$$

3. TESTING METHOD

Given $\{Y_i, \boldsymbol{W}_i\}_{i=1}^n$ *i.i.d.* as $(Y, \boldsymbol{W})$, $\boldsymbol{W}_i = (Z_i, \boldsymbol{X}_{1i}^{\mathrm{T}}, \boldsymbol{X}_{2i}^{\mathrm{T}})^{\mathrm{T}}$, denote by $\left\{\boldsymbol{W}_{1[i:n]}\right\}_{i=1}^n$ the $Z-$concomitants of $\{\boldsymbol{W}_i\}_{i=1}^n$, i.e. for a generic data set $\{\zeta_i\}_{i=1}^n$, $\zeta_{[i:n]} = \zeta_j$ iff $Z_{n:i} = Z_j$, where $Z_{n:1} \leq Z_{n:2} \leq ... \leq Z_{n:n}$ are order statistics of $\{Z_i\}_{i=1}^n$.

The sample analogue of (3) is $\hat{\boldsymbol{\theta}}_n(u) = \left(\hat{\boldsymbol{b}}_n^{-\mathrm{T}}(u), \hat{\boldsymbol{b}}_n^{+\mathrm{T}}(u), \hat{\boldsymbol{d}}_n^{\mathrm{T}}(u)\right)^{\mathrm{T}}$, $u \in [K/n, 1-K/n)$, where $K = k_1 + k_2$, and

$$\begin{aligned}\hat{\boldsymbol{\theta}}_n(u) &= \underset{\boldsymbol{b}^-,\boldsymbol{b}^+,\boldsymbol{d}}{\arg\min}\left\{\sum_{i=1}^{\lfloor nu\rfloor}\left(Y_{[i:n]} - \boldsymbol{X}_{1[i:n]}^{\mathrm{T}}\boldsymbol{b}^- - \boldsymbol{X}_{2[i:n]}^{\mathrm{T}}\boldsymbol{d}\right)^2\right. \qquad (6)\\ &\qquad\qquad \left. + \sum_{i=1+\lfloor nu\rfloor}^{n}\left(Y_{[i:n]} - \boldsymbol{X}_{1[i:n]}^{\mathrm{T}}\boldsymbol{b}^+ - \boldsymbol{X}_{2[i:n]}^{\mathrm{T}}\boldsymbol{d}\right)^2\right\} \\ &= \hat{\boldsymbol{M}}_n^{-1}(u)\,\hat{\boldsymbol{S}}_n(u), \;\; u \in [K/n, 1-K/n),\end{aligned}$$

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

$\boldsymbol{M}(u)$, with components $\hat{M}_{n\ell j}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{\ell[i:n]}\boldsymbol{X}_{j[i:n]}^{\mathrm{T}}$ estimating $M_{\ell j}(u)$ in $\boldsymbol{M}(u)$, $\ell, j = 1, 2$. This suggests test statistics for H_0 based on some functional of

$$\begin{aligned} \hat{\boldsymbol{\eta}}_n(u) &= \left(\hat{\boldsymbol{b}}_n^- - \hat{\boldsymbol{b}}_n^+\right)(u) \\ &= \boldsymbol{R}\hat{\boldsymbol{M}}_n^{-1}(u)\,\hat{\boldsymbol{S}}_n(u), \\ &= \boldsymbol{\eta}_0(u) + \boldsymbol{R}\hat{\boldsymbol{M}}_n^{-1}(u)\,\hat{\boldsymbol{N}}_n(u), \;\; u \in [K/n, 1-K/n) . \end{aligned} \tag{7}$$

with $\hat{\boldsymbol{N}}_n(u) = \left(\hat{N}_{n1}^{\mathrm{T}}(u), \hat{N}_{n1}^{\mathrm{T}}(1) - \hat{N}_{n1}^{\mathrm{T}}(u), \hat{N}_{n2}^{\mathrm{T}}(1)\right)^{\mathrm{T}}$, $\hat{N}_{nj}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{j[i:n]}U_{[i:n]}$, $j=1,2$.

The asymptotic distribution of $\sqrt{n}\hat{\boldsymbol{N}}_n$ is obtained applying results for partial sums of concomitants in Bhattacharya (1974, 1976), extended by Sen (1976), Stute (1993, 1997) and Davydov and Egorov (2000), among others.

Henceforth, for any matrix A, $\|A\|^2 = \bar{g}\left(A^{\mathrm{T}}A\right)$ is the spectral norm, where $\bar{g}(C)$ is the maximum eigenvalue of the matrix C, and "$\rightarrow_d$ " means convergence in distribution of random variables, random vectors or random elements in a Skorohov space $\mathbb{D}[a,b]$, $0 \leq a < b \leq 1$. Define $\boldsymbol{N}_\infty(u) = \left(N_{\infty 1}^{\mathrm{T}}(u), N_{\infty 1}^{\mathrm{T}}(1) - N_{\infty 1}^{\mathrm{T}}(u), N_{\infty 2}^{\mathrm{T}}(1)\right)^{\mathrm{T}}$, where $N_{\infty j}$ is a $k_j \times 1$ vector of a centered Gaussian process with $\mathbb{E}\left(N_{\infty\ell}(u)N_{\infty j}^{\mathrm{T}}(v)\right) = \mathbb{E}\left(\boldsymbol{X}_\ell\boldsymbol{X}_j^{\mathrm{T}}U^2 1_{\{F_Z(Z)\leq u\wedge v\}}\right)$, $\ell, j = 1, 2$, and $u, v \in (0,1)$. Assume,

A3: $\mathbb{E}\|\boldsymbol{X}_jU\|^2 < \infty$, $j = 1, 2$.

Theorem 1: *Assuming A1, A2,*

$$\sup_{u\in(0,1)} \left\| \left(\hat{\boldsymbol{M}}_n - \boldsymbol{M}\right)(u) \right\| = o(1) \; a.s. \tag{8}$$

and if A3 is also assumed,

$$\sqrt{n}\hat{\boldsymbol{N}}_n \rightarrow_d \boldsymbol{N}_\infty \text{ in } \mathbb{D}[0,1] . \tag{9}$$

Therefore, using (7) under H_0, and conditions in Theorem 1,

$$\sqrt{n}\hat{\boldsymbol{\eta}}_n \rightarrow_d \boldsymbol{\eta}_\infty \; in \; \mathbb{D}[\epsilon, 1-\epsilon], \text{ for } \epsilon \in (0,1),$$

where $\boldsymbol{\eta}_\infty(u) =_d \boldsymbol{R}\boldsymbol{M}^{-1}(u)\boldsymbol{N}_\infty(u)$, "$=_d$ " means equality in distribution. Weak convergence of $\sqrt{n}\hat{\boldsymbol{\eta}}_n$ in $\mathbb{D}[0,1]$ is not possible, as shown by Chibisov (1964) for the standard empirical process (see subsection 2.5 in Gaenssler and Stute, 1979 for discussion). Thus,

$$\mathbb{E}\left(\boldsymbol{\eta}_\infty(u)\boldsymbol{\eta}_\infty^{\mathrm{T}}(v)\right) = \boldsymbol{\Sigma}_0(u,v) = \boldsymbol{R}\boldsymbol{M}^{-1}(u)\boldsymbol{\Omega}_0(u,v)\boldsymbol{M}^{-1}(v)\boldsymbol{R}^{\mathrm{T}}, \;\; u, v \in (0,1),$$

with $\mathbf{\Omega}_0(u,v) = \mathbb{E}\left(\boldsymbol{N}_\infty(u)\boldsymbol{N}_\infty^{\mathrm{T}}(v)\right)$. In order to apply the UI testing principle, we must standardize $\hat{\boldsymbol{\eta}}_n$. Therefore, we need to estimate,

$$\mathbf{\Omega}(u,u) = \begin{bmatrix} \Omega_{11}(u) & \mathbf{0} & \Omega_{12}(u) \\ \mathbf{0} & \Omega_{11}(1) - \Omega_{11}(u) & \Omega_{12}(1) - \Omega_{12}(u) \\ \Omega_{21}(u) & \Omega_{21}(1) - \Omega_{21}(u) & \Omega_{22}(1) \end{bmatrix},$$

where $\Omega_{\ell j}(u) = \mathbb{E}\left(\boldsymbol{X}_\ell \boldsymbol{X}_j^{\mathrm{T}} V^2 1_{\{F_Z(Z)\le u\}}\right)$, $j = 1, 2$, and $V = Y - \boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{b}_0^+(1) - \boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{d}(1)$ are the errors of the best linear predictor. Under H_0, $\mathbf{\Omega} = \mathbf{\Omega}_0$. Asssume

A4: $Rank\left(\mathbf{\Omega}(u,u)\right) = 2k_1 + k_2$ for all $u \in (0,1)$.

The natural estimator of $\Omega_{\ell j}(u)$ is $\hat{\Omega}_{n\ell j}(u) = n^{-1}\sum_{i=1}^{\lfloor nu\rfloor} \boldsymbol{X}_{\ell[i:n]}\boldsymbol{X}_{j[i:n]}^{\mathrm{T}}\hat{V}_{[i:n]}^2$, $\ell, j = 1, 2,$, where $\hat{V}_i = Y_i - \boldsymbol{X}_{1i}^{\mathrm{T}}\hat{\boldsymbol{b}}_n^+(1) - \boldsymbol{X}_{2i}^{\mathrm{T}}\hat{\boldsymbol{d}}_n(1)$ are the OLS residuals using all the data set.

Tests are based on functionals of the empirical process,

$$\hat{\alpha}_n(u) = \hat{\boldsymbol{\eta}}_n^{\mathrm{T}}(u)\,\hat{\mathbf{\Sigma}}_n^{-1}(u,u)\,\hat{\boldsymbol{\eta}}_n(u),\ \ u \in [K/n, 1 - K/n).$$

where $\hat{\mathbf{\Sigma}}_n(u,u) = \boldsymbol{R}^{\mathrm{T}}\hat{\boldsymbol{M}}_n^{-1}(u)\hat{\mathbf{\Omega}}_n(u,u)\hat{\boldsymbol{M}}_n^{-1}(u)\boldsymbol{R}$, and $\hat{\mathbf{\Omega}}_n(u,u)$ is the estimator of $\mathbf{\Omega}(u,u)$ with components $\hat{\Omega}_{n\ell j}(u)$. A sufficient condition for consistency of $\hat{\mathbf{\Omega}}_n$ is

A5: $\mathbb{E}\left\|\boldsymbol{X}_j\right\|^4 < \infty$, $j = 1, 2$, and $\mathbb{E}\left\|V\right\|^4 < \infty$.

This condition can be relaxed assuming that $\mathbb{E}\left(\boldsymbol{X}_\ell\boldsymbol{X}_j^{\mathrm{T}}V^2 1_{\{F_Z(Z)\le u\}}\right) = \mathbb{E}\left(V^2\right)M_{\ell j}(u)$, $\ell, j = 1, 2$. The test rejects H_0 for large values of

$$\hat{\varphi}_{n\epsilon} = n \max_{K+\lfloor n\epsilon\rfloor \le j < n-K-\lfloor n(1-\epsilon)\rfloor} \hat{\alpha}_n\left(\frac{j}{n}\right),\ \text{for } \epsilon \in (0,1).$$

The trimming parameter ϵ is introduced to rid of data points corresponding to the extreme $Z's$ quantiles. When the alternative is non-parametric, ϵ should be chosen as close to zero as possible in order to detect any possible alternative. However, too small ϵ can produce serious size distortions (see Section 4). The asymptotic distribution of $\hat{\varphi}_{n\epsilon}$ is derived as an immediate consequence of Theorem 1. Define

$$\varphi_{\infty\epsilon} =_d \sup_{u\in[\epsilon,1-\epsilon]} \alpha_\infty(u),$$

where,

$$\left\{\alpha_\infty(u)\right\}_{u\in(0,1)} \overset{d}{=} \left\{\boldsymbol{\eta}_\infty^{\mathrm{T}}(u)\mathbf{\Sigma}_0^{-1}(u,u)\,\boldsymbol{\eta}_\infty(u)\right\}_{u\in(0,1)}.$$

Theorem 2: *Assume A1 − A5. Under H_0, for $\epsilon \in (0,1)$,*

$$\hat{\varphi}_{n\epsilon} \rightarrow_d \varphi_{\infty\epsilon}.$$

Therefore, a test with significance level α 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 $\varphi_{\infty\epsilon}$.

Next, we study the power of the test in the direction of fix and local alternatives,

$$H_{n1} : \boldsymbol{\beta}(Z) = \bar{\boldsymbol{\beta}}_0 + \frac{\boldsymbol{\tau}(Z)}{\sqrt{n}} \text{ a.s.,} \tag{10}$$

for a vector of constants $\bar{\boldsymbol{\beta}}_0$ and an unknown function $\boldsymbol{\tau} : \mathbb{R} \rightarrow \mathbb{R}^{k_1}$ such that, for all $u \in (0,1)$, $T_j(u) = \mathbb{E}\left[\boldsymbol{X}_j \boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{\tau}(Z) 1_{\{F_Z(Z) \leq u\}}\right]$ is bounded, $j = 1,2$. Then, define $\boldsymbol{T}(u) = [T_1^{\mathrm{T}}(u), T_1^{\mathrm{T}}(1) - T_1^{\mathrm{T}}(u), T_2^{\mathrm{T}}(1)]^{\mathrm{T}}$ and the random processes,

$$\left\{\alpha_\infty^1(u)\right\}_{u\in(0,1)} \overset{d}{=} \left\{\boldsymbol{\eta}_\infty^{1\mathrm{T}}(u)(u)\boldsymbol{\Sigma}_0^{-1}(u,u)\,\boldsymbol{\eta}_\infty^1(u)\right\}_{u\in(0,1)},$$

with $\{\boldsymbol{\eta}_\infty^1(u)\}_{u\in(0,1)} =_d \left\{\boldsymbol{R}\boldsymbol{M}^{-1}(u)\left(\boldsymbol{N}_\infty + \boldsymbol{T}\right)(u)\right\}_{u\in(0,1)}$. To study the power of the test under H_{n1}, we need to assume,

A6: $\mathbb{E}\left\|\boldsymbol{X}_j \boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{\tau}(Z)\right\| < \infty$, for $j = 1,2$.

Theorem 3: *Assume A1 − A6, for $\epsilon \in (0,1)$. Under $H_{\eta 1}$,*

$$\hat{\varphi}_{n\epsilon} \rightarrow_p \infty, \tag{11}$$

and under H_{n1},

$$\hat{\varphi}_{n\epsilon} \rightarrow_d \sup_{u\in[\epsilon,1-\epsilon]} \alpha_\infty^1(u). \tag{12}$$

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

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

This suggests choosing ϵ as small as possible in order to detect alternatives with coefficients only varying at $Z's$ extreme values.

Remark 3 *Other functionals of $\hat{\alpha}_n$ can be used to perform the test. In particular, Andrews and Ploberger (1994), Example 1, page 1404, discuss an optimal significance test,* *in Wald's (1943) sense, of the parameter vector* $\boldsymbol{\phi}_0 = \left(\phi_{01}, ..., \phi_{0k_1}\right)^{\mathrm{T}}$ *in a discontinuos regression design (RDD) model with* $\boldsymbol{\beta}_0(z) = \left(\beta_{01}(z), ..., \beta_{0k_1}(z)\right)^{\mathrm{T}}$ *and* $\beta_{0j}(z) = \phi_{0j} 1_{\{z \le \pi_0\}}$, $j = 1, ..., k_1$, *where* π_0 *is a nuisance parameter. They use this example to illustrate sufficient conditions for asymptotic optimal tests when some nuisance parameter is only prensent under the alternative. The hypothesis of interest is* H_0 *in the direction* $H_{n\phi}$: $\boldsymbol{\phi}_0 = \boldsymbol{\kappa}_0 / \sqrt{n}$ *for some* $\kappa_0 \in \mathbb{R}^{k_1}$, *assuming that* U *is independent of* $(\boldsymbol{X}_1, \boldsymbol{X}_2, Z)$. *This approach suggests using as test statistic*

$$\hat{\rho}_{nG_\epsilon}(c) = \frac{n}{(1+c)^{\frac{1+k_1}{2}}} \int_0^1 \exp\left(\frac{c}{2(1+c)} \hat{\alpha}_n(u)\right) dG_\epsilon(u),$$

for optimal testing in Wald's sense, where $G_\epsilon : [\epsilon, 1-\epsilon] \to \mathbb{R}^+$ *is a given weight function and* $c > 0$ *is a scalar constant that depends on the weights, such that the weighted average power is maximum. The statistic*

$$\lim_{c \to 0} \frac{\hat{\rho}_{nG_\epsilon}(c) - 1}{c} = n \int_0^1 \bar{\alpha}_n(u)\, dG_\epsilon(u)$$

is suitable for alternatives $H_{n\phi}$ *close to the null, while* $\hat{\varphi}_{n\epsilon}$ *is designed to detect distant alternatives.*

The distribution of $\varphi_{\infty\epsilon}$ depends on unknown features of the underlying data generating process under general conditions, but can be implemented with the assistance of a bootstrap technique. We use a wild bootstrap resample $\{Y_i^*, \boldsymbol{W}_i\}_{i=1}^n$, with $Y_i^* = \boldsymbol{X}_{1i}^{\mathrm{T}} \hat{\boldsymbol{\beta}}_0(1) + \boldsymbol{X}_{1i}^{\mathrm{T}} \hat{\delta}_0(1) + \hat{V}_i^*$ where, $\hat{V}_i^* = \hat{V}_i \xi_i$, and $\{\xi_i\}_{i=1}^n$ are *i.i.d.* as ξ, which satisfies,

A7: $\mathbb{E}(\xi) = 0$, $\mathbb{E}\left(\xi^2\right) = 1$ and $|\xi| \le C < \infty$ a.s.

The bootstrap test statistic is,

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

where $\hat{\alpha}_n^*(u) = \hat{\eta}_n^{*\mathrm{T}}(u) \hat{\boldsymbol{\Sigma}}_n^{-1}(u,u) \hat{\eta}_n^*(u)$, $\hat{\boldsymbol{\eta}}_n^*(u) = \boldsymbol{R} \hat{\boldsymbol{M}}_n^{-1}(u) \hat{\boldsymbol{N}}_n^*(u)$, $\hat{\boldsymbol{N}}_n^*(u) = \Big(\hat{N}_{n1}^{*\mathrm{T}}(u), \hat{N}_{n1}^{*\mathrm{T}}(1) - \hat{N}_{n1}^{*\mathrm{T}}(u), \hat{N}_{n2}^{*\mathrm{T}}(1)\Big)^{\mathrm{T}}$, and $\hat{N}_{nj}^*(u) = n^{-1} \sum_{i=1}^{\lfloor nu \rfloor} \boldsymbol{X}_{j[i:n]} \hat{V}_{[i:n]}^*$, $j = 1, 2$. The bootstrap critical value at the $\alpha - \textit{level}$ of significance is $\hat{c}_{n\epsilon}^*(\alpha) = \inf\{c : \mathbb{P}_\xi(\hat{\varphi}_{n\epsilon}^* \le c) \ge 1 - \alpha\}$, where $\mathbb{P}_\xi$ is the induced probability function of ξ. So, the bootstrap test is given by the binary variable $\hat{\Phi}_{n\epsilon}^*(\alpha) = 1_{\{\hat{\varphi}_{n\epsilon} > \hat{c}_{n\epsilon}^*(\alpha)\}}$. The next theorem justifies the bootstrap test.

Theorem 4: *Assume A1 − A5, and A7. Under H_0, for any $\epsilon > 0$,*

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

and there exists a $C > 0$ such that, under H_1,

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

This implies that $\lim_{n\to\infty} \mathbb{E}\left[\hat{\Phi}^{*}_{n\epsilon}(\alpha)\right] = \alpha$ under H_0, and $\lim_{n\to\infty} \mathbb{E}\left[\hat{\Phi}^{*}_{n\epsilon}(\alpha)\right] = 1$ under H_1. The test can also be based on the bootstrap $p-values$, $\hat{p}^{*}_{n\epsilon} = \mathbb{P}_{\xi}\left(\hat{\varphi}^{*}_{n\epsilon} \geq \hat{\varphi}_{n\epsilon}\right)$.

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

<u>**Algorithm 1**</u>

i. Generate b sets of random numbers $\left\{\xi^{(j)}_i\right\}_{i=1}^{n}$ *i.i.d.* as ξ, and the corresponding resamples $\{Y^{*(j)}_i, \boldsymbol{W}_i\}_{i=1}^n$, $j = 1, ..., b$, with b large.

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

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

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

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

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

iv. Use the test $\hat{\Phi}^{*(b)}_{n\epsilon}(\alpha) = 1_{\left\{\hat{\varphi}_{n\epsilon} > \hat{c}^{*(b)}_{n\epsilon}(\alpha)\right\}} = 1_{\left\{\hat{p}^{*(b)}_{n\epsilon} < \alpha\right\}}$.

The greater the b, the better the bootstrap approximations.

When the alternative hypothesis is nonparametric, one should choose the smaller possible ϵ in order to detect as many alternatives as possible, but a too small ϵ may produce serious size distorsions. In order to keep the type I error under control, given a nominal level α, we can choose the smallest ϵ that minimizes the actual level error. To this end, we propose a data-driven calibration method, inspired by Politis et. al. (1999) Section 9.4.1. We think of the actual level of the test, ω, as a function of ϵ, i.e. $h : \epsilon \to \omega$. If h were known, we could calculate the actual error level $e(\epsilon) = |h(\epsilon) - \alpha|$. If the underlying joint distribution of $(Y, \boldsymbol{W})$, F, were known, we could simulate samples according to F, and estimate $h(\epsilon)$ as the fraction of times that the corresponding test rejects H_0 for the given ϵ. Since F is unknown, we can use some estimator $\hat{F}_n$ that is consistent for F at least under H_0. A natural choice is the empirical distribution of $\{Y_i, \boldsymbol{W}_i\}_{i=1}^n$, but we could use wild bootstrap resamples that impose H_0 instead, as in Algorithm 1 step **i**.

In order to save computing time and choosing ϵ as small as possible, we fix the maximum error level that we are prepared to bear, e_0, e.g. $e_0 = 10^{-3}$. The following algorithm provides the data-driven calibrated smallest ϵ, $\hat{\epsilon}_n$, that ensures an error level less or equal to e_0 into a given inteval $[1/n, \ell_0/n]$ for suitably chosen small ℓ_0, e.g. $\ell_0 = \lfloor n/3 \rfloor$. Such $\hat{\epsilon}_n$ may not exist, in which case we choose the ϵ minimizing the error level in the interval.

Algorithm 2

i. Fix e_0 and ℓ_0.

ii. Fix b_0 and generate *i.i.d.* resamples $\left\{Y_i^{\dagger(j)}, \boldsymbol{W}_i\right\}_{i=1}^n$ from $\hat{F}_n$, $j = 1, ..., b_0$.

iii. Set $\ell := 1$.

iv. If $\ell = \ell_0 + 1$, compute $\hat{\epsilon}_n := n^{-1} \arg\min_{\ell \in [1,\ell_0]} |\hat{e}_n(\ell/n)|$ and stop.

v. For each resample $\left\{Y_i^{\dagger(j)}, \boldsymbol{W}_i\right\}_{i=1}^n$, compute the corresponding test $\hat{\Phi}^{*(b)}_{n(\ell/n)j}(\alpha)$ using Algorithm 1, $j = 1, ..., b_0$.

vi. Compute $\hat{h}_n(\ell/n) := b_0^{-1} \sum_{j=1}^{b_0} \hat{\Phi}^{*(b)}_{n(\ell/n)j}(\alpha)$ and record the corresponding estimated error level $\hat{e}_n(\ell/n) := \left|\hat{h}_n(\ell/n) - \alpha\right|$.

vii. If $\hat{e}_n(\ell/n) \leq e_0$, $\hat{\epsilon}_n := \ell/n$ and stop. Otherwise, $\ell := \ell + 1$ and go to **iv**.

Then, we use the test $\hat{\Phi}^{*(b)}_{n\hat{\epsilon}_n}(\alpha)$ in Algorithm 1. This data-driven trimming choice is computationally expensive. Of course, we could choose $\hat{\epsilon}_n$ minimizing $\hat{e}_n(\epsilon)$ over the interval, but it will be even more costly, and the resulting $\hat{\epsilon}_n$ will probably be larger. In this respect, the bigger e_0 (ℓ_0), the smaller (bigger) computational cost. A formal justification of the test $\hat{\Phi}^{*(b)}_{n\hat{\epsilon}_n}(\alpha)$ is beyond the scope of this paper, but we show in next section (Table 8) that it works very well in practice.

Under strong regularity conditions below, we can avoid trimming, and critical values of the test can be tabulated. Suppose for simplicity that $k_2 = 0$, i.e. there are no constant coefficients in the model. Assume,

A8: Z is independent of $(U, \boldsymbol{X}_1)$, $\mathbb{E}\left(U^2 | \boldsymbol{X}_1\right) = \sigma^2$, and $M_{11}(1)$ is non-singular.

This assumption is not acceptable in practice, but allows to discuss the relation of our proposal to related ones for time series parameter instability tests, as well as the behaviour of our test statistic when ϵ is too small. Under $A8$, $M_{11}(u) = uM_{11}(1)$, $\Omega_{11}(u) = \sigma^2 \cdot u \cdot M_{11}(1)$, and applying the same arguments as in Remark 1,

$$\boldsymbol{\eta}_\infty(u) = \boldsymbol{R}\boldsymbol{M}^{-1}(u)\boldsymbol{N}_\infty(u) = M_{11}^{-1}(1)\frac{N_{1\infty}(u) - uN_{1\infty}(1)}{u(1-u)} \text{ a.s.,} \tag{13}$$

and $\left\{N_{1\infty}(u)\right\}_{u\in(0,1)} =_d \left\{\sigma \cdot M_{11}^{1/2}(1)\boldsymbol{B}_0(u)\right\}_{u\in(0,1)}$, where $\boldsymbol{B}_0$ is a vector of independent Brownian bridges, i.e. $\boldsymbol{B}_0$ is a Gaussian process with mean zero and $\mathbb{E}\left(\boldsymbol{B}_0(u)\boldsymbol{B}_0^{\mathrm{T}}(u)\right) = (u \wedge v - uv) \cdot \boldsymbol{I}_{k_1}$, for all $u, v \in (0,1)$. Henceforth,

$$\varphi_{\epsilon\infty} \overset{d}{=} \sup_{u\in[\epsilon,1-\epsilon]} \frac{\boldsymbol{B}_0^{\mathrm{T}}(u)\boldsymbol{B}_0(u)}{u(1-u)},$$

which has been tabulated by James et al. (1987) for $\boldsymbol{B}_0$ scalar and different values of ϵ, and by Andrews (1993) in the multivariate case.

Under $A8$, one can exploit the information in (13) and, after estimating σ^2 by $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \hat{V}_{ni}^2$, use as test statistic,

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

with

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

which resembles the classical UI tests, but without trimming. This statistic, suitably standardized, converges to a extremum value distribution applying Darling and Erdős (1956) type results. To this end, we need an alternative condition replacing $A3$ by,

A9: $\mathbb{E}\left\|\boldsymbol{X}_1\right\|^{2+\delta} < \infty$ and $\mathbb{E}\left|U\right|^{2+\delta} < \infty$ for some $\delta > 0$.

This is stronger than $A3$. These types of moment conditions were proposed by Shorak (1979), relaxing those in Darling and Erdős (1956). Henceforth, $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$, and E is a random variable such that $\mathbb{P}(E \le x) = \exp(-2\exp(-x))$, $a(x) = \sqrt{2\log x}$ 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 level accuracy. We also consider,

$$\begin{aligned}\tilde{\varphi}_n^{(1)} &= \sum_{j=K}^{n-K-1} \tilde{\alpha}_n\left(\frac{j}{n}\right),\\ \tilde{\varphi}_n^{(2)} &= \max_{K\le j<n-K} \frac{j(n-j)}{n}\,\tilde{\alpha}_n\left(\frac{j}{n}\right).\end{aligned}$$

The next theorem, provides the limiting distribution of $\tilde{\varphi}_n^{(j)}$, $j = 0, 1, 2$ under H_0.

Theorem 5: *Assume A1, A2, A8 and A9, under* H_0,

$$a(\log n)\sqrt{\tilde{\varphi}_n^{(0)}} - b_{1+k_1}(\log n) \xrightarrow{d} E, \tag{14}$$

$$\tilde{\varphi}_n^{(1)} \xrightarrow{d} \int_0^1 \frac{\boldsymbol{B}_0^{\mathrm{T}}(u)\boldsymbol{B}_0(u)}{u(1-u)} du, \tag{15}$$

$$\tilde{\varphi}_n^{(2)} \xrightarrow{d} \sup_{u\in(0,1)} \boldsymbol{B}_0^{\mathrm{T}}(u)\boldsymbol{B}_0(u). \tag{16}$$

This suggests that the asymptotic distribution of $\hat{\varphi}_{n\epsilon}$ changes at $\epsilon = 0$. Tests based on critical values of the asymptotic approximation (14) are expected to exhibit poor level accuracy. See simulations in Section 5. The critical values of the random variables on the right hand side of (15) and (16) have been tabulated by Scholz and Stephens (1987) and Kiefer (1959), respectively.

Remark 4 *Under A8 we can construct Neyman smooth and optimal functional LR tests, in the direction of local alternatives (10), based on the principal components of the* $\hat{\boldsymbol{\eta}}_n'$*s transformation,*

$$\tilde{\boldsymbol{\eta}}_n(u) = \frac{u(1-u)}{\hat{\sigma}} \hat{M}_{n11}^{1/2}(1)\hat{\boldsymbol{\eta}}_n(u), \ u \in [K/n, 1-K/n).$$

Thus, by (12), (25) and (26), under H_{n1}, $\sqrt{n}\tilde{\boldsymbol{\eta}}_n \to_d \tilde{\boldsymbol{\eta}}_\infty =_d \boldsymbol{B}_0 + \boldsymbol{\omega}_0$ *with* $\boldsymbol{\omega}_0(u) = \mathbb{E}\left[\boldsymbol{\tau}(Z)1_{\{u<Z<1-u\}}\right]$. *The principal components random vectors of* $\boldsymbol{B}_0$ *are*

$$\boldsymbol{\zeta}_j = \frac{1}{\sqrt{\vartheta_j}} \int_0^1 \boldsymbol{B}_0(u)\mu_j(u)du,$$

where $\vartheta_j = (j\pi)^{-2}$ *and* $\mu_j(u) = \sqrt{2}\sin(j\pi u)$, $j \in \mathbb{N}$, *are the eigenvalues and (orthonormal) eigenfunctions of the Brownian Bridge covariance kernel* $\Upsilon(u,v) = u \wedge v - uv$. *Therefore,* $\{\boldsymbol{\zeta}_j\}_{j\in\mathbb{N}}$ *are i.i.d.* $N_{k_1}(0, \boldsymbol{I}_{k_1})$. *The sample version of* $\boldsymbol{\zeta}_j$ *is*

$$\hat{\boldsymbol{\zeta}}_{nj} = \sqrt{\frac{n}{\vartheta_j}} \int_0^1 \tilde{\boldsymbol{\eta}}_n(u)\mu_j(u)du.$$

The distribution of the infinite dimensional random vectors $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_j)_{j>0}$ *and* $\hat{\boldsymbol{\zeta}}_n = \left(\hat{\boldsymbol{\zeta}}_{nj}\right)_{j>0}$ *are uniquely determined by their finite dimensional distributions. Thus, by the continuous mapping theorem,* $\hat{\boldsymbol{\zeta}}_n \to_d \boldsymbol{\zeta}$ *under* H_0, *and* $\hat{\boldsymbol{\zeta}}_n \to_d \boldsymbol{\zeta} + \boldsymbol{\rho}_0$ *under* H_{n1}, *where* $\boldsymbol{\rho}_0 = (\boldsymbol{\rho}_{0j})_{j>0}$ *with* $\boldsymbol{\rho}_{0j} = \vartheta_j^{-1/2}\int_0^1 \boldsymbol{\omega}_0(u)\mu_j(u)du$. *This suggests a Neyman-type test that rejects* H_0 *for large values of*

$$\hat{Q}_{nm} = \sum_{j=1}^{m} \hat{\boldsymbol{\zeta}}_{nj}^{\mathrm{T}}\hat{\boldsymbol{\zeta}}_{nj},$$

for a fixed m*. Thus, under* H_0, $\hat{Q}_{nm} \to_d \chi^2_{mk_1}$ *and under* H_{n1}, $\hat{Q}_{nm} \to_d \chi^2_{mk_1}\left(\sum_{j=1}^m \boldsymbol{\rho}_{0j}^{\mathrm{T}} \boldsymbol{\rho}_{0j}\right)$, *where* $\chi^2_\ell(\Lambda)$ *denotes a non-centered chi-square with* $\ell-$*degrees of freedom and non-centrality parameter* Λ*. Tests based on a single principal component were proposed by Durbin and Knott (1972) in the classical goodness-of-fit (GOF) tests, which was extended by Schoenfeld (1977) to linear combinations of principal components. Stute (1997) applied this test to specification testing of regression models, which in turns has been extended in different directions. The functional Neyman-Pearson LR test, introduced by Grenander (1950) for the classical GOF problem can also be applied in our context. This has been applied by Sowell (1996) for optimal parameter instability testing in time series using the CUSUM of residuals, and by Stute (1997) for optimal regression specification testing using the CUSUM of residuals concomitants. This approach has been further extended to other contexts by Delgado et al. (2005), and Delgado and Stute (2008), among others. Suppose, for notational convenience, that* $k_1 = 1$*. The optimal functional LR test, in the direction (10), consists of rejecting* H_0 *in favour of* H_{1n} *at the* α *significance level, when*

$$G = \frac{1}{\sqrt{\sum_{j=1}^\infty \vartheta_j^{-1} \boldsymbol{\rho}_{0j}^2}} \sum_{j=1}^\infty \vartheta_j^{-1} \boldsymbol{\rho}_{0j} \int_0^1 \tilde{\boldsymbol{\eta}}_\infty(u) \mu_j(u) du \geq z_{1-\alpha},$$

with $z_{1-\alpha}$ *the* $(1-\alpha)-th$ *quantile of the standard normal. The feasible test statistic is, for large* m,

$$\hat{G}_{nm} = \frac{1}{\sqrt{\sum_{j=1}^m \vartheta_j^{-1} \boldsymbol{\rho}_{0j}^2}} \sum_{j=1}^m \vartheta_j^{-1} \boldsymbol{\rho}_{0j} \int_0^1 \tilde{\boldsymbol{\eta}}_n(u)\, \mu_j(u)\, du.$$

4. FINITE SAMPLE PROPERTIES

We generate samples $\{Y_i, Z_i, 1, X_{12i}, ..., X_{1k_1i}, X_{21i}, ..., X_{2k_2i}\}_{i=1}^n$ with

$$Y_i = \beta_{01}(Z_i) + \sum_{j=2}^{k_1} \beta_{0j}(Z_i) X_{1ji} + \sum_{j=1}^{k_2} \delta_{0j} X_{2ji} + U_i,\ i = 1, ..., n,$$

with $\{Z_i\}_{i=1}^n$ *i.i.d.* uniform in $[0,1]$, $X_{\ell ji} = Z_i + e_{\ell ji}$, with $e_{\ell ji}'s$ *i.i.d.* uniformly in $[0,1]$, $j = 1, 2$, and

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

with ε_i *i.i.d.* $N(0,1)$; that is, $Var(U_i) = 1$, and κ governs how severe the conditional heteroskedasticity is. We generate random coefficients

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

for all $j = 1, ..., k_1$, i.e. $Var(\beta_{0j}(Z)) = \lambda^2$, with the following models,

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

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 the ρ, the smaller the departure from linearity is. Model c) is harder to fit than a) or b) using smooth methods with moderate sample sizes, and d) is a RDD model that cannot be estimated using smoothing methods. We set $\pi_0 = 0.4$, but we have also tried other values and the results do not change substantially except when the jump is placed in extreme low quantiles ($\pi_0 \leq 0.1$). Figure 1 represents $\boldsymbol{\eta}_0$ for the different models and different λ values.

FIGURE 1 ABOUT HERE

The simulation study is implemented to provide evidence on the ϵ's choice effect, the accuracy of the bootstrap test, and the relative performance of our test with respect to existing alternatives. Monte Carlo and bootstrap replications are set to 1.000.

Figure 2 provides the rejection rate for different $\epsilon's$ with $\alpha = 0.05$ under H_0, and under the "smooth" alternatives a) and c) with $\lambda = 0.25, 0.5$. The type I error is out of control when ϵ is close to zero, but the level accuracy is excellent for ϵ around 0.1. The power does not change much for $n = 100$, and is almost 1 for $n = 200$, in models a) and c) for different $\epsilon's$ values. Figure 3 illustrates the behavior of the test in the RDD model d). When π_0 is very small ($\pi_0 = 0.1$), the test is powerful for $\epsilon \leq 0.1$. Similar comments apply for fairly small $\epsilon's$ ($\pi_0 = 0.25$), the test is powerful for $\epsilon \leq 0.25$. In both cases the power decreases as ϵ increases for $\epsilon \geq \pi_0$. When $\pi_0 = 0.4$, the power of the test is unaffected by $\epsilon's$ choice. Of course, the power always increases with λ.

FIGURE 2 AND 3 ABOUT HERE

In order to check the level accuracy of the bootstrap test, Table 1 compares the rejection rate using asymptotic critical values (Theorem 5) and their bootstrap approximations in a model holding A8, using statistics $\tilde{\varphi}_n^{(j)}$ in a pure varying coefficient model, i.e., with $k_2 = 0$. The bootstrap test exhibits very good level accuracy for the three test statistics. As expected, the asymptotic test based on $\tilde{\varphi}_n^{(0)}$ shows poor size properties compared with the others, particularly for small n. However, the level accuracy of the asymptotic tests based on $\tilde{\varphi}_n^{(1)}$ and $\tilde{\varphi}_n^{(2)}$ is fairly good, but worse than the corresponding bootstrap tests, as expected.

TABLE 1 ABOUT HERE

Next, we perform the comparison with existing tests in the context of the PLR model. We consider the omnibus specification test proposed by Stute (1997), which is based on the CUSUM of residuals type process,

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

We use the Kolmogorov-Smirnov type statistic,

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

While the CUSUM test is able to detect any alternative to the linear regression specification hypothesis, with fairly modest power, our test is directional, designed to detect varying coefficients alternatives. We also consider the LR type bootstrap test of Cai et al. (2000), based on the test statistic $\hat{T}_n = (RSS_0/RSS_1) - 1$ that compares the restricted and unrestricted sums of squared residuals, RSS_0 and RSS_1, respectively. The bandwidth is chosen using the modified multifold cross-validation criterion suggested in Cai et. al. (2000) paper. Smooth LR type tests are asymptotically distribution free by assuming that the bandwidth converges to zero at a suitable rate as the sample size diverges (see Fan and Huang, 2005; or Cai et al., 2017). Cai et al. (2017) points out that the asymptotic approximation of $\hat{T}_n$ is poor and bandwidth dependent. Thus, they recommend using bootstrapped critical values. We report the bootstrap test they suggest, with the same bandwidth choice they propose.

Table 2 reports results for $k_1 = 1$, $k_2 = 1, 2, 3$, $\lambda = 0.25$ and $\kappa = 1$. It shows that, under H_1, our 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 that of the CUSUM test. The bootstrap smoothing based test has power similar to ours in all models, except for the RDD model, due to the poor performance of the Nadaraya-Watson estimator in this case.

TABLE 2 ABOUT HERE

Table 3 reports results for $k_1 = 2, 3, 4$, $k_2 = 1$, $\lambda = 0.25$ and $\kappa = 1$. Note that, again, our directional test works better than the omnibus CUSUM as k_1 increases. For instance, when $k_1 = 4$ and under model d), the power of our test is almost twice that of the CUSUM test. The test using $\hat{T}_n$ also suffers from the curse of dimensionality, and performs worse than the others that only need to estimate the model under H_0.

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

TABLE 3 ABOUT HERE

In the next set of simulations we apply the test to check the linearity hypothesis when $k_1 = 1$, $k_2 = 1$ and $X_2 = Z$. That is, H_0 is equivalent to omnibus specification testing of the simple regression model $\mathbb{E}\left(Y|\, Z\right) = \bar{\beta}_{01} + Z\delta_{01}$ a.s. Our test is omnibus for the linear regression specification hypothesis, and competes with the CUSUM test. Since β_{01} is not identifiable, tests based on comparing fits under the null and the alternative, like $\hat{T}_n$, cannot be implemented. We consider model b) with different ρ values. Table 4 shows that our test rejects almost double than the CUSUM for ρ large.

TABLE 4 ABOUT HERE

We also consider the test for model checking of non-linear regression models. We consider omnibus specification testing of $\mathbb{E}\left(Y|\, Z\right) = \bar{\beta}_{01} + \sum_{\ell=1}^{L} Z^{\ell}\delta_{0\ell}$ a.s. This corresponds to applying our test to model (2) with $k_1 = 1$, and $g_j(z) = z^j$, $j = 1, ..., L$. Table 5 reports the rejection rate for model b) with $\rho = 15$, which produces a sensitive departure from linearity for $L = 1, 2, 3, 4$. Our tests performs much better than the CUSUM test for $L = 1, 2$, and both have little power for $L = 4$.

TABLE 5 ABOUT HERE

Next, we consider the performance of the test as a specification test of interactive effects in model (2) with $k_1 \geq 1$, $L = 1$, $g_1(z) = 1$ and $g_2(z) = z$. That is, our test is implemented for testing that the partial effect of Z and $\boldsymbol{X}_1$ have a particular functional form. In particular, that $\boldsymbol{\beta}_{0j}\left(Z\right) = \bar{\beta}_{0j} + \delta_{0j}Z$ a.s., for $j \geq 1$. Table 6 reports the rejection rate for CUSUM and our test in model b) with $\lambda = 0.5$, different ρ values $k_2 = 0$ and $k_1 = 2, 3, 4$. Our test performs better than CUSUM, particularly for ρ large.

TABLE 6 ABOUT HERE

Now, examine testing the specification of interactive effects in the context of model (2) with $k_1 \geq 2$, $L = 1, 2, 3$, $g_0(z) = 1$ and $g_j(z) = z^j$, $j = 1, ..., L$. We consider testing $\boldsymbol{\beta}_{0j}(Z) = \bar{\beta}_{0j} + \sum_{\ell=1}^{L} Z^{\ell}\delta_{0(j+\ell-1)}$ a.s., $j \geq 1$. Table 7 reports the rejection rate for both ours and the CUSUM tests under model b) with $\lambda = 0.5$, $\rho = 15$, $k_1 = 3$ and $k_2 = 0$. Our test also performs better in general.

TABLE 7 ABOUT HERE

Finally, we show the performance of the data-driven calibration method for the optimal choice of ϵ described in Algorithm 2. We set $\ell_0 = \lfloor n/3 \rfloor$ and $e_0 = 10^{-3}$. We have used resamples $\left\{Y_i^{*(j)}, \boldsymbol{W}_i\right\}_{i=1}^{n}$, $j = 1, .., b_0$ in the step **i** of Algorithm 2 with $b_0 = b = 1.000$, which imposes H_0, as in step **i** of Algorithm 1. The data-driven calibrated ϵ is compared with different prespecified values of ϵ. Table 8 provides the rejection rate under H_0 with $k_1 = 2$, $k_1 = 1$ and $\kappa = 1$. This shows that using the calibration method we are able to almost reach the 5% significance level, which is not the case for the prespecified $\epsilon's$.

TABLE 8 ABOUT HERE

5. AN APPLICATION FOR MODELING EDUCATION RETURNS

We complement the previous Monte Carlo study with a real data application to model education returns using intelligence quotient (IQ) as a proxy variable of "ability". This is based on the work of Blackburn and Neumark (1995), which is replicated in Wooldridge's (2009a) 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. The simplest parametric model, using IQ as proxy of "ability", is

$$\log WAGE = \bar{\beta}_{01} + \bar{\beta}_{02} EDUC + \bar{\beta}_{03} IQ + \boldsymbol{X}_2^{\mathrm{T}} \boldsymbol{\delta}_0 + U, \tag{17}$$

where $WAGE$ are USD monthly earnings, $EDUC$ is years of education, and $\boldsymbol{X}_2^{\mathrm{T}} = (EXPER, TENURE, MARRIED, SOUTH, URBAN, BLACK)^{\mathrm{T}}$, $EXPER$ 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 $\boldsymbol{\delta}_0 = (\delta_{01}, ..., \delta_{06})^{\mathrm{T}}$. The "ability" is in the error term U, which is correlated with $EDUC$ for obvious reasons. The variable IQ in model (17) is a proxy of ability, which is valid if $\mathbb{E}(U|\, EDUC, IQ, \boldsymbol{X}_2) = \mathbb{E}(U|\, IQ, \boldsymbol{X}_2) = \gamma_{01} + \gamma_{02} \cdot IQ + \boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{\gamma}_0$, where γ_{01}, γ_{02} and $\boldsymbol{\gamma}_0 = (\gamma_{03}, ..., \gamma_{08})^{\mathrm{T}}$ are unknown parameters. In this case, the partial effect of $EDUC$ on $WAGE$, $\bar{\beta}_{02}$, can be consistently estimated using OLS in (17), though estimators of $\bar{\beta}_{03}$ and $\boldsymbol{\delta}_0$ are typically inconsistent. The OLS estimators of $\bar{\beta}_{02}$ and $\bar{\beta}_{03}$ in model (17), heteroskedasticity robust SE in parenthesis, are 0.054 (0.006) and 0.0036 (0.001), respectively.

The OLS estimator of the partial effect of $EDUC$ on $WAGE$ in (17) is inconsistent when either there are interactive effects or IQ enters nonlinearly into the model. A reasonable alternative to (17) is the SVC model, where

$$\log(WAGE) = \beta_{01}(IQ) + \beta_{02}(IQ) \cdot EDUC + \boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{\delta}_0 + U. \tag{18}$$

Figure 4 provides estimates of β_{01} in (18) and β_{02} using the Cai et al. (2000) procedure, with the same cross-validation bandwidth choice they suggest. We also provide OLS estimates of the parametric specification $\beta_{0j}(IQ) = \bar{\beta}_{0j}^{(1)} + \bar{\beta}_{0j}^{(2)} IQ + \bar{\beta}_{0j}^{(3)} IQ^2$, $j = 1, 2$.

FIGURE 4 ABOUT HERE

The $p-values$ for testing $H_0 : Var(\beta_{0j}(IQ)) = 0$, $j = 1, 2$ versus $H_1 : Var(\beta_{0j}(IQ) > 0$ for some $j = 1, 2$, or $H_2 : Var(\beta_{01}(IQ)) = 0$ and $Var(\beta_{02}(IQ)) > 0$, in model (18), are reported in Table 9.

TABLE 9 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 our 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 smooth 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 $EDUC$ partial effect, by testing H_0 in the model

$$\log(WAGE) = (\beta_{01}(IQ) + \delta_{07} IQ) + (\beta_{02}(IQ) + \delta_{08} IQ)\, EDUC + \boldsymbol{X}_2^{\mathrm{T}}\delta_0 + U.$$

TABLE 10 ABOUT HERE

In this case, see Table 10, 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, $EDUC$ and a simple interactive effect of $EDUC$ with IQ, cannot be rejected.

6. CONCLUSIONS AND FINAL REMARKS

We have proposed a test for coefficients constancy in SVC models based on a UI type statistic that compares the OLS coefficient estimates using subsamples of concomitants, after trimming out some observations. The test is implemented with the assistance of a wild bootstrap method, and is justified under fairly general regularity conditions. We proposed a data-driven method for calibrating the amount of trimming that minimizes the error level of the test. Under restrictive conditions, the trimming can be avoided, and the critical values can be tabulated. Under these assumptions, we proposed a Neyman-type smooth tests, and an optimal functional LR test in the direction of local alternatives, based on the principal components of the UI test statistic's empirical process.

Simulation results provided evidence of the good performance of our test in finite samples compared to a smooth LR test, and a CUSUM-type test designed for omnibus model specification testing. Simulations also showed that, unlike our test, the two competitors suffer from the curse of dimensionality, and that the LR smooth test exhibits a lack of power under alternatives with discontinuous varying coefficients. We have also included a real data application to model education partial effects controlling by IQ in a returns of education model.

Ordering the varying coefficient variable is essential for implementing our test. When the coefficients depend on a $q \times 1$ random vector $\boldsymbol{Z} = (Z_1, ..., Z_q)^{\mathrm{T}}$, i.e.

$$Y = \boldsymbol{X}_1^{\mathrm{T}} \boldsymbol{\beta}_0 (\boldsymbol{Z}) + \boldsymbol{X}_2^{\mathrm{T}} \boldsymbol{\delta}_0 + U,$$

with $\boldsymbol{\beta}_0 : \mathbb{R}^q \to \mathbb{R}^{k_1}$, the test requires ordering the data according to $\boldsymbol{Z}$ somehow. In this scenario, single-index models have proven to be an efficient way of coping with the data ordering issue, i.e. $Z = g_{\boldsymbol{\psi}_0} (\boldsymbol{Z})$ a.s. in (1) , for some unknown parameter $\boldsymbol{\psi}_0 \in \Psi \subset \mathbb{R}^q$. Xia and Li (1999) proposed a $\sqrt{n}-$consistent estimator of $\boldsymbol{\psi}_0$, $\hat{\boldsymbol{\psi}}_n$, using kernel smoothing. The test can be implemented by ordering the data according to $\hat{Z}_{ni} = g_{\hat{\boldsymbol{\psi}}_n} (\boldsymbol{Z}_i)$, $i = 1, ..., n$, and using the corresponding concomitants in (6). However, the corresponding test statistic will converge to a random variable with a different distribution than $\varphi_{\infty\epsilon}$

under H_0, because of the $\boldsymbol{\psi}_0$'s estimation effect. The test can still be implemented using wild bootstrap. Of couse, the resulting test depends on the amount of smoothing chosen for estimating $\boldsymbol{\psi}_0$. A formal justification of the test in this situation is beyond the scope of this article.

A fairly straightforward extension consists of allowing endogenous explanatory variables using the instrumental variables approach, see e.g. Cai et al. (2017). Extensions to nonlinear and multiple equations structural systems are also directly applicable.

APPENDIX

Since $F_Z(Z)$ is distributed $U(0,1)$ for all continuous F_Z, we assume w.l.o.g. that Z is distributed as an $U[0,1]$.

Proof of Theorem 1. A typical uniformity argument shows that

$$\sup_{u\in(0,1)}\left\|\left(\tilde{M}_{\ell jn}-M_{\ell j}\right)(u)\right\|=o(1)\text{ a.s.}$$

with

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

Then, (8) follows by noticing that $\hat{M}_{\ell jn}(u)=\tilde{M}_{\ell jn}(Z_{n:\lfloor nu\rfloor})$ and that, since Z is bounded on $[0,1]$, $\sup_{u\in[0,1]}\left|Z_{n:\lfloor nu\rfloor}-u\right|=o(1)$ a.s. applying the Glivenko-Cantelli theorem for the uniform quantile function (e.g. Csörgő, 1983, Remark 1). (9) follows from Davydov and Egorov (2000) Theorem 1. ■

Henceforth, $\bar{\boldsymbol{\theta}}_0=\left(\bar{\boldsymbol{\beta}}_0^{\mathrm{T}},\bar{\boldsymbol{\beta}}_0^{\mathrm{T}},\boldsymbol{\delta}_0^{\mathrm{T}}\right)^{\mathrm{T}}$.

Proof of Theorem 2. Define $\hat{\Omega}_{n\ell j}^{0}=n^{-1}\sum_{i=1}^{\lfloor nu\rfloor}\boldsymbol{X}_{\ell i}\boldsymbol{X}_{ji}^{\mathrm{T}}V_i^2$, $\ell,j=1,2$. First, notice that $\hat{\boldsymbol{\theta}}_n(1)=\bar{\boldsymbol{\theta}}_0+o(1)$ a.s., by (8) and $n^{-1}\sum_{i=1}^{\lfloor nu\rfloor}\boldsymbol{X}_{ij}V_i=o(1)$ a.s. under $A5$. Then, applying the same arguments to prove (8)

$$\sup_{u\in(0,1)}\left\|\left(\hat{\Omega}_{nj\ell}-\hat{\Omega}_{nj\ell}^{0}\right)(u)\right\|=\sup_{u\in(0,1)}\left\|\frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor}\boldsymbol{X}_{ij}\boldsymbol{X}_{i\ell}^{\mathrm{T}}\left(\hat{V}_i^2-V_i^2\right)\right\|=o(1)\text{ a.s.},$$

and $\sup_{u\in(0,1)}\left\|\left(\hat{\Omega}_{n\ell j}^{0}-\Omega_{\ell j}\right)(u)\right\|=o(1)$ a.s. $\ell,j=1,2$. Therefore,

$$\sup_{u\in(0,1)}\left\|\left(\hat{\boldsymbol{\Omega}}_0-\boldsymbol{\Omega}\right)(u)\right\|=o(1)\text{ a.s.},\tag{19}$$

and theorem follows applying (9) and the CMT. ■

Henceforth, $\boldsymbol{X}_i(u)=[\boldsymbol{X}_{1i}^{\mathrm{T}}(u),\boldsymbol{X}_{1i}^{\mathrm{T}}(1)-\boldsymbol{X}_{1i}^{\mathrm{T}}(u),\boldsymbol{X}_{2i}^{\mathrm{T}}]^{\mathrm{T}}$, with $\boldsymbol{X}_{1i}(u)=\boldsymbol{X}_{1i}1_{\left\{Z_i\le Z_{\lfloor nu\rfloor:n}\right\}}$.

Proof of Theorem 3. Under $H_{1\eta}$, by Theorem 1,

$$\frac{\tilde{\varphi}_{n\epsilon}}{n} \xrightarrow{p} \sup_{\epsilon \leq u \leq 1-\epsilon} \boldsymbol{\eta}_0^{\mathrm{T}}(u) \left[\boldsymbol{R}\boldsymbol{M}^{-1}(u)\boldsymbol{\Omega}(u,v)\boldsymbol{M}^{-1}(v)\right]^{-1} \boldsymbol{R}^{\mathrm{T}}\boldsymbol{\eta}_0(u) > 0,$$

which proves (11). In order to prove, (12), notice that under H_{n1},

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}_0\right)(u) = \hat{\boldsymbol{M}}_n^{-1}(u)\left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i(u)\boldsymbol{X}_i^{\mathrm{T}}(u) \cdot \left(\boldsymbol{\tau}^{\mathrm{T}}(Z_i), \boldsymbol{\tau}^{\mathrm{T}}(Z_i), \mathbf{0}\right)^{\mathrm{T}} + \sqrt{n}\hat{\mathbf{N}}_n(u)\right],$$

and, under $A6$,

$$\sup_{u\in(0,1)} \left\| n^{-1}\sum_{i=1}^{n} \boldsymbol{X}_i(u)\boldsymbol{X}_i^{\mathrm{T}}(u) \cdot \left(\boldsymbol{\tau}^{\mathrm{T}}(Z_i), \boldsymbol{\tau}^{\mathrm{T}}(Z_i), \mathbf{0}\right)^{\mathrm{T}} - \boldsymbol{T}(u) \right\| = o(1) \text{ a.s.}$$

using the same arguments to prove (8) in Theorem 1. Then, apply (8), (9) and the continuous mapping theorem (CMT) to complete the proof. ■

Proof of Theorem 4. It suffices to show that for any $c > 0$,

$$\mathbb{P}_{\xi}\left(\hat{\varphi}_{n\epsilon}^{*} \leq c\right) \to \mathbb{P}\left(\varphi_{\infty\epsilon} \leq c\right) \text{ a.s.}, \tag{20}$$

Notice that uniformly in $u \in (0,1)$,

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

Following the strategy of proof in Stute et al. (1998) (SGQ), (20) follows by showing that, conditional to the sample, $\sqrt{n}\hat{\mathbf{N}}_n^{*}$ converges in distribution to $\boldsymbol{N}_{\infty}$ a.s., i.e. for almost all sample $\{Y_i, \boldsymbol{W}_i\}_{i=1}^{n}$, 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 $u_1, u_2 \in (0,1)$,

$$\mathbb{E}_{\xi}\left[\hat{\mathbf{N}}_n^{*}(u_1)\,\hat{\mathbf{N}}_n^{*\mathrm{T}}(u_2)\right] = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i(u_1)\boldsymbol{X}_i^{\mathrm{T}}(u_2)\hat{V}_i^2 = \boldsymbol{\Omega}(u_1,u_2) + o(1) \text{ a.s.}$$

using (19). Then, fixing $u_1, ..., u_q$, by the Cramér-Wold device, it suffices to show that,

$$\mathbb{P}_{\xi}\left\{\sqrt{n}\sum_{j=1}^{q} \boldsymbol{c}_j^{\mathrm{T}}\hat{\mathbf{N}}_n^{*}(u_j) \leq \varepsilon\right\} \to \mathbb{P}\left\{\sum_{j=1}^{q} \boldsymbol{c}_j^{\mathrm{T}}\boldsymbol{N}_{\infty}(u_j) \leq \varepsilon\right\} \text{ a.s.}, \tag{21}$$

for any $\varepsilon < \infty$, $\boldsymbol{c}_j \in \mathbb{R}^{2k_1+k_2}$, $j = 1, .., q$. Write $\vartheta_i = \sum_{j=1}^{q} a_j \boldsymbol{c}^{\mathrm{T}}\boldsymbol{X}_i(u_j)$; then,

$$\sqrt{n}\sum_{j=1}^{q} a_j \boldsymbol{c}^{\mathrm{T}}\hat{\mathbf{N}}_n^{*}(u_j) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \vartheta_i \hat{V}_i \xi_i.$$

Then (21) follows by checking the Linderberg condition

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

Since $|\xi| \leq C$,

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

In order to show tightness, it suffices to check Billingsley (1968) Theorem 15.7 as in SGQ Lemma A3. Define, for $\boldsymbol{c} \in \mathbb{R}^{2k_1+k_2}$,

$$\hat{\omega}^*_{n\boldsymbol{bi}}(u) = \sqrt{n}\boldsymbol{c}^{\mathrm{T}} \hat{\mathbf{N}}^*_n(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{c}^{\mathrm{T}} \boldsymbol{X}_i(u)\hat{V}_i \xi_i.$$

In order to prove tightness, we must show, as in SGQ Lemma 3, that for any $\boldsymbol{c} = (\boldsymbol{c}_1^{\mathrm{T}}, \boldsymbol{c}_2^{\mathrm{T}}, \boldsymbol{c}_3^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathbb{R}^{k_1}$, $\boldsymbol{c}_3 \in \mathbb{R}^{k_2}$ and $K/n \leq u_0 \leq u_1 \leq u_2 < 1 - K/n$,

$$\mathbb{E}_\xi\left([\hat{\omega}^*_{n\boldsymbol{bi}}(u_1) - \hat{\omega}^*_{n\boldsymbol{bi}}(u_0)]^2 [\hat{\omega}^*_{n\boldsymbol{bi}}(u_2) - \hat{\omega}^*_{n\boldsymbol{bi}}(u_1)]^2 \right) \leq C\left[\hat{G}_n(u_2) - \hat{G}_n(u_0)\right]^2, \tag{22}$$

where $C < \infty$ is a generic constant, $\hat{G}_n$ is monotone, and $\hat{G}_n \to G$ a.s. First notice that for any $u_\ell \geq u_j$, $u_\ell, u_j \in [K/n, 1 - K/n)$,

$$\hat{\omega}^*_{n\boldsymbol{bi}}(u_\ell) - \hat{\omega}^*_{n\boldsymbol{bi}}(u_j) = (\boldsymbol{c}_1 - \boldsymbol{c}_2)^{\mathrm{T}} \boldsymbol{X}_{1i} \hat{V}_i \xi_i 1_{\left\{Z_{\lfloor nu_j \rfloor:n} \leq Z_i \leq Z_{\lfloor nu_\ell \rfloor:n}\right\}}.$$

Then, applying Lemma 5.1 of Stute (1997),

$$LHS(22) \leq \frac{3\left\|\boldsymbol{c}_1 - \boldsymbol{c}_2\right\|^2}{n^2}\sum\sum_{i \neq j} \mathbb{E}_\xi \lambda_i^2 \mathbb{E}_\xi \gamma_j^2,$$

$\lambda_i = \|\mathbf{X}_{1i}\| \hat{V}_i \xi_i 1_{\{Z_{\lfloor nu_0 \rfloor:n} \leq Z_i \leq Z_{\lfloor nu_1 \rfloor:n}\}}$, $\gamma_i = \|\mathbf{X}_{1i}\| \hat{V}_i \xi_i 1_{\{Z_{\lfloor nu_1 \rfloor:n} \leq Z_i \leq Z_{\lfloor nu_2 \rfloor:n}\}}$. Then,

$$\begin{aligned} LHS(22) &\leq \frac{C}{n^2}\sum\sum_{i \neq j} \|\mathbf{X}_{1i}\|^2 \|\mathbf{X}_{1j}\|^2 \hat{V}_i^2 \hat{V}_j^2 1_{\{Z_{\lfloor nu_1 \rfloor:n} \leq Z_i \leq Z_{\lfloor nu_2 \rfloor:n}\}} 1_{\{Z_{\lfloor nu_0 \rfloor:n} \leq Z_i \leq Z_{\lfloor nu_1 \rfloor:n}\}} \\ &\leq C\left[\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{X}_{1i}\|^2 \hat{V}_i^2 1_{\{Z_{\lfloor nu_0 \rfloor:n} \leq Z_i \leq Z_{\lfloor nu_2 \rfloor:n}\}}\right]^2 = C\left[\hat{G}_n(u_2) - \hat{G}_n(u_0)\right]^2, \end{aligned}$$

where C is a generic constant, and

$$\hat{G}_n(u) = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{X}_{1i}\|^2 \hat{V}_i^2 1_{\{Z_i \leq Z_{\lfloor nu \rfloor:n}\}}$$

is monotone and $\sup_{u\in(0,1)}\left\|\left(\hat{G}_n - G\right)(u)\right\| = o(1)$ a.s., with $G(u) = \mathbb{E}\left(\left\|\mathbf{X}_1\right\|^2 U^2 1_{\{Z\le u\}}\right)$ a.s., using a Glivenko-Cantelli argument as in the proof of Theorem 1. ■

Proof of theorem 5. Define

$$\begin{aligned}\hat{\boldsymbol{\eta}}_n^{\dagger}(u) &= M_{11}^{-1}(1)\frac{\hat{N}_{1n}(u) - u\hat{N}_{1n}(1)}{u(1-u)},\\ \tilde{\alpha}_n^{\dagger}(u) &= \hat{\boldsymbol{\eta}}_n^{\dagger\mathrm{T}}(u)\,\frac{M_{11}(1)\,u(1-u)}{\sigma^2}\hat{\boldsymbol{\eta}}_n^{\dagger}(u)\end{aligned}$$

Now notice that

$$\begin{aligned}u(1-u)M_{11}(1)\left(\hat{\boldsymbol{\eta}}_n - \hat{\boldsymbol{\eta}}_n^{\dagger}\right)(u) &= (1-u)\left(uM_{11}(1)\hat{M}_{11}^{-1}(u) - \boldsymbol{I}_{k_1}\right)\hat{N}_{1n}(u) \\ &+ u\left((1-u)M_{11}(1)\left[\hat{M}_{11}(1) - \hat{M}_{11}(u)\right]^{-1} - \boldsymbol{I}_{k_1}\right)\left(\hat{N}_{1n}(1) - \hat{N}_{1n}(u)\right).\end{aligned} \tag{23}$$

Let $\{c_n\}_{n\ge1}$ be a sequence such that $c_n + (nc_n) \to 0$ as $n\to\infty$. Applying Theorem 0 in Wellner (1978) to this context,

$$\begin{aligned}\sup_{k_1/n+c_n\le u<(n-k_1)/n}\left\|M_{11}(1)u\tilde{M}_{11}^{-1}(u) - \boldsymbol{I}_{k_1}\right\| &= o_p(1) \\ \sup_{k_1/n\le u<(n-k_1)/n-c_n}\left\|(1-u)\,M_{11}(1)\left[\tilde{M}_{11}(1) - \tilde{M}_{11}(u)\right]^{-1} - \boldsymbol{I}_{k_1}\right\| &= o_p(1).\end{aligned} \tag{24}$$

Since $\hat{M}_{11}(u) = \tilde{M}_{11}\left(Z_{n:\lfloor nu\rfloor}\right)$ and $\sup_{u\in[0,1]}\left|Z_{n:\lfloor nu\rfloor} - u\right| = o(1)$ a.s., by (23) and (24), and $\sup_{u\in[0,1]}\left|\hat{N}_{1n}(u)\right| = O_p(n^{-1/2})$ by (9),

$$\sup_{k_1+c_n\le u<n-k_1-c_n}\left\|u(1-u)M_{11}(1)\left(\hat{\boldsymbol{\eta}}_n^{\dagger} - \hat{\boldsymbol{\eta}}_n\right)(u)\right\| = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{25}$$

This implies that

$$\max_{k_1+nc_n\le j<n-k_1-nc_n}\left|\frac{j(n-j)}{n}\left[\left(\tilde{\alpha}_n^{\dagger} - \tilde{\alpha}_n\right)\left(\frac{j}{n}\right)\right]\right| = o_p(1)$$

proves (16), applying the CMT, after noticing that

$$\tilde{\varphi}_{nc_n}^{(2)} = \max_{k_1+nc_n\le j<n-k_1-nc_n}\frac{j(n-j)}{n}\tilde{\alpha}_n\left(\frac{j}{n}\right),$$

and $\tilde{\varphi}_n^{(2)}$ are asymptotically equivalent, and that

$$\sqrt{n}M_{11}^{1/2}(1)\frac{u(1-u)}{\sigma}\hat{\boldsymbol{\eta}}_n^{\dagger} \xrightarrow{d} \boldsymbol{B}_0 \text{ in } \mathbb{D}[0,1]. \tag{26}$$

Now, in view of (26) applying an extension of the Anderson-Darling result to the multivariate case, e.g. Scholz and Stephens (1987) or Csörgő and Horváth (1997) Corollary 1.1.1 for general weight functions,

$$\tilde{\varphi}_n^{(1)\dagger} \xrightarrow{d} \int_0^1 \frac{\boldsymbol{B}_0^{\mathrm{T}}(u)\boldsymbol{B}_0(u)}{u(1-u)}du,$$

which proves (15). And applying the extension of the Darling-Erdős theorem in Shorack (1979) to the vector case, as in Horváth (1993),

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

which proves (14). ■

REFERENCES

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

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 1383–1414.

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.

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.

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.

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

Chou, S.-Y., Liu, J.-T., Huang, C.J., 2004. Health insurance and savings over the life cycle - a semiparametric smooth coefficient estimation. J. Appl. Econom. 19, 295–322.

Csörgő, M. (1983). Quantile processes with statistical applications. SIAM (Vol. 42). Philadephia.

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

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.

Delgado, M. A., Hidalgo, J., and Velasco, C. (2005). Distribution free goodness-of-fit tests for linear processes. Ann. Stat. 33, 2568-2609.

Delgado, M. A., and Stute, W. (2008). Distribution-free specification tests of conditional models. J. Econom. 143, 37-55.

Durbin, J., & Knott, M. (1972). Components of Cramér–von Mises statistics. J. R. Stat. Soc. Ser. B. 34, 290-307.

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.

Frölich, M., 2008. Parametric and non-parametric regression in the presence of endogenous control variables. Int. Stat. Rev. 76, 214–227.

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

Grenander, U., 1950. Stochastic processes and statistical inference. Ark. för Mat., 1(3), 195-277.

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

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.

Kiefer, J., 1959. K-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests. Ann. Math. Stat. 30, 420-447.

Lee, Y., Stoyanov, A., Zubanov, N., 2019. Olley and Pakes-style Production Function Estimators with Firm Fixed Effects. Oxf. Bull. Econ. Stat. 81, 79–97.

Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Rev. Econ. Stud. 70, 317–341.

Li, Q., Huang, C.J., Li, D., Fu, T.-T., 2002. Semiparametric smooth coefficient models. J. Bus. Econ. Stat. 20, 412–422.

Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry, Econometrica, 64, 1263–1297

Politis, D. N., Romano, J. P., & Wolf, M. (1999). Subsampling. Springer.

Schoenfeld, D. A. (1977). Asymptotic properties of tests based on linear combinations of the orthogonal components of the Cramer-von Mises statistic. Ann. Stat. 5, 1017-1026.

Scholz, F. W., Stephens, M. A. (1987). K-sample Anderson–Darling tests. J. Am. Stat. Assoc. 82(399), 918-924.

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

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

Sowell, F., 1996. Optimal tests for parameter instability in the generalized method of moments framework. Econometrica, 1085-1107.

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.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Am. Math. Soc. 54, 426–482.

Wang, H., Xia, Y., 2009. Shrinkage estimation of the varying coefficient model. J. Am. Stat. Assoc. 104, 747–757.

Wellner, J. A., 1978. Limit theorems for the ratio of the empirical distribution function to the true distribution function. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 45(1), 73-88.

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

Wooldridge, J.M., 2009b. On estimating firm-level production functions using proxy variables to control for unobservables. Econ. Lett. 104, 112–114.

Xia, Y., Li, W.K., 1999. On single-index coefficient regression models. J. Am. Stat. Assoc. 94, 1275–1285.

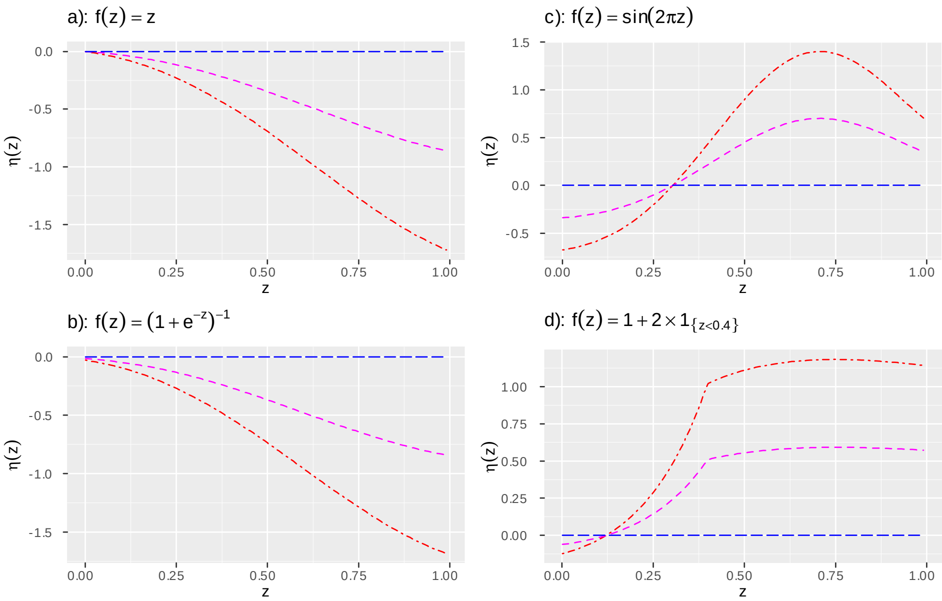

Figure 1: *Representation of η_0 for different models when $\lambda = 0$ (blue curve), $\lambda = 0.25$ (purple curve), and $\lambda = 0.5$ (red curve).*

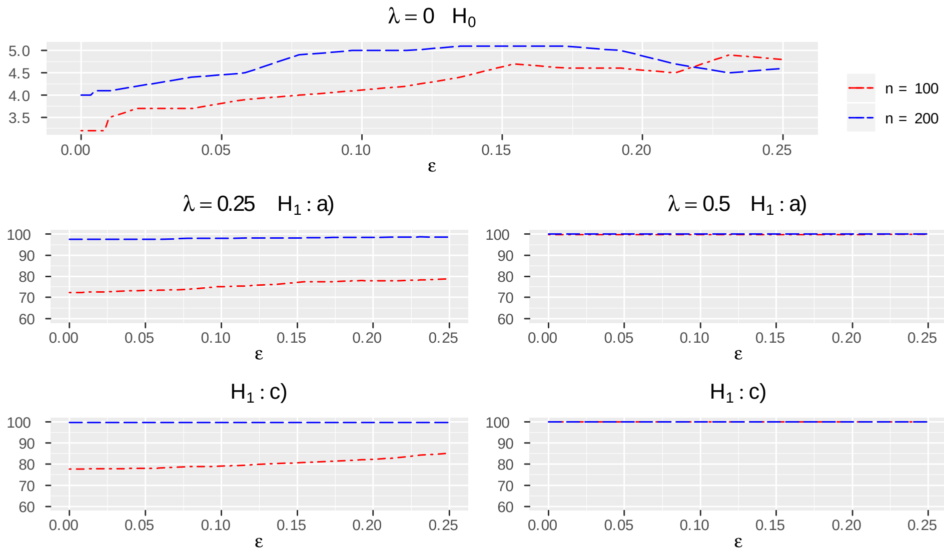

Figure 2: *Representation of $\hat{\Phi}_{n\epsilon}(\alpha)$ for the null and different alternatives.*

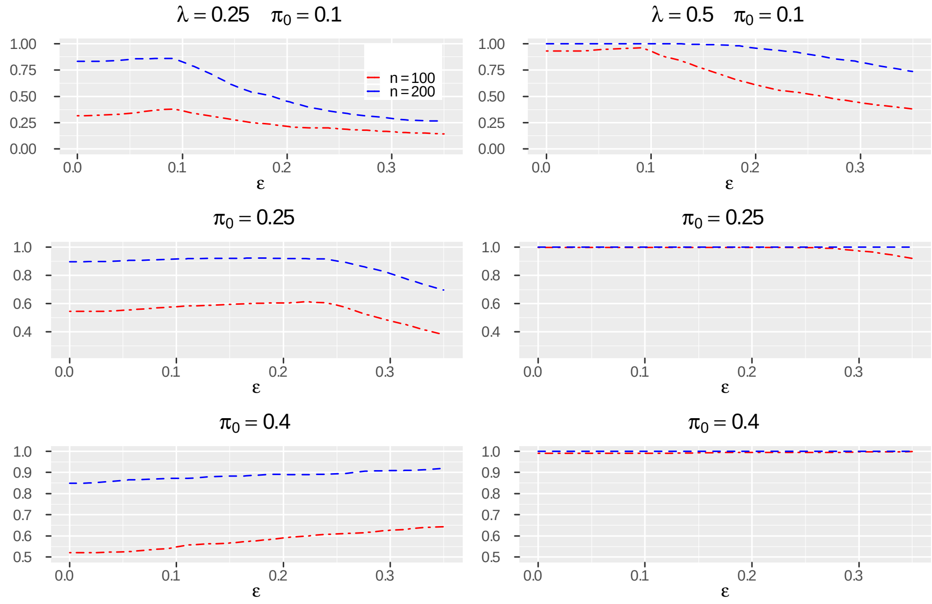

Figure 3: *Representation of* $\hat{\Phi}_{n\epsilon}(\alpha)$ *under alternative* d) *for different* π_0 *values.*

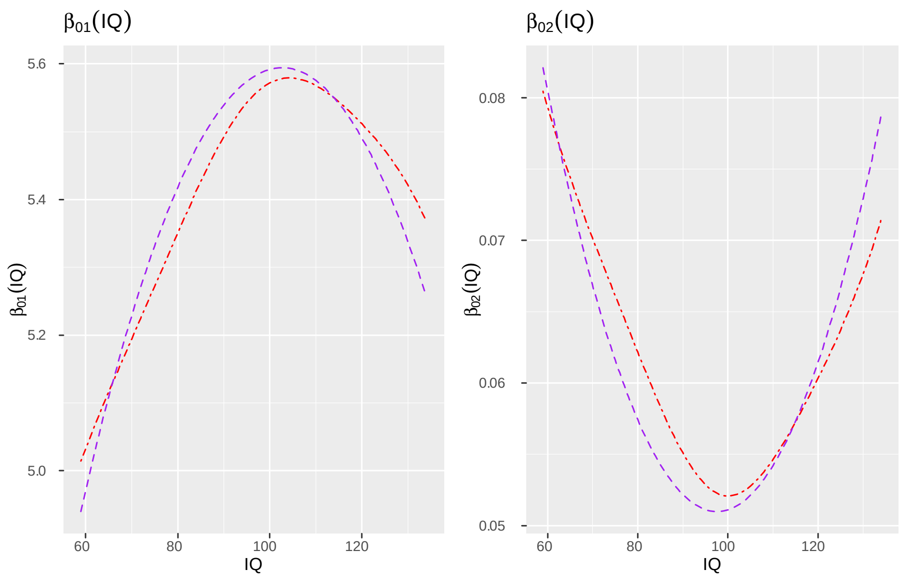

Figure 4: *Representation of* $\beta_{00}(IQ)$ *and* $\beta_{01}(IQ)$ *for the estimates of the varying coefficients using kernels (red curve), and OLS estimates of the parametric specification (purple curve).*

α	1%				5%				10%			
k_1	1	2	3	4	1	2	3	4	1	2	3	4
$\tilde{\varphi}_n^{(0)}$ (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 $\kappa = 0$)*

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 ($k_1 = 1$, $\lambda = 0.25$ and $\kappa = 1$)*

Model	H_0			$H_1: a$			$H_1: c$			$H_1: d$		
k_1	2	3	4	2	3	4	2	3	4	2	3	4
$\hat{\varphi}_{n0.02}$												
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 ($k_2 = 1$, $\lambda = 0.25$ and $\kappa = 1$)*

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 ($k_1 = 1$, $k_2 = 1$ and $\kappa = 1$)*

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 ($k_1 = 1$, $k_2 = 1$, $\rho = 15$ and $\kappa = 1$)*

ρ	1			2			3			15		
k_1	2	3	4	2	3	4	2	3	4	2	3	4
$\hat{\varphi}_{n0.02}$												
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 $\kappa = 1$)*

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 ($k_1 = 3$, $k_2 = 0$, $\lambda = 0.5$ $\rho = 15$ and $\kappa = 1$)*

ϵ	Data-driven	0.02	0.06	0.07	0.13	0.15	0.18	0.20	0.24	0.26	0.28	0.35
50	5.9	3.95	3.95	3.95	3.95	3.95	3.95	5.26	3.95	3.95	3.95	5.26
100	5.2	1.32	2.63	1.32	1.32	2.63	3.95	5.26	5.26	3.95	5.26	3.95

Table 8. *Percentage of times H_0 was rejected, 5% of significance ($k_1 = 2$, $k_2 = 1$, and $\kappa = 1$)*

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