This is a postprint version of the following published document:


DOI: 10.1016/j.jeconom.2020.07.041

© Elsevier, 2020

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Testing Constancy in Varying Coefficient Models

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

Abstract

This article proposes a coefficients constancy test in semi-varying coefficient models that 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 a 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 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; Union-intersection 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.

†Corresponding author: e-mail: delgado@test-econ.uc3m.es.
1. INTRODUCTION

This article proposes a coefficient constancy test for semi-varying coefficient (SVC) models, where some partial effects are non(semi)-parametric functions of one of the explanatory variables. The test can be used as a significance test in linear regression, able to detect non-parametric alternatives where the partial effect is of unknown functional form. It can also be applied as a model-check technique for interactive effects in linear in parameters regression models. Existing tests are based on smooth estimates of the varying coefficients using kernels, most of them comparing the sum of square residuals under the null and under the alternative, assuming that the varying coefficients are smooth enough. See for instance 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. Unlike the preceding proposals, our test does not need to use smoothing, is consistent under discontinuous alternatives, and detects alternatives converging to the null at the parametric rate $n^{-1/2}$. We adapt the union-intersection (UI) testing procedure in time series regression analysis for testing parameter stability with respect to "time", which compares least square estimates of each pair of subsamples before and after each moment of time. See for example Hawkins (1989) or Horváth and Shao (1995). We apply the UI principle after sorting the data according to the varying coefficient’s explanatory variable.

Our test belongs to the family of cumulative sums (CUSUM) of concomitants, introduced by Bhattacharya (1974) and motivated to test the constancy of conditional expectations. This work was extended by Andrews (1997) and Stute (1997) to specification testing of conditional distributions and regression functions, respectively, which has formed a basis for specification testing of many other models defined by means of general conditional moment restrictions. While UI parameter stability tests in time series regression analysis are based on CUSUMs of sequential observations, our UI test is based on CUSUMs of concomitants (induced order statistics).

The proposed test is omnibus, consistent in the direction of any possible non-parametric alternative, in pure varying coefficient models with no constant coefficients, and also in general SVC models where the varying coefficients’ variable is independent in mean of the explanatory variables’ cross-products. 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 a parameter should be chosen small enough 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. In general, under restrictive assumptions, the 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. Under fairly weak assumptions, we justify implementing the test with the assistance of a wild bootstrap procedure.

SVC models have been applied in economics with different motivations. 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, tests for coefficients’ constancy, consistent under non-parametric alternatives, are useful either for significance testing, or for partial effects’ model checks.

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 on an application of our proposal for modelling ed-
ucation returns controlling for unobserved individual ability using IQ as a 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 $W = (Z, X_1^\top, X_2^\top)^\top$ are related according to the SVC model

$$ Y = X_1^\top \beta_0 (Z) + X_2^\top \delta_0 + U, \quad (1) $$

where "$\top$" means transpose, $U$ is an unobserved error term such that $\mathbb{E}(U|W) = 0 \text{ a.s.}$, $X_j = (X_{j1}, ..., X_{jk})^\top$ 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 $\beta_0 = (\beta_{01}, ..., \beta_{0k_1})^\top : \mathbb{R} \to \mathbb{R}^{k_1}$ consists of possibly non-smooth unknown functions, and $\delta_0 = (\delta_{01}, ..., \delta_{0k_2})^\top$ is a $k_2 \times 1$ vector of unknown parameters.

A constancy test of $\beta_0$, taking $X_{11} = 1$, is in fact a significance test of $Z$ in a multiple regression model, consistent under non-parametric alternatives, where the $Z$'s partial effect is of unknown functional form. Since $X_2$ may be a vector of functions of $Z$ and $X_1$, a $\beta_0$'s constancy test can also be used as a model check of the partial effects of $X_1$.

Consider

$$ X_2 = (X_{11}g_1^\top(Z), ..., X_{1k_1}g_{k_1}^\top(Z))^\top, \quad \delta_0 = (\delta_{01}^\top, ..., \delta_{0k_1}^\top)^\top \text{ and } k_2 = \sum_{j=1}^{k_1} m_j, $$

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

$$ \mathbb{E}(Y|Z, X_1) = X_1^\top [\beta_0(Z) + r_\delta(Z)] \text{ a.s.} \quad (2) $$

with non-parametric $\beta_0$ and parametric $r_\delta(\cdot) = (g_1^\top(\cdot)\delta_{01}, ..., g_{k_1}^\top(\cdot)\delta_{0k_1})^\top$, for some $\delta_0 = (\delta_{01}^\top, ..., \delta_{0k_1}^\top)^\top \in \mathbb{R}^{k_2}$. Therefore, assuming (2), testing $H_0$ is equivalent to check that $\mathbb{E}(Y|Z, X_1) = X_1^\top [\bar{\beta}_0 + r_\delta(Z)]$ a.s. for some $(\bar{\beta}_0^\top, \delta_0^\top)^\top \in \mathbb{R}^{k_1+k_2}$, in the direction (2) for non-parametric $\beta_0$, where $Z$ can be some component of $X_1$. When $Z$ is not a component of $X_1$, a constancy test for $\beta_0$ in (2) is in fact a specification test of a parametric model for the partial effects of $X_1$, $\bar{\beta}_0 + r_\delta(Z)$. In the returns of education application in Section 5, the focus of interest is the partial effect of education.
(EDUC), where Z is the intelligence quotient (IQ), and the rest of explanatory variables are worker characteristics such as tenure, experience, being married, race, etc. As usual, the dependent variable is the wage (WAGE) in logs. In this model, IQ is a proxy variable of unobserved worker characteristics related to her/his relative "ability", which is correlated with EDUC. First, we use our test as a significance test for IQ and, second, as a specification test for a parametric model of the interactive effects between EDUC and IQ.

The null hypothesis is formally stated as

$$H_0 : \beta_0 (Z) = \bar{\beta}_0 \ a.s.,$$

where $\bar{\beta}_0 := E(\beta_0 (Z)) = (\bar{\beta}_{01}, ..., \bar{\beta}_{0k_1})^T \ a.s.$, which can be equivalently expressed as $H_0 : \text{Var} (\beta_0 (Z)) = E((\beta_0 (Z) - \bar{\beta}_0)^2 = 0$ for all $j = 1, ..., k_1$. As mentioned in the previous section, the existing tests are based on smooth estimates of $\beta_0$, usually kernels, and they are not valid under non-smooth alternatives. In contrast, the test we propose does not estimate $\beta_0$, and is consistent under alternatives where $\beta_0$ is discontinuous.

This hypothesis nests the case $k_2 = 0, k_1 = 1$ with $X_{11} = 1$, i.e. when $Y = \beta_0 (Z) + U$, a pure non-parametric model. In this case, Y and Z are independent in mean under $H_0$, which can be tested using Bhattacharya’s (1974) CUSUM of concomitants test that inspires our proposal.

Define $S(u) = (S_1^T(u), S_1^T(1) - S_1^T(u), S_2^T(1))^T$

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

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

**A1:** $F_Z$ is continuous.

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

5
Our test is based on comparing the vector of functions $\mathbf{b}_0^+$ and $\mathbf{b}_0^-$, where,

$$\theta_0(u) = (\mathbf{b}_0^{-\top}(u), \mathbf{b}_0^{+\top}(u), \mathbf{d}_0^u(u))^{\top}$$

$$= \arg \min_{\mathbf{b}^-, \mathbf{b}^+, \mathbf{d}} \mathbb{E} \left( Y - (\mathbf{X}_1 1_{(F_Z(z) \leq u)})^{\top} \mathbf{b}^- - (\mathbf{X}_1 1_{(F_Z(z) > u)})^{\top} \mathbf{b}^+ - \mathbf{X}_2^{\top} \mathbf{d} \right)^2 \quad (3)$$

$$= \mathbf{M}^{-1}(u) \mathbf{S}(u), \quad u \in (0, 1).$$

Then, $\theta_0(u)$ are the coefficients of the best linear predictor of $Y$ given $\mathbf{X}_2$ and $\mathbf{X}_1$ splitted into two truncated variables according to the $u$-th $Z$'s quantile, $F_Z^{-1}(u)$. We can express $\mathbf{b}_0^\pm$ in terms of $\mathbf{b}_0$, as

$$\theta_0(u) = \mathbf{M}^{-1}(u) \cdot \mathbb{E} \left( \mathbf{m}(u) \left( \mathbf{\beta}_0^+(Z), \mathbf{\beta}_0^-(Z), \mathbf{\delta}_0^u \right)^{\top} \right), \quad (4)$$

with $\mathbf{M}(u) = \mathbb{E} (\mathbf{m}(u))$. Notice that, under $H_0$, the objective function in (3) is

$$\mathbb{E} \left( (Y - \mathbf{X}_1^{\top} \mathbf{b}^- - \mathbf{X}_2^{\top} \mathbf{d})^2 1_{(F_Z(z) \leq u)} \right) + \mathbb{E} \left( (Y - \mathbf{X}_1^{\top} \mathbf{b}^+ - \mathbf{X}_2^{\top} \mathbf{d})^2 1_{(F_Z(z) > u)} \right) \quad (5)$$

$$= \mathbb{E} (U^2) + \mathbb{E} \left( \left( (\mathbf{X}_1^{\top} (\bar{\mathbf{\beta}}_0 - \mathbf{b}^-) + \mathbf{X}_2^{\top} (\delta_0 - \mathbf{d})) 1_{(F_Z(z) \leq u)} \right)^2 \right)$$

$$+ \mathbb{E} \left( \left( (\mathbf{X}_1^{\top} (\bar{\mathbf{\beta}}_0 - \mathbf{b}^+) + \mathbf{X}_2^{\top} (\delta_0 - \mathbf{d})) 1_{(F_Z(z) > u)} \right)^2 \right)$$

$$\geq \mathbb{E} (U^2) \quad \text{for all} \quad (\mathbf{b}^-, \mathbf{b}^+, \mathbf{d}) \in \mathbb{R}^{2k_1 + k_2} \quad \text{and} \quad u \in (0, 1),$$

Therefore, from either (4) or (5), a necessary condition for $H_0$ is that $\mathbf{b}_0^-(u) = \mathbf{b}_0^+(u) = \bar{\mathbf{b}}_0$ uniformly in $u \in (0, 1)$, where $\bar{\mathbf{b}}_0$ is in fact the coefficients' vector of the best linear predictor of $Y$ given $(\mathbf{X}_1, \mathbf{X}_2)$. The test statistic, introduced in the next section, is a functional of the sample version of

$$\mathbf{\eta}_0(u) = (\mathbf{b}_0^- - \mathbf{b}_0^+) (u) = \mathbf{RM}^{-1}(u) \mathbf{S}(u),$$

with $\mathbf{R} = \begin{bmatrix} \mathbf{I}_{k_1} & - \mathbf{I}_{k_1} & 0 \end{bmatrix}$, $\mathbf{I}_m$ is the $m \times m$ identity matrix, which detects any alternative to $H_0$ of the form,

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

However, tests consistent for $H_0$ under $H_{1\eta}$ may have trivial power in infinite many other directions. That is, in general, $H_{1\eta}$ is necessary, but not sufficient, for

$$H_1 : \text{Var} \left( \mathbf{\beta}_{0j}(Z) \right) > 0 \quad \text{for some} \quad j = 1, \ldots, k_1.$$

That is, tests consistent under $H_{1\eta}$ may not be consistent under $H_1$, i.e. omnibus for $H_0$. However, $H_{1\eta}$ is equivalent to $H_1$ 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 $\mathbb{E}(X_1X_j^T|Z) = M_{1j}(1)$ a.s., $j = 1, 2$. Therefore,
\[
S_1(u) = uM_{11}(1)\mathbb{E}(\beta_0(Z)1_{\{F_Z(Z)\leq u\}}) + uM_{21}(1)\delta_0.
\]
Reasoning as in Andrews (1993) Lemma A.5, define $\nu = (v_1^1, v_2^1, v_3^1)^\top = M^{-1}(u)S(u)$. Then, $M(u)\nu = S(u)$, and
\[
\begin{pmatrix}
  uM_{11}(1) & 0 \\
  0 & (1 - u)M_{11}(1)
\end{pmatrix}
\begin{pmatrix}
  v_1 \\
  v_2
\end{pmatrix}
= \begin{pmatrix}
  S_1(u) \\
  S_1(1) - S_1(u)
\end{pmatrix}
- \begin{pmatrix}
  uM_{21}(1)v_3 \\
  (1 - u)M_{21}(1)v_3
\end{pmatrix}.
\]
Therefore,
\[
\begin{pmatrix}
  v_1 \\
  v_2
\end{pmatrix}
= \begin{pmatrix}
  M_{11}^{-1}(1)S_1(u)/u \\
  M_{11}^{-1}(1)[S_1(1) - S_1(u)]/(1 - u)
\end{pmatrix}
- \begin{pmatrix}
  M_{11}^{-1}(1)M_{21}(1)v_3 \\
  M_{11}^{-1}(1)M_{21}(1)v_3
\end{pmatrix},
\]
and
\[
\eta_0(u) = RM^{-1}(u)S(u)
= v_1 - v_2
= M_{11}^{-1}(1)S_1(u) - uS_1(1)
= \frac{1}{u(1 - u)} \int_{-\infty}^{F_Z^{-1}(u)} (\beta_0(z) - \bar{\beta}_0) F_z(dz).
\]
Hence, $\eta_0(u) = 0$ for all $u \in (0, 1)$ iff $\beta_0(Z) = \bar{\beta}_0$ a.s.

Remark 2 Suppose that all the coefficients are varying ($k_2 = 0$), that is, (1) is a pure varying coefficient model. Then, for all $u \in (0, 1)$,
\[
\eta_0(u) = M_{11}^{-1}(u)S_1(u) - [M_{11}(1) - M_{11}(u)]^{-1}[S_1(1) - S_1(u)]
= [M_{11}(1) - M_{11}(u)]^{-1}M_{11}(1)M_{11}^{-1}(u)[S_1(u) - M_{11}(u)M_{11}^{-1}(1)S_1(1)].
\]
Therefore, for all $u \in (0, 1)$
\[
\eta_0(u) = 0 \iff S_1(u) - M_{11}(u)M_{11}^{-1}(1)S_1(1) = 0
\iff \int_{\{F_Z(Z)\leq u\}} J(Z)\{\beta_0(Z) - \mathbb{E}[J(Z)]^{-1}\mathbb{E}[J(Z)\beta_0(Z)]\}dP = 0.
\]
with $J(Z) = \mathbb{E}(X_1X_1^T|Z)$. Hence, if $J(Z)$ is non-singular a.s.,
\[
\eta_0(u) = 0 \text{ all } u \in (0, 1) \iff \beta_0(Z) = \mathbb{E}[J(Z)]^{-1}\mathbb{E}[J(Z)\beta_0(Z)] \text{ a.s.}
\iff \beta_0(Z) = \bar{\beta}_0 \text{ a.s.}
Therefore, when either \( X_1 X_j^T \), \( j = 1, 2 \), and \( Z \) are independent in mean, or when all parameters are varying (\( k_2 = 0 \)), \( H_1 \) and \( H_{1\eta} \) are equivalent. That is, under these conditions, the test presented in next section, designed to be consistent under \( H_{1\eta} \), is omnibus.

3. TESTING METHOD

Given \( \{Y_i, W_i\}_{i=1}^n \) i.i.d. as \( (Y, W) \), \( W_i = (Z_i, X_{1i}^T, X_{2i}^T)^T \), \( \{Y_{[i:n]}, X_{1[i:n]}, X_{2[i:n]}\}_{i=1}^n \) denotes the \( Z \)-concomitant of \( \{Y_i, X_1, X_2\}_{i=1}^n \), i.e. for a generic data set \( \{\zeta_i\}_{i=1}^n, \zeta_{[i:n]} = \zeta_j \) iff \( Z_{ni} = Z_j \), where \( Z_{n1} \leq Z_{n2} \leq \ldots \leq Z_{nn} \) are order statistics of \( \{Z_i\}_{i=1}^n \).

The sample analogue of (3) is
\[
\hat{\theta}_n(u) = \arg\min_{b^-b^+,d} \left\{ \sum_{i=1}^{[nu]} (Y_{[i:n]} - X_{1[i:n]}^T b^- - X_{2[i:n]}^T d)^2 + \sum_{i=1+|nu|}^n (Y_{[i:n]} - X_{1[i:n]}^T b^+ - X_{2[i:n]}^T d)^2 \right\}
\]
\[
= \hat{M}_n^{-1}(u) \hat{S}_n(u), \quad u \in [K/n, 1-K/n),
\]
where \([\cdot]\) means smallest nearest integer, \( \hat{S}_n(u) = \left( \hat{S}_{n1}(u), \hat{S}_{n1}(1) - \hat{S}_{n1}(u), \hat{S}_{n2}(1) \right)^T \), where \( \hat{S}_{nj}(u) = n^{-1} \sum_{i=1}^{[nu]} X_{j[i:n]} Y_{[i:n]} \), estimates \( S_j(u) \), \( j = 1, 2 \), and \( \hat{M}_n(u) \) estimates \( M(u) \), with components \( \hat{M}_{nj}\tilde{j}(u) = n^{-1} \sum_{i=1}^{[nu]} X_{\tilde{j}[i:n]} X_{j[i:n]}^T \) estimating \( M_{nj}(u) \), \( \tilde{j}, j = 1, 2 \). This suggests test statistics for \( H_0 \) based on some functional of
\[
\hat{n}_n(u) = \left( \hat{b}_n^- - \hat{b}_n^+ \right)(u)
\]
\[
= R \hat{M}_n^{-1}(u) \hat{S}_n(u),
\]
\[
= n_0(u) + R \hat{M}_n^{-1}(u) \hat{N}_n(u), \quad u \in [K/n, 1-K/n),
\]
with \( \hat{N}_n(u) = \left( \hat{N}_{n1}(u), \hat{N}_{n1}(1) - \hat{N}_{n1}(u), \hat{N}_{n2}(1) \right)^T \), \( \hat{N}_{nj}(u) = n^{-1} \sum_{i=1}^{[nu]} X_{j[i:n]} U_{[i:n]} \), \( j = 1, 2 \).

The main difference with respect to UI parameter stability tests in time series regression is that, rather than sequential observations, we use the concomitants (induced order statistics) with respect to \( Z \).

The asymptotic distribution of \( \sqrt{n} \hat{N}_n \) is obtained applying results for CUSUMs of concomitants in Bhattacharya (1974, 1976), extended by Sen (1976), Stute (1993, 1997) and Davydov and Egorov (2000), among others.
Henceforth, for any matrix $A$, $\|A\|^2 = \bar{g}(A^t A)$ is the spectral norm, where $\bar{g}(B)$ is the maximum eigenvalue of the matrix $B$, and "$\rightarrow_d$" means convergence in distribution of random variables, random vectors or random elements in a Skorohod space $\mathbb{D}[a,b]$, $0 \leq a < b \leq 1$, and "$\rightarrow_d$" means convergence in probability. Define $N_\infty(u) = (N_{1\infty}(u), N_{2\infty}(1) - N_{1\infty}(u), N_{2\infty}(1))^T$, where $N_{1\infty}$ is a $k_j \times 1$ vector of a centered Gaussian process with $\mathbb{E}(N_{1\infty}(u)N_{1\infty}^T(v)) = \mathbb{E}(X_jX_j^T U^2 1_{ \{ F_2(Z) \leq u_\wedge v \} } )$, $\ell, j = 1, 2$, and $u, v \in [0,1]$.

Assume,

**A3:** $\mathbb{E}\|X_jU\|^2 < \infty$, $j = 1, 2$.

**Theorem 1:** Assuming A1, A2,

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

and if A3 is also assumed,

$$\sqrt{n} \tilde{N}_n \rightarrow_d N_\infty \text{ in } \mathbb{D}[0,1]. \quad (9)$$

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

$$\sqrt{n} \tilde{n}_n \rightarrow_d \eta_\infty \text{ in } \mathbb{D}[\epsilon, 1 - \epsilon], \text{ for } \epsilon \in (0,1),$$

where $\eta_\infty(u) \overset{d}{=} R M^{-1}(u) N_\infty(u)$, and "$\overset{d}{=}"$ means equality in distribution. Weak convergence of $\sqrt{n} \tilde{n}_n$ 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).

Therefore, the $\eta_\infty$'s matrix of variance and covariance functions is

$$\mathbb{E}(\eta_\infty(u)\eta_\infty^T(v)) = \Sigma_0(u,v) = R M^{-1}(u)\Omega_0(u,v)M^{-1}(v)R^T, \ u, v \in (0,1),$$

with $\Omega_0(u,v) = \mathbb{E}(N_\infty(u)N_\infty^T(v))$. In order to apply the UI testing principle, we must standardize $\tilde{n}_n$. Therefore, we need to estimate,

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

where $\Omega_{ij}(u) = \mathbb{E}(X_jX_j^T U^2 1_{ \{ F_2(Z) \leq u_\wedge v \} } )$, $j = 1, 2$, and $V = Y - X_1^T \tilde{\beta}_0 - X_2^T \tilde{\delta}_0$ are the errors of the best linear predictor of $Y$ given $(X_1, X_2)$. Under $H_0$, $\Omega = \Omega_0$. Assume
A4: \( \text{Rank} \left( \Omega(u, u) \right) = 2k_1 + k_2 \) for all \( u \in (0, 1) \).

The natural estimator of \( \Omega_{ij}(u) \) is \( \hat{\Omega}_{n,ij}(u) = n^{-1} \sum_{i=1}^{[nu]} X_{ij[i:n]} X_{ij[i:n]}^\top \hat{V}_{ij}^2 \), for \( \ell, j = 1, 2 \), where \( \hat{V}_i = Y_i - X_{ij1} \tilde{\beta}_n - X_{ij2} \tilde{\delta}_n \) are the OLS residuals using all the data set, i.e.

\[
\left( \tilde{\beta}_n^\top, \tilde{\delta}_n^\top \right) = \arg \min_{b, d} \sum_{i=1}^{n} (Y_i - X_{ij1} b - X_{ij2} d)^2.
\]

Tests are based on functionals of the empirical process,

\[
\tilde{a}_n(u) = \tilde{\eta}_n^\top(u) \tilde{\Sigma}_n^{-1}(u, u) \tilde{\eta}_n(u), \quad u \in [K/n, 1 - K/n).
\]

where \( \hat{\Sigma}_n(u, u) = R^\top \hat{M}_n^{-1}(u) \hat{\Omega}_n(u, u) \hat{M}_n^{-1}(u) R \), and \( \hat{\Omega}_n(u, u) \) is the estimator of \( \Omega(u, u) \) with components \( \hat{\Omega}_{n,ij}(u) \), \( \ell, j = 1, 2 \). In order to show that \( \hat{\Omega}_n \) is consistent, we also need to assume

A5: \( \mathbb{E} \| X_j \|^4 < \infty, \quad j = 1, 2 \), and \( \mathbb{E} \| V \|^4 < \infty \).

The test rejects \( H_0 \) for large values of

\[
\hat{\varphi}_{ne} = n \max_{K + |uj| \leq \lfloor n(1 - \epsilon) \rfloor - K} \hat{\alpha}_n \left( \frac{j}{n} \right), \quad \text{for } \epsilon \in (0, 1/2).\]

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

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

where,

\[
\left\{ \alpha_{\infty}(u) \right\}_{u \in (0, 1)} = \left\{ \eta_{\infty}^\top(u) \Sigma_0^{-1}(u, u) \eta_{\infty}(u) \right\}_{u \in (0, 1)}.
\]

Theorem 2: Assume A1 – A5. For \( \epsilon \in (0, 1/2) \), under \( H_0 \),

\[
\hat{\varphi}_{ne} \rightarrow^d \varphi_{\infty}.
\]

Therefore, a test with significance level \( \alpha \) is given by the binary random variable

\[
\hat{\Phi}_{ne}(\alpha) = 1_{\{\hat{\varphi}_{ne} > c_{\epsilon}(\alpha)\}}, \quad \text{where } c_{\epsilon}(\alpha) \text{ is the } (1 - \alpha) - \text{th quantile of } \varphi_{\infty}.
\]

Next, we study the power of the test. Consider local alternatives,

\[
H_{n1}: \beta(Z) = \beta_0 + \frac{\tau(Z)}{\sqrt{n}} \text{ a.s.},
\]

for a vector of constants \( \beta_0 \) and an unknown function \( \tau : \mathbb{R} \rightarrow \mathbb{R}^{k_1} \) such that,
A6: $\mathbb{E}\|X_jX_1^\top\tau(Z)\| < \infty$, for $j = 1, 2$.

Define the random processes,

$$\left\{ \alpha_\infty^j(u) \right\}_{u \in (0,1)} \overset{d}{=} \left\{ \eta_\infty^j(u) \Sigma_0^{-1}(u, u) \eta_\infty^j(u) \right\}_{u \in (0,1)},$$

with $\left\{ \eta_\infty^j(u) \right\}_{u \in (0,1)} \overset{d}{=} \left\{ RM^{-1}(u) (N_\infty + T)(u) \right\}_{u \in (0,1)}$, $T(u) = [T_1^j(u), T_1^j(1) - T_1^j(u), T_j^2(1)^\top]$ and $T_j(u) = \mathbb{E}[X_jX_1^\top\tau(Z)1_{(F_Z(Z) \leq u)}]$ for $j = 1, 2$.

**Theorem 3:** Assume A1 – A6. For $\epsilon \in (0, 1/2)$, under $H_{1\eta}$,

$$\hat{\varphi}_{ne} \to_p \infty,$$

(11)

and under $H_{n1}$,

$$\hat{\varphi}_{ne} \to \sup_{u \in [1, 1-\epsilon]} \alpha_\infty^1(u).$$

(12)

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

$$\gamma(u) = T^\tau(u)M^{-1}(u)R^\tau\Sigma_0^{-1}(u, u)RM^{-1}(u)T(u).$$

This suggests choosing $\epsilon$ 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 $\phi_0 = (\phi_{01}, \ldots, \phi_{ok_1})^\top$ in a discontinuous regression design (RDD) model with $\beta_0(z) = (\beta_{01}(z), \ldots, \beta_{ok_1}(z))^\top$ and $\beta_{0j}(z) = \phi_{0j}1_{(z \leq \pi_0)}$, $j = 1, \ldots, k_1$, where $\pi_0$ is a nuisance parameter. They use this example to illustrate conditions for asymptotic optimal tests when some nuisance parameter is only present under the alternative. The hypothesis of interest is $H_0$ in the direction $H_{n\phi}: \phi_0 = \kappa_0^\top/\sqrt{n}$ for some $\kappa_0 \in \mathbb{R}^{k_1}$, assuming that $U$ is independent of $(X_1, X_2, Z)$. This approach suggests using as test statistic

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

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

$$\lim_{c \to 0} \frac{\hat{\rho}_{nG_\epsilon}(c) - 1}{c} = n \int_0^1 \hat{\alpha}_n(u) dG_\epsilon(u).$$
is suitable for alternatives $H_{n,:}$ close to the null, while $\varphi_{ne}$ is designed to detect distant alternatives.

The distribution of $\varphi_{\infty}$ 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 that imposes the restriction $H_0$ based on OLS residuals using the sample whole $\{\hat{V}_i\}_{i=1}^n$. The bootstrap resample is $\{Y_i^*, W_i\}_{i=1}^n$, with $Y_i^* = X_i^T \beta_n + \hat{X}_i^T \delta_n + \hat{V}_i^*$, and $\hat{V}_i^* = \hat{V}_i^\xi_i$, where $\{\xi_i\}_{i=1}^n$ are i.i.d. as $\xi$, and independent of $\{Y_i, W_i\}_{i=1}^n$, such that,

**A7:** $E_\xi (\xi) = 0$, $E_\xi (\xi^2) = 1$ and $|\xi| \leq C < \infty$ a.s.

Henceforth, $P_\xi$ denotes the probability function of $\xi$, $E_\xi$ the corresponding expectation, and $C$ is a generic constant. Notice that, $E_\xi(\hat{V}_i^*) = 0$ and $E_\xi(\hat{V}_i^{*2}) = \hat{V}_i^2$.

The bootstrap version of $\hat{\theta}_n(u)$ based on $\{Y_i^*, W_i\}_{i=1}^n$ is

$$\hat{\theta}_n(u) = \left( \hat{b}_n^{-*T}(u), \hat{b}_n^{+*T}(u), \hat{d}_n^{*T}(u) \right)^T = \arg \min_{b^-, b^+, d} \left\{ \sum_{i=1}^{[mu]} (Y_{i:n}^* - X_{1i:n}^* b^- - X_{2i:n}^* d)^2 + \sum_{i=1+[mu]}^{n} (Y_{i:n}^* - X_{1i:n}^* b^+ - X_{2i:n}^* d)^2 \right\}$$

$$= (\hat{\beta}_n^T, \hat{\beta}_n^T, \hat{\delta}_n^T) + \hat{M}_n^{-1}(u) \hat{N}_n(u), \ u \in (0, 1),$$

with $\hat{N}_n(u) = \left( \hat{N}_n^{*1}(u), \hat{N}_n^{*1}(1) - \hat{N}_n^{*1}(u), \hat{N}_n^{*1}(1)^T \right)$, and $\hat{N}_n(u) = n^{-1} \sum_{i=1}^{[mu]} X_{j:n} \hat{V}_{i:n}$, $j = 1, 2$. The bootstrap test statistic is

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

where $\hat{\alpha}_n^*(u) = \hat{\eta}_n^{*1}(u) \hat{\Sigma}_n^{-1}(u, u) \hat{\eta}_n^{*1}(u)$ and $\hat{\eta}_n(u) = R \hat{\theta}_n^*(u) = R \hat{M}_n^{-1}(u) \hat{N}_n(u)$. The bootstrap critical value at the $\alpha$ level of significance is

$$\hat{c}_{ne}(\alpha) = \inf \left\{ c : P_\xi (\hat{\varphi}_{ne} \leq c) \geq 1 - \alpha \right\}.$$

So, the bootstrap test is given by the binary variable $\hat{\Phi}_{ne}(\alpha) = 1_{\{\hat{\varphi}_{ne} > \hat{c}_{ne}(\alpha)\}}$. The next theorem justifies the bootstrap test.
Theorem 4: Assume A1 – A5, and A7. For \( \epsilon \in (0, 1/2) \), under \( H_0 \),

\[
\lim_{n \to \infty} \mathbb{P}_\xi (\hat{\varphi}_{ne}^* \leq c) = \mathbb{P}(\varphi_{\infty} \leq c) \quad \text{a.s.,}
\]

for any \( c > 0 \). And there exists a \( C < \infty \) such that, under \( H_1 \),

\[
\lim_{n \to \infty} \mathbb{P}_\xi (\hat{\varphi}_{ne}^* > C) = 0 \quad \text{a.s.}
\]

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

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

**Algorithm 1**

**i.** Generate \( b \) sets of random numbers \( \{ \xi^{(j)} \}_{j=1}^n \) i.i.d. as \( \xi \), and the corresponding resamples \( \{ Y_i^{(j)}, W_i \}_{i=1}^n, j = 1, \ldots, b \), with \( b \) large.

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

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

\[
\hat{c}_{ne}^{(b)*}(\alpha) = \inf \left\{ c : \frac{1}{b} \sum_{j=1}^b 1\{ \hat{\varphi}_{ne_j}^{(b)*} < c \} \geq 1 - \alpha \right\},
\]

and the corresponding \( p \)-values, \( \hat{p}_{ne}^* \), by

\[
\hat{p}_{ne}^{(b)*} = \frac{1}{b} \sum_{j=1}^b 1\{ \hat{\varphi}_{ne_j}^{(b)*} \geq \hat{\varphi}_{ne} \}.
\]

**iv.** Use the test \( \hat{\Phi}_{ne}^{(b)} (\alpha) = 1\{ \hat{\varphi}_{ne} > \hat{c}_{ne}^{(b)*}(\alpha) \} \).

The greater the \( b \), the better the bootstrap approximations.

When the alternative hypothesis is non-parametric, one should choose the smaller possible \( \epsilon \) in order to detect as many alternatives as possible, but a too small \( \epsilon \) may produce serious size distortions. In order to keep the type I error under control, given a nominal level \( \alpha \), we can choose the smallest \( \epsilon \) 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, \( \omega \), as a function of \( \epsilon \), 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 \((Y, W)\), \(F\), 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 \(\epsilon\). 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 \(\{Y_i, W_i\}_{i=1}^n\), but we could use the wild bootstrap resamples that impose \(H_0\) instead, e.g. as in Algorithm 1 step i.

In order to save computing time and choosing \(\epsilon\) 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 \(\hat{n}\), that ensures an error level less or equal to \(e_0\) into a given interval \([1/n, \ell_0/n]\) for suitably chosen small \(\ell_0\), e.g. \(\ell_0 = \lfloor n/3 \rfloor\). Such \(\hat{\epsilon}_n\) may not exist, in which case we choose the \(\epsilon\) minimizing the error level in the interval.

**Algorithm 2**

i. Fix \(e_0\) and \(\ell_0\).

ii. Fix \(b_0\) and generate i.i.d. resamples \(\left\{ Y_i^{(j)}, W_i \right\}_{i=1}^n \) from \(\hat{F}_n\), \(j = 1, \ldots, b_0\).

iii. Set \(\ell := 1\).

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

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

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

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

Then, we use the test \(\hat{\Phi}_{n\hat{\epsilon}_n}^{*(b)}(\alpha)\) in Algorithm 1. This data-driven trimming choice is computationally expensive. Of course, we could choose \(\hat{\epsilon}_n\) minimizing \(\hat{\epsilon}_n (\epsilon)\) over the whole interval, but it will be even more costly, and the resulting \(\hat{\epsilon}_n\) will probably be larger. In this respect, the bigger \(e_0 (\ell_0)\), the smaller (bigger) computational cost. A formal justification of the test \(\hat{\Phi}_{n\hat{\epsilon}_n}^{*(b)}(\alpha)\) is beyond the scope of this paper, but we show in the next section (Table 8) that it works very well in practice.
Under the 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 $(U, X_1)$, $\mathbb{E}(U^2 | X_1) = \sigma^2$, and $M_{11}(1)$ is non-singular.

This assumption is not acceptable in practice, but allows us 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 $\epsilon$ 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,

$$\eta_{11}(u) = RM^{-1}(u)N_{11}(u) = M_{11}^{-1}(1)\frac{N_{11}(u) - uN_{11}(1)}{u(1-u)} \text{ a.s.}, \tag{13}$$

and $\{N_{11}(u)\}_{u \in (0,1)} \overset{d}{=} \{\sigma \cdot M_{11}^{1/2}(1)B_0(u)\}_{u \in (0,1)}$, where $B_0$ is a vector of independent Brownian bridges, i.e. $B_0$ is a Gaussian process with mean zero and $\mathbb{E}(B_0(u)B_0^*(u)) = (u \wedge v - uv) \cdot I_{k_1}$, for all $u, v \in (0, 1)$. Henceforth,

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

which has been tabulated, for different $\epsilon$’s values, by James et al. (1987) for $B_0$ scalar, and by Andrews (1993) in the multivariate case.

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

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

with

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

which resembles the classical UI tests, but without trimming. This statistic, suitably transformed, 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:** $\mathbb{E} \|X_1\|^{2+\delta} < \infty$ and $\mathbb{E} |U|^{2+\delta} < \infty$ for some $\delta > 0$.

These types of moment conditions were proposed by Shorak (1979), relaxing those in Darling and Erdős (1956). We also consider,

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

$$\hat{\varphi}_n^{(2)} = \max_{K \leq j < n-K} \frac{j(n-j)}{n} \tilde{\alpha}_n \left( \frac{j}{n} \right).$$
Henceforth, $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$, and $E$ is a random variable such that $\Pr(E \leq x) = \exp(-2\exp(-x))$, $a(x) = \sqrt{2\log x}$ and $b_m(x) = 2\log x + (m/2)\log \log x - \log \Gamma(m/2)$.

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

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

$$a(\log n)\sqrt{\tilde{\varphi}^{(0)}_n} - b_{n1}(\log n) \to_d E,$$  

$$\tilde{\varphi}^{(1)}_n \to_d \int_0^1 B_0(u)B_0(u)\frac{du}{u(1-u)},$$  

$$\tilde{\varphi}^{(2)}_n \to_d \sup_{u \in (0,1)} B_0^T(u)B_0(u).$$

This suggests that the asymptotic distribution of $\tilde{\varphi}_{ne}$ 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 $\tilde{\eta}_n$’s transformation,

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

Thus, by (12), (25) and (26), under $H_{n1}$, $\sqrt{n}\tilde{\eta}_n \to_d \tilde{\eta}_{\infty} \overset{d}{=} B_0 + \omega_0$ with $\omega_0(u) = \mathbb{E}[(\tau(Z)1_{u < Z < 1 - u})]$. The principal components random vectors of $B_0$ are

$$\zeta_j = \frac{1}{\vartheta_j} \int_0^1 B_0(u)\mu_j(u)du,$$

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, $\{\zeta_j\}_{j \in \mathbb{N}}$ are i.i.d. $N_{k_1}(0, I_{k_1})$. The sample version of $\zeta_j$ is

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

The distribution of the infinite dimensional random vectors $\zeta = (\zeta_j)_{j > 0}$ and $\hat{\zeta}_n = (\hat{\zeta}_{nj})_{j > 0}$ are uniquely determined by their finite dimensional distributions. Thus, by the continuous mapping theorem, $\hat{\zeta}_n \to_d \zeta$ under $H_0$, and $\hat{\zeta}_n \to_d \zeta + \rho_0$ under $H_{n1}$, where
\( \mathbf{\rho}_0 = (\mathbf{\rho}_{0j})_{j > 0} \) with \( \mathbf{\rho}_{0j} = \vartheta_j^{-1/2} \int_0^1 \mathbf{\omega}_0(u) \mu_j(u) du \). This suggests a Neyman-type test that rejects \( H_0 \) for large values of

\[
\hat{Q}_{nm} = \sum_{j=1}^m \mathbf{x}_n^T \mathbf{\hat{c}}_n,
\]

for a fixed \( m \). Thus, under \( H_0 \), \( \hat{Q}_{nm} \rightarrow_d \chi^2_{m-1} \) and under \( H_n \), \( \hat{Q}_{nm} \rightarrow_d \chi^2_{m-1} \left( \sum_{j=1}^m \mathbf{\rho}_{0j}^T \mathbf{\rho}_{0j} \right) \),

where \( \chi^2_{\ell}(\Lambda) \) denotes a non-centered chi-square with \( \ell \) degrees of freedom and non-centrality parameter \( \Lambda \). 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 also 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 \( H_{n1} \) at the significance level \( \alpha \),

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

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

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

### 4. FINITE SAMPLE PROPERTIES

We generate samples \( \{Y_i, Z_i, 1, X_{12i}, ..., X_{1k_1 i}, X_{21i}, ..., X_{2k_2 i}\}_{i=1}^n \) with

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

\( \{Z_i\}_{i=1}^n \) 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 \([0, 1]\), \( \ell = 1, 2 \), and

\[
U_i = \frac{\varepsilon_i \exp(\kappa Z_i/2)}{\sqrt{\text{Var}(\varepsilon_i \exp(\kappa Z_i/2))}},
\]
with \( \varepsilon_i \) i.i.d. \( N(0,1) \); that is, \( \text{Var}(U_i) = 1 \), and \( \kappa \) governs how severe the conditional heteroskedasticity is. We generate random coefficients 

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

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

a) \( f(z) = z \),

b) \( f(z) = [1 + \exp(-\rho z)]^{-1} \),

c) \( f(z) = \sin(2\pi z) \),

d) \( f(z) = 1 + 2 \cdot 1_{\{z \leq \pi_0\}} \).

Model a) is a simple linear model and b) is a nonlinear alternative. The two models are almost indistinguishable for \( \rho = 1 \) when \( z \in (0,1) \). The lower the \( \rho \), 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 \( \pi_0 = 0.4 \) in all simulations, but we have also tried other values and the results do not change substantially except when the jump is placed in fairly low quantiles (\( \pi_0 \leq 0.1 \)). Figure 1 represents \( \eta_0 \) for the different models and different \( \lambda \) values.

The simulation study is implemented to provide evidence on the \( \epsilon \)'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 \( \epsilon \) is close to zero, but the level accuracy is excellent for \( \epsilon \) around 0.1. The power does not change much for \( n = 100 \), and is close to 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 \( \pi_0 \) 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 \( \epsilon \) 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 \( \lambda \).

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 under A8, using statistics $\hat{\phi}_n^{(j)}$ 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 $\hat{\phi}_n^{(0)}$ shows poor size properties compared to the others, particularly for small $n$. However, the level accuracy of the asymptotic tests based on $\hat{\phi}_n^{(1)}$ and $\hat{\phi}_n^{(2)}$ 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 multivariate CUSUM of residuals type process

$$
\hat{\psi}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i \prod_{j=2}^{k_1} 1\{x_{1j} \leq x_j\} \prod_{m=1}^{k_2} 1\{x_{2m} \leq x_{k_1 + m}\}, \quad x = (x_2, \ldots, x_{k_1 + k_2})^T,
$$

using the Kolmogorov-Smirnov type statistic,

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

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 $\hat{T}_n = (\text{RSS}_0/\text{RSS}_1) - 1$ that compares the restricted and unrestricted sums of squared residuals, $\text{RSS}_0$ and $\text{RSS}_1$, respectively. The unrestricted estimates use kernels and 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, and they recommend a bootstrap test. 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 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 reports results for $k_1 = 2, 3; k_2 = 1; 3; 4$; $k_1 = 1; 3; 4$, and $X = 1$. 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 $H_0$.

In the next set of simulations we apply the test to check the linearity hypothesis when $k_1 = 1$, $k_2 = 1$ and $X = Z$. That is, $H_0$ is equivalent to omnibus specification testing of the simple regression model $E(Y|Z) = \beta_{01} + Z\delta_{01}$ a.s. Our test is omnibus for the linear regression specification hypothesis, and competes with the CUSUM test. Since $\beta_{01}$ 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 $\rho$ values. Table 4 shows that our test rejects almost double than the CUSUM for $\rho$ large.

We also consider the test for model checking of nonlinear regression models. We consider omnibus specification testing of $E(Y|Z) = \beta_{01} + \sum_{\ell=1}^{L} Z^\ell\delta_{0\ell}$ a.s. This corresponds to applying our test to model (2) with $k_1 = 1$, and $g_j(z) = z^j$, $j = 1, \ldots, L$. 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 $L = 4$.

Next, we consider the performance of the test as a specification test of the 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 $X_1$ have a particular functional form. In particular, that $\beta_{0j} (Z) = \tilde{\beta}_{0j} + \delta_{0j} Z$ 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 $\rho$ values, $k_2 = 0$ and $k_1 = 2, 3, 4$. Our test performs better than CUSUM, particularly for $\rho$ large.

**TABLE 6 ABOUT HERE**

Now, we 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 $g_j(z) = z^j$, $j = 1, \ldots, L$. We consider testing $\beta_{0j} (Z) = \tilde{\beta}_{0j} + \sum_{\ell=1}^{L} Z^\ell \delta_{0(j+\ell-1)}$ 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 $\epsilon$ described in Algorithm 2. We set $\ell_0 = \lfloor n/3 \rfloor$ and $e_0 = 10^{-3}$. We have used resamples $\{Y_4^{(i)}, W_i\}_{i=1}^{b_0}$, $j = 1, \ldots, b_0$ 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 $\epsilon$ is compared with different prespecified values of $\epsilon$. Table 8 provides the rejection rate under $H_0$ with $k_1 = 2$, $k_2 = 1$ 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

$$\log WAGE = \tilde{\beta}_{01} + \tilde{\beta}_{02} EDUC + \tilde{\beta}_{03} IQ + X_2^\top \delta_0 + U,$$  \hspace{1cm} (17)
where $WAGE$ are USD monthly earnings, $EDUC$ are years of education, and $X_2^T = (EXPER, TENURE, MARRIED, SOUTH, URBAN, BLACK)^T$, $EXPER$ 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 $\delta_0 = (\delta_{01}, ..., \delta_{06})^T$. The "ability" is in the error term $U$, which is correlated with $EDUC$ for obvious reasons. The variable $IQ$ in model (17) is a valid proxy of ability if $E(U|EDUC, IQ, X_2) = E(U|IQ, X_2)$.

In this case, the partial effect of $EDUC$ on $WAGE$, $\beta_{02}$, can be consistently estimated using OLS in (17), though estimators of $\beta_{03}$ and $\delta_0$ are typically inconsistent. The OLS estimators of $\beta_{02}$ and $\beta_{03}$ 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 between $EDUC$ and $IQ$, or when $IQ$ enters nonlinearly into the model.

A reasonable alternative to (17) is the following SVC model,

$$
\log (WAGE) = \beta_{01}(IQ) + \beta_{02}(IQ) \cdot EDUC + X_2^T\delta_0 + U.
$$

Figure 4 provides estimates of $\beta_{01}$ and $\beta_{02}$ in (18) 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 $\beta_{0j}(IQ) = \beta_{0j}^{(1)} + \beta_{0j}^{(2)}IQ + \beta_{0j}^{(3)}IQ^2$, $j = 1, 2$.

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. We also report the smoothing LR test of Cai et al. (2000).

Here, the omnibus test based on $\hat{\phi}_n$ 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

$$
\log (WAGE) = (\beta_{01}(IQ) + \delta_{07}IQ) + (\beta_{02}(IQ) + \delta_{08}IQ) \cdot EDUC + X_2^T\delta_0 + U.
$$
In this case, see Table 10, we are unable to reject the specification of the interactive effects 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 $Z = (Z_1, ..., Z_q)^T$, i.e.

$$Y = X_1^T \beta_0(Z) + X_2^T \delta_0 + U,$$

with $\beta_0 : \mathbb{R}^q \to \mathbb{R}^{k_1}$, the test requires ordering the data according to $Z$ somehow. In this scenario, single-index models have proven to be an efficient way of coping with the data ordering issue, i.e. $Z = g_{\psi_0}(Z)$ a.s. in (1), where $g_{\psi_0} : \mathbb{R}^q \to \mathbb{R}$ is a known function, and $\psi_0 \in \Psi \subset \mathbb{R}^q$ is an unknown parameter vector. Xia and Li (1999) proposed a $\sqrt{n}$-consistent estimator of $\psi_0$, $\hat{\psi}_n$, using kernel smoothing. The test can be implemented by ordering the data according to $\hat{Z}_{ni} = g_{\hat{\psi}_n}(Z_i)$, $i = 1, ..., n$, and using the corresponding
concomitants in (6). However, the corresponding test statistic will converge to a random variable with a different distribution than \( \phi_{\infty} \) under \( H_0 \), because of the \( \psi_0 \)'s estimation effect. The test can still be implemented using wild bootstrap. Of course, the resulting test depends on the amount of smoothing chosen for estimating \( \psi_0 \). 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 \( F_Z(Z) \overset{d}{=} U(0,1) \) under \( A1 \), we assume w.l.o.g that \( Z \overset{d}{=} U[0,1] \).

**Proof of Theorem 1.** A typical uniformity argument shows that

\[
\sup_{u \in (0,1)} \left\| \left( \hat{M}_{tj} - M_j \right) (u) \right\| = o(1) \text{ a.s.}
\]

with

\[
\hat{M}_{tj}(u) = \frac{1}{n} \sum_{i=1}^{n} X_{ti} X_{ji} 1_{\{Z_i \leq u\}}, \quad \ell, j = 1, 2.
\]

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

Henceforth, \( \tilde{\theta}_0 = (\tilde{\beta}_0, \tilde{\beta}_0^T, \delta_0^T)^T \) and \( \tilde{\theta}_n = (\tilde{\beta}_n^T, \tilde{\beta}_n^T, \tilde{\delta}_n^T)^T \).

**Proof of Theorem 2.** Define \( \hat{\Omega}_{n}^0 = n^{-1} \sum_{i=1}^{[nu]} X_{j1} X_{j1}^T V_i^2 \), \( \ell, j = 1, 2 \). First, notice that \( \tilde{\theta}_n = \tilde{\theta}_0 + o(1) \) a.s. by (8) and \( n^{-1} \sum_{i=1}^{[nu]} X_{ij} V_i = o(1) \) a.s. under \( A5 \). Then, applying the same arguments to prove (8)

\[
\sup_{u \in (0,1)} \left\| \left( \hat{\Omega}_{n\ell j} - \hat{\Omega}_{n\ell j}^0 \right) (u) \right\| = \sup_{u \in (0,1)} \left\| \frac{1}{n} \sum_{i=1}^{[nu]} X_{ij} X_{i\ell}^T \left( \hat{V}_i^2 - V_i^2 \right) \right\| = o(1) \text{ a.s., (19)}
\]

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

\[
\sup_{u \in (0,1)} \left\| \left( \hat{\Omega}_0 - \Omega \right) (u) \right\| = o(1) \text{ a.s., (20)}
\]

and theorem follows applying (9) and the CMT.

Henceforth, \( X_i(u) = [X_i^T(u), X_i^T(1) - X_i^T(u), X_i^{[1]}_2] \), with \( X_i^T(u) = X_i^T 1_{\{Z_i \leq Z_{n:[nu]}\}} \).
Proof of Theorem 3. Under $H_{1\eta}$, by Theorem 1,
\[
\frac{\hat{\varphi}_{n\epsilon}}{n} \xrightarrow{p} \sup_{\epsilon \leq u \leq 1-\epsilon} \eta_0^*(u) [RM^{-1}(u)\Omega(u,v)M^{-1}(v)]^{-1} R^T \eta_0(u) > 0,
\]
which proves (11). In order to prove (12), notice that under $H_{n1}$,
\[
\sqrt{n} \left( \tilde{\theta}_n - \theta_0 \right) (u) = \tilde{\mathcal{M}}_n^{-1}(u) \left[ \frac{1}{n} \sum_{i=1}^{n} X_i(u)X_i^T(u) \cdot (\tau^T(Z_i), \tau^T(Z_i), 0)^T + \sqrt{n} \tilde{\mathcal{N}}_n(u) \right],
\]
and, under A6,
\[
\sup_{u \in (0,1)} \left\| n^{-1} \sum_{i=1}^{n} X_i(u)X_i^T(u) \cdot (\tau^T(Z_i), \tau^T(Z_i), 0)^T - T(u) \right\| = o(1) \text{ a.s.}
\]
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. Notice that uniformly in $u \in (0,1)$,
\[
\hat{\eta}_n^*(u) = R [M (u) + o(1)]^{-1} \tilde{\mathcal{N}}_n(u) \text{ a.s.},
\]
Following the strategy of proof in Stute et al. (1998) (SGQ), the theorem follows by showing that, conditional to the sample, $\sqrt{n}\tilde{\mathcal{N}}_n^*$ converges in distribution to $\mathcal{N}_1^\infty$ a.s., i.e. for almost all sample $\{Y_i, W_i\}_{i=1}^n$, both under $H_0$ and under $H_1$, where $\mathcal{N}_1^\infty$ is a Gaussian process centered at zero and with matrix of variance and covariance functions
\[
E \left[ \mathcal{N}_1^\infty (u_1) \mathcal{N}_1^\infty (u_2) \right] = \Omega (u_1, u_2), \text{ } u_1, u_2 \in (0,1), \text{ i.e. } \mathcal{N}_1^\infty \stackrel{d}{=} \mathcal{N}_\infty \text{ under } H_0.
\]
To this end, it suffices to show the convergence of the finite dimensional distributions (fdis) and tightness. For fdis convergence, first notice that for $u_1, u_2 \in (0,1)$, $E_{\xi} \left[ \tilde{\mathcal{N}}_n^*(u_1) \right] = 0$ and
\[
n \cdot E_{\xi} \left[ \tilde{\mathcal{N}}_n^*(u_1) \tilde{\mathcal{N}}_n^*(u_2) \right] = \hat{\Omega}_n (u_1, u_2) = \Omega (u_1, u_2) + o(1) \text{ a.s.}
\]
using (19) and (20). Then, fixing $u_1, \ldots, u_q \in (0,1)$, by the Cramér-Wold device, it suffices to show that,
\[
\mathbb{P}_{\xi} \left\{ \sqrt{n} \sum_{j=1}^{q} c_j^T \tilde{\mathcal{N}}_n^*(u_j) \leq \epsilon \right\} \rightarrow \mathbb{P} \left\{ \sum_{j=1}^{q} c_j^T \mathcal{N}_\infty^1 (u_j) \leq \epsilon \right\} \text{ a.s.,} \tag{21}
\]
for any $\epsilon < \infty$, $c_j \in \mathbb{R}^{2k_1+k_2}$, $j = 1, \ldots, q$. Write $\vartheta_i = \sum_{j=1}^{q} a_j c_j^T X_i(u_j)$; then,
\[
\sqrt{n} \sum_{j=1}^{q} a_j c_j^T \tilde{\mathcal{N}}_n^*(u_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \vartheta_i \tilde{V}_i \xi_i.
\]
Hence, (21) follows by checking the Linderberg condition

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

Since $|\xi| \leq C < \infty$, for a generic constant $C$,

$$L_n(\delta) \leq \frac{C^2}{n} \sum_{i=1}^{n} \theta_i^2 \hat{V}_i^2 \mathbb{1}_{\{ |\theta_i| \geq \delta \sqrt{n} \}}$$

$$= \frac{C^2}{n} \sum_{i=1}^{n} \theta_i^2 \hat{V}_i^2 \mathbb{1}_{\{ |\theta_i| \geq \frac{C}{n} \}} + o(1) \text{ a.s.}$$

$$= o(1) \text{ a.s. as } n \to \infty.$$ 

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

$$\hat{\omega}_{n_{bi}}^*(u) = \sqrt{n} c^T \hat{N}_{n}^T(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c^T X_i(u) \hat{V}_i \xi_i.$$ 

We must show, as in SGQ Lemma 3, that for any $c = (c_1, c_2, c_3)^T$, $c_1, c_2 \in \mathbb{R}^{k_1}$, $c_3 \in \mathbb{R}^{k_2}$ and $K/n \leq u_0 \leq u_1 \leq u_2 < 1 - K/n$,

$$\mathbb{E}_{\xi} \left\{ [\hat{\omega}_{n_{bi}}^*(u_1) - \hat{\omega}_{n_{bi}}^*(u_0)]^2 [\hat{\omega}_{n_{bi}}^*(u_2) - \hat{\omega}_{n_{bi}}^*(u_1)]^2 \right\} \leq C \left[ \hat{G}_n(u_2) - \hat{G}_n(u_0) \right]^2,$$  \hspace{1cm} (22)

\( \hat{G}_n \) is monotone, and \( \hat{G}_n \to G \) a.s. First notice that for any \( u_\ell \geq u_j, u_\ell, u_j \in [K/n, 1 - K/n] \),

$$\hat{\omega}_{n_{bi}}^*(u_\ell) - \hat{\omega}_{n_{bi}}^*(u_j) = (c_1 - c_2)^T X_{i_{\ell}} \hat{V}_i \xi_i \mathbb{1}_{\{ z_{(n_\ell j : i_{\ell})}^j \leq z_i \leq z_{(n_\ell j : i_{\ell})} \}}.$$ 

Then, applying Stute’s (1997) Lemma 5.1,

$$LHS(22) \leq \frac{3}{n^2} \sum_{i \neq j} \mathbb{E}_{\xi} \lambda_i \gamma_j^2,$$

$$\lambda_i = \|X_{i_{\ell}}\| \|\hat{V}_i \xi_i \mathbb{1}_{\{ z_{(n_\ell j : i_{\ell})}^j \leq z_i \leq z_{(n_\ell j : i_{\ell})} \}}, \; \gamma_i = \|X_{i_{\ell}}\| \|\hat{V}_i \xi_i \mathbb{1}_{\{ z_{(n_\ell j : i_{\ell})}^j \leq z_i \leq z_{(n_\ell j : i_{\ell})} \}}.$$ 

Then,

$$LHS(22) \leq \frac{C}{n^2} \sum_{i \neq j} \|X_{i_{\ell}}\|^2 \|X_{i_{j}}\|^2 \|\hat{V}_i \xi_i \|^2 \mathbb{1}_{\{ z_{(n_\ell j : i_{\ell})}^j \leq z_i \leq z_{(n_\ell j : i_{\ell})} \}} \mathbb{1}_{\{ z_{(n_\ell j : i_{\ell})}^j \leq z_i \leq z_{(n_\ell j : i_{\ell})} \}}$$

$$\leq C \left[ \frac{1}{n} \sum_{i=1}^{n} \|X_{i_{\ell}}\|^2 \|\hat{V}_i \xi_i \|^2 \mathbb{1}_{\{ z_{(n_\ell j : i_{\ell})}^j \leq z_i \leq z_{(n_\ell j : i_{\ell})} \}} \right]^2 \leq C \left[ \hat{G}_n(u_2) - \hat{G}_n(u_0) \right]^2,$$ 

and

$$\hat{G}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \|X_{i_{\ell}}\|^2 \|\hat{V}_i \xi_i \|^2 \mathbb{1}_{\{ z_i \leq z_{(n_\ell j : i_{\ell})} \}}.$$
is monotone and \( \sup_{u \in (0, 1)} \left\| \left( \hat{G}_n - G \right)(u) \right\| = o(1) \) a.s., with \( G(u) = \mathbb{E}(\|X_1\|^2 U^2 1(Z \leq u)) \) a.s., using a Glivenko-Cantelli argument as in the proof of Theorem 1.

**Proof of theorem 5.** Define

\[
\hat{\theta}_n^T(u) = M_{11}^{-1}(1) \frac{\hat{N}_n(u) - u \hat{N}_n(1)}{u(1 - u)},
\]

\[
\hat{\alpha}_n^T(u) = \Theta_n^T(u) \frac{M_{11}^{-1}(1) u(1 - u)}{\sigma^2} \hat{\theta}_n^T(u).
\]

Now notice that

\[
u(1 - u) M_{11}(1) (\hat{\theta}_n - \hat{\theta}_n^T)(u) = (1 - u) \left( u M_{11}(1) \hat{M}_{11}^{-1}(u) - I_{k_1} \right) \hat{N}_n(u) + u \left( (1 - u) M_{11}(1) \left[ \hat{M}_{11}(1) - \hat{M}_{11}(u) \right]^{-1} - I_{k_1} \right) \left( \hat{N}_n(1) - \hat{N}_n(u) \right).
\]

Let \( \{c_n\}_{n \geq 1} \) be a sequence such that \( nc_n \to \infty \) as \( n \to \infty \). Applying Theorem 0 in Wellner (1978) to this context,

\[
sup_{k_1/n + c_n \leq u < (n - k_1)/n} \left\| M_{11}(1) u \hat{M}_{11}^{-1}(u) - I_{k_1} \right\| = o_p(1) \tag{24}
\]

\[
sup_{k_1/n \leq u < (n - k_1)/n - c_n} \left\| (1 - u) M_{11}(1) \left[ \hat{M}_{11}(1) - \hat{M}_{11}(u) \right]^{-1} - I_{k_1} \right\| = o_p(1).
\]

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

\[
sup_{k_1/n + c_n \leq u < (n - k_1)/n - c_n} \left\| u(1 - u) M_{11}(1) \left( \hat{\theta}_n^T - \hat{\theta}_n \right)(u) \right\| = o_p \left( \frac{1}{\sqrt{n}} \right). \tag{25}
\]

This implies that

\[
\max_{k_1/n + c_n \leq j < n - k_1 - n c_n} \left| \frac{j(n - j)}{n} \left( \hat{\alpha}_n^T - \hat{\alpha}_n \right) \left( \frac{j}{n} \right) \right| = o_p(1)
\]

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

\[
\tilde{\varphi}_n^{(2)} = \max_{k_1/n + c_n \leq j < n - k_1 - n c_n} \frac{j(n - j)}{n} \hat{\alpha}_n \left( \frac{j}{n} \right),
\]

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

\[
\sqrt{n} M_{11}^{1/2}(1) \frac{u(1 - u)}{\sigma} \hat{\theta}_n^T \to_d B_0 \text{ in } [0, 1]. \tag{26}
\]

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

\[
\tilde{\varphi}_n^{(1)} \to_d \int_0^1 \frac{B_0^T(u) B_0(u)}{u(1 - u)} du,
\]

27
which proves (15). Finally, (14) is proved applying Shorack’s (1979) extension of the Darling-Erdős theorem to the vector case, as in Horváth (1993).

REFERENCES


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


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


Figure 1: Representation of $\eta_0(z)$ for different models when $\lambda = 0$ (blue curve), $\lambda = 0.25$ (purple curve), and $\lambda = 0.5$ (red curve).

Figure 2: Representation of $\hat{\Phi}_n^*(\alpha)$ for the null and different alternatives.
Figure 3: Representation of $\hat{\Phi}_{n \varepsilon}(\alpha)$ under alternative d) for different $\pi_0$ values.

Figure 4: Representation of $\beta_{01}(IQ)$ and $\beta_{02}(IQ)$ for the estimates of the varying coefficients using kernels (red curve), and OLS estimates of the parametric specification (purple curve).
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$1%$</th>
<th>$5%$</th>
<th>$10%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>$\hat{\varphi}_n^{(0)}$ (bootstrap)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>200</td>
<td>1.2</td>
<td>1.1</td>
<td>0.5</td>
</tr>
<tr>
<td>500</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>$\hat{\varphi}_n^{(1)}$ (bootstrap)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.6</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>100</td>
<td>1.2</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>200</td>
<td>1.2</td>
<td>1.2</td>
<td>0.7</td>
</tr>
<tr>
<td>500</td>
<td>1.0</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>$\hat{\varphi}_n^{(2)}$ (bootstrap)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.7</td>
<td>1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>200</td>
<td>1.2</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>500</td>
<td>1.2</td>
<td>1.0</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>$\hat{\varphi}_n^{(0)}$ (asymptotic)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.0</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>100</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>200</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>500</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>$\hat{\varphi}_n^{(1)}$ (asymptotic)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>200</td>
<td>1.3</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td>500</td>
<td>0.8</td>
<td>1.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>$\hat{\varphi}_n^{(2)}$ (asymptotic)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.2</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>100</td>
<td>0.3</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>200</td>
<td>0.7</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>500</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Table 1. Percentage of times $H_0$ was rejected ($k_2 = 0$ and $\kappa = 0$)
<table>
<thead>
<tr>
<th>Model</th>
<th>$H_0$</th>
<th>$H_1: a$</th>
<th>$H_1: c$</th>
<th>$H_1: d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_2$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

$$\hat{\phi}_{n,0.02}$$

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>3.3</td>
<td>4.6</td>
<td>11.5</td>
<td>7.1</td>
</tr>
<tr>
<td>100</td>
<td>4.0</td>
<td>5.0</td>
<td>26.6</td>
<td>12.4</td>
</tr>
<tr>
<td>200</td>
<td>4.5</td>
<td>3.6</td>
<td>49.1</td>
<td>22.4</td>
</tr>
</tbody>
</table>

$$\hat{\phi}_n$$

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>4.5</td>
<td>4.6</td>
<td>12.7</td>
<td>4.8</td>
</tr>
<tr>
<td>100</td>
<td>4.6</td>
<td>5.4</td>
<td>26.8</td>
<td>7.8</td>
</tr>
<tr>
<td>200</td>
<td>4.4</td>
<td>4.7</td>
<td>48.1</td>
<td>11.7</td>
</tr>
</tbody>
</table>

$$\hat{T}_n$$

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>4.7</td>
<td>6.6</td>
<td>15.8</td>
<td>7.6</td>
</tr>
<tr>
<td>100</td>
<td>3.8</td>
<td>6.2</td>
<td>32.1</td>
<td>12.1</td>
</tr>
<tr>
<td>200</td>
<td>4.9</td>
<td>5.1</td>
<td>57.7</td>
<td>28.3</td>
</tr>
</tbody>
</table>

Table 2. Percentage of times $H_0$ was rejected, 5% of significance ($k_1 = 1$, $\lambda = 0.25$ and $\kappa = 1$)
<table>
<thead>
<tr>
<th>Model</th>
<th>$H_0$</th>
<th>$H_1 : a$</th>
<th>$H_1 : b$</th>
<th>$H_1 : c$</th>
<th>$H_1 : d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\phi}_{n0.02}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\phi}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{T}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

Table 3. Percentage of times $H_0$ was rejected, 5% of significance ($k_2 = 1$, $\lambda = 0.25$ and $\kappa = 1$)

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.25</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\phi}_{n0.02}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\phi}_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>200</td>
</tr>
</tbody>
</table>

Table 4. Percentage of times $H_0$ was rejected, 5% of significance ($k_1 = 1$, $k_2 = 1$ and $\kappa = 1$)
Table 5. Percentage of times $H_0$ was rejected, 5% of significance ($k_1 = 1, k_2 = 1, \rho = 15$ and $\kappa = 1$)

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.25</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\phi}_{n0.02}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>11.8</td>
<td>7.2</td>
</tr>
<tr>
<td>100</td>
<td>25.2</td>
<td>11.8</td>
</tr>
<tr>
<td>200</td>
<td>56.3</td>
<td>24.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\phi}_n$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>7.7</td>
<td>5.3</td>
</tr>
<tr>
<td>100</td>
<td>13.7</td>
<td>6.0</td>
</tr>
<tr>
<td>200</td>
<td>24.8</td>
<td>6.3</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\phi}_{n0.02}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>3.6</td>
<td>3.3</td>
<td>2.8</td>
<td>4.2</td>
</tr>
<tr>
<td>100</td>
<td>3.7</td>
<td>5.4</td>
<td>3.0</td>
<td>4.8</td>
</tr>
<tr>
<td>200</td>
<td>3.8</td>
<td>3.8</td>
<td>4.9</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\phi}_n$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.5</td>
<td>6.1</td>
<td>7.0</td>
<td>4.6</td>
</tr>
<tr>
<td>100</td>
<td>5.3</td>
<td>7.1</td>
<td>4.8</td>
<td>6.4</td>
</tr>
<tr>
<td>200</td>
<td>4.5</td>
<td>5.9</td>
<td>5.1</td>
<td>4.9</td>
</tr>
</tbody>
</table>
Table 7. Percentage of times $H_0$ was rejected, 5% of significance ($k_1 = 3, k_2 = 0, \lambda = 0.5, \rho = 15$ and $\kappa = 1$)

<table>
<thead>
<tr>
<th>$L$</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{c}_{n0.02}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>13.8</td>
<td>7.0</td>
<td>4.6</td>
</tr>
<tr>
<td>100</td>
<td>38.8</td>
<td>17.3</td>
<td>6.9</td>
</tr>
<tr>
<td>200</td>
<td>84.1</td>
<td>47.4</td>
<td>10.4</td>
</tr>
</tbody>
</table>

| $\hat{c}_n$ |  |  |  |
| 50  | 9.4 | 7.7 | 8.7 |
| 100 | 15.5 | 8.5 | 6.3 |
| 200 | 34.1 | 14.5 | 7.5 |

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

<table>
<thead>
<tr>
<th>$\hat{c}$</th>
<th>Data-driven</th>
<th>0.02</th>
<th>0.06</th>
<th>0.07</th>
<th>0.13</th>
<th>0.15</th>
<th>0.18</th>
<th>0.20</th>
<th>0.24</th>
<th>0.26</th>
<th>0.28</th>
<th>0.35</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>5.9</td>
<td>3.95</td>
<td>3.95</td>
<td>3.95</td>
<td>3.95</td>
<td>3.95</td>
<td>5.26</td>
<td>3.95</td>
<td>3.95</td>
<td>3.95</td>
<td>3.95</td>
<td>5.26</td>
</tr>
<tr>
<td>100</td>
<td>5.2</td>
<td>1.32</td>
<td>2.63</td>
<td>1.32</td>
<td>1.32</td>
<td>2.63</td>
<td>3.95</td>
<td>5.26</td>
<td>5.26</td>
<td>3.95</td>
<td>5.26</td>
<td>3.95</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Test</th>
<th>$H_1 : \text{Var}(\beta_{00}(IQ)) &gt; 0$ or $\text{Var}(\beta_{01}(IQ)) &gt; 0$</th>
<th>$H_2 : \text{Var}(\beta_{00}(IQ)) = 0$ and $\text{Var}(\beta_{01}(IQ)) &gt; 0$</th>
<th>$H_2 : \text{Var}(\beta_{00}(IQ)) &gt; 0$ and $\text{Var}(\beta_{01}(IQ)) = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{c}_{n0.003}$</td>
<td>0.012</td>
<td>0.017</td>
<td>0.08</td>
</tr>
<tr>
<td>$\hat{c}_n$</td>
<td>0.734</td>
<td>0.734</td>
<td>0.734</td>
</tr>
<tr>
<td>$\hat{T}_n$</td>
<td>0.041</td>
<td>0.009</td>
<td>0.009</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Test</th>
<th>$H_1 : \text{Var}(\beta_{00}(IQ)) &gt; 0$ or $\text{Var}(\beta_{01}(IQ)) &gt; 0$</th>
<th>$H_2 : \text{Var}(\beta_{00}(IQ)) = 0$ and $\text{Var}(\beta_{01}(IQ)) &gt; 0$</th>
<th>$H_2 : \text{Var}(\beta_{00}(IQ)) &gt; 0$ and $\text{Var}(\beta_{01}(IQ)) = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{c}_{n0.003}$</td>
<td>0.6489</td>
<td>0.405</td>
<td>0.484</td>
</tr>
<tr>
<td>$\hat{c}_n$</td>
<td>0.491</td>
<td>0.653</td>
<td>0.543</td>
</tr>
</tbody>
</table>

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