*mathematics*

MDPI

*Article*

# Changepoint in Error-Prone Relations

**Michal Pešta**

Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, 18675 Prague, Czech Republic; michal.pesta@mff.cuni.cz

**Abstract:** Linear relations, containing measurement errors in input and output data, are considered. Parameters of these so-called errors-in-variables models can change at some unknown moment. The aim is to test whether such an unknown change has occurred or not. For instance, detecting a change in trend for a randomly spaced time series is a special case of the investigated framework. The designed changepoint tests are shown to be consistent and involve neither nuisance parameters nor tuning constants, which makes the testing procedures effortlessly applicable. A changepoint estimator is also introduced and its consistency is proved. A boundary issue is avoided, meaning that the changepoint can be detected when being close to the extremities of the observation regime. As a theoretical basis for the developed methods, a weak invariance principle for the smallest singular value of the data matrix is provided, assuming weakly dependent and non-stationary errors. The results are presented in a simulation study, which demonstrates computational efficiency of the techniques. The completely data-driven tests are illustrated through problems coming from calibration and insurance; however, the methodology can be applied to other areas such as clinical measurements, dietary assessment, computational psychometrics, or environmental toxicology as manifested in the paper.

**Keywords:** changepoint; errors-in-variables; hypothesis testing; structural break; non-stationarity; dependence; weak invariance principle; singular value; calibration; insurance

check for updates

**Citation:** Pešta, M. Changepoint in Error-Prone Relations. *Mathematics* **2021**, *9*, 89. https://doi.org/10.3390/math9010089

Received: 12 December 2020
Accepted: 29 December 2020
Published: 4 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction and Main Aims

If measured input and output data are supposed to be in some linear relations, then it is of particular interest to detect whether impact of the input characteristics has changed over time on the output observables. Moreover, only *error-prone surrogates* of the unobservable input-output characteristics are in hand instead of a precise measurement. Despite the fact that the relations and, consequently, suitable underlying stochastic models are linearly defined, the possible estimates and the corresponding inference may be highly non-linear [1]. It becomes even more challenging to handle measurement errors in input and output data simultaneously, when the linear relations are subject to change at some unknown time point—*changepoint*.

There is a vast literature aimed at linear relations modeled through so-called *measurement error models* or *errors-in-variables models* (for an overview, see [2–5], or [6]), but very little has been explored in the changepoint analysis for these models yet. A change in regression has been explored thoroughly, cf. [7] or [8]. However, such a framework does not cover the case of measurement error models. Maximum likelihood approach [9,10] and Bayesian approach [11,12] to the changepoint estimation in the measurement error models were applied, both requiring parametric distributional assumptions on the errors. The changepoint in the input data only was estimated in [13]. A change in the variance parameter of the normally distributed errors within the measurement error models was investigated in [14]. All of these mentioned contributions dealt with the changepoint estimation solely. Our *main goal* is to test for a possible change in the parameters relating the input and output data, both encumbered by some errors. Consequently, if a change is detected, we aim to estimate it. By our best knowledge, we are not aware of any similar

results even for the independent and identically distributed errors. Additionally to that, our changepoint tests are supposed to be *nuisance-parameter-free*, *distributional-free*, and to allow for very general error structures.

### 1.1. Outline

The paper is organized as follows: In the next section, our data model for the changepoint in errors-in-variables is introduced and several practical motivations for such a model are given. Section 3 contains a spectral weak invariance principle for weakly dependent and non-stationary random variables. It serves as the main theoretical tool for the consequent inference. The technical assumptions are discussed as well. Two test statistics for the changepoint detection are proposed in Section 4. Consequently, their asymptotic behavior is derived under the null as well as under the alternative hypothesis. Moreover, a consistent changepoint estimator is introduced. Section 5 contains a simulation study that compares finite sample performance of the investigated tests. It numerically emphasizes the advantages of the proposed detection procedures. A practical application of the developed approach to a calibration problem is presented in Section 6.1. On the other hand, an actuarial application concerning randomly spaced time series is performed in Section 6.2. Afterwards, our conclusion follows. Proofs are given in the Appendix A.

## 2. Changepoint in Errors-In-Variables

*Errors-in-variables* (EIV) or also called *measurement error* model

$$\boldsymbol{X} = \boldsymbol{Z} + \boldsymbol{\Theta} \tag{$\mathcal{M}$}$$

and

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{$\mathcal{H}_0$}$$

is considered, where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown *regression parameters* possibly subject to change, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{Y} \in \mathbb{R}^{n\times 1}$ consist of *observable random* variables ($\boldsymbol{X}$ are covariates and $\boldsymbol{Y}$ is a response), $\boldsymbol{Z} \in \mathbb{R}^{n\times p}$ consists of *unknown constants* and has full rank, $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{\Theta} \in \mathbb{R}^{n\times p}$ are *random errors*. This setup can be extended to a *multivariate case*, where $\boldsymbol{\beta} \in \mathbb{R}^{p\times q}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times q}$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times q}$, $q \geq 1$, see Section 3.3.

The EIV model $\mathcal{M}$–$\mathcal{H}_0$ with non-random unknown constants $\boldsymbol{Z}$ is sometimes called *functional EIV* model [2,15]. On the other hand, a different approach may handle $\boldsymbol{Z}$ as random covariates, which is called *structural EIV* model [9]. As stated by [16]: 'However, functional models played an important role in the study of measurement error models and in statistics more generally.' In addition, here, we will concentrate on the functional EIV model not because of this matter-of-fact quote, but because we wish to demonstrate a distributional-free approach, where 'no, or only minimal, assumptions are made about the distribution of the $\boldsymbol{X}$s' [4], as challenged in the introduction. Nevertheless with respect to derivation of the forthcoming theory for the functional EIV model, changing some technical assumptions would allow proving suitable results for the structural case as well.

To estimate the unknown parameter $\boldsymbol{\beta}$, one usually minimizes the *Frobenius matrix norm* of the errors $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$, see [17]. This approach leads to a *total least squares* (TLS) estimate $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X} - \lambda_{min}([\boldsymbol{X}, \boldsymbol{Y}]^\top [\boldsymbol{X}, \boldsymbol{Y}])\boldsymbol{I}_p)^{-1}\boldsymbol{X}^\top \boldsymbol{Y}$, where $\lambda_{min}(\boldsymbol{M})$ is the smallest eigenvalue of an matrix $\boldsymbol{M}$ and $\boldsymbol{I}_p$ is a $(p \times p)$ identity matrix. Geometrically speaking, the Frobenius norm tries to minimize the *orthogonal distance* between the observations and the fitted hyperplane. Therefore, the TLS are usually known as *orthogonal regression*. One can generalize this method by replacing the Frobenius norm by *any unitary invariance matrix norm*, which surprisingly yields the same TLS estimate, with interesting invariance and equivariance properties [18]. The TLS estimate is shown to be strongly and weakly consistent [1,19,20] as well as to be asymptotically normal [21–23] under various conditions.

We aim to detect a possible change in the linear relation parameter $\boldsymbol{\beta}$. The interest lies in testing the *null hypothesis* $\mathcal{H}_0$ of all observations $Y_i$'s being random variables with expectations $\boldsymbol{Z}_{i,\bullet}\boldsymbol{\beta}$'s. Our goal is to test against the alternative of the first $\tau$ outcome

observations have expectations $\boldsymbol{Z}_{i,\bullet}\boldsymbol{\beta}$'s and the remaining $n-\tau$ observations come from distributions with expectations $\boldsymbol{Z}_{i,\bullet}(\boldsymbol{\beta}+\boldsymbol{\delta})$'s, where $\boldsymbol{\delta} \neq \boldsymbol{0}$. A 'row-column' notation for a matrix $\boldsymbol{M}$ is used in this manner: $\boldsymbol{M}_{i,\bullet}$ denotes the $i$th row of $\boldsymbol{M}$ and $\boldsymbol{M}_{\bullet,j}$ corresponds to the $j$th column of $\boldsymbol{M}$. Furthermore, if $i \in \mathbb{N}_0$, then $\boldsymbol{M}_i$ stays for the first $i$ rows of $\boldsymbol{M}$ and $\boldsymbol{M}_{-i}$ represents the remaining $n-i$ rows of $\boldsymbol{M}$, when the first $i$ rows are deleted. Now more precisely, our *alternative hypothesis* is

$$\boldsymbol{Y}_\tau = \boldsymbol{Z}_\tau\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\tau \quad \text{and} \quad \boldsymbol{Y}_{-\tau} = \boldsymbol{Z}_{-\tau}(\boldsymbol{\beta}+\boldsymbol{\delta}) + \boldsymbol{\varepsilon}_{-\tau}. \tag{$\mathcal{H}_A$}$$

Here, $\boldsymbol{\delta} \equiv \boldsymbol{\delta}(n) \neq \boldsymbol{0}$ is an unknown vector parameter representing the size of change and is possibly depending on $n$. The *changepoint* $\tau \equiv \tau(n) < n$ is also an unknown scalar parameter, which depends on $n$ as well, although $\boldsymbol{\beta}$ is considered to be independent of $n$. One may also think of the changepoint in errors-in-variables framework as *segmented regression with measurement errors*, cf. [10].

### 2.1. Intercept and Fixed Regressors

Please note that the EIV model $\mathcal{M}$–$\mathcal{H}_0$ has no intercept and all the covariates are encumbered by some errors. To overcome such a restriction, one can think of an extended regression model, where some explanatory variables are *subject to error* and some are measured *precisely*. i.e., $\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{\gamma} + \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{W}$ are *observable true* and $\boldsymbol{Z}$ are *unobservable true* constants, both with full rank. Regression parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ remain unknown. Then, the non-random (fixed) *intercept* can be incorporated into the regression model by setting one column of the matrix $\boldsymbol{W}$ equal to $[1,\ldots,1]^\top$. Consequently, we may *project out* exact observations using projection matrix $\boldsymbol{R} := \boldsymbol{I}_n - \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top$. Notice that $\boldsymbol{R}$ is symmetric and idempotent. Finally, one may work with $\boldsymbol{R}\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{R}\boldsymbol{\varepsilon}$ instead of $\mathcal{H}_0$.

### 2.2. Motivations

The proposed class of models—*errors-in-variables with changepoint*—is very rich and general. Our approach and results are motivated in the context of several applications taken from chemistry, biological sciences, medicine, epidemiological studies, and finance.

Application 1: Assessing agreement in clinical measurement

Direct measurement of cardiac stroke volume or blood pressure without adverse effects is difficult or even impossible. The true values remain unknown. Indirect methods are, therefore, used instead. When a new measurement technique is developed, it has to be evaluated by comparison with an established technique rather than with the true quantity [24]. Clinicians need to test whether both measurement techniques agree sufficiently. Thereafter, the old technique may be replaced by the new one.

Application 2: Nutritional epidemiology

Data from a nutritional study were analyzed in [10], where the relation between dietary folate intake (calories adjusted μg/day) on plasma homocysteine concentration (μmol/liter of blood) was investigated. There exists a suspicion that serum homocysteine is significantly elevated when ingested folate is below a certain changepoint. Moreover, the analysis used estimates of folate that were developed with a food frequency questionnaire, which is recognized to be imperfect.

Application 3: Psychometric testing

Let us think of two psychometric instruments: unspeeded 15-item vocabulary tests and highly speeded 75-item vocabulary tests, cf. [25]. The results of both tests are error-prone. Within a group of people, there is a speculation that individuals with an unspeeded test's result exceeding some unknown level should perform dramatically better in the highly speeded test.

Application 4: Environmental toxicology

A threshold limiting value in toxicology is the dose of a toxin or a substance under which there is harmless or insignificant influence on some response. In a dose-response relationship, both of them are measured with errors. In addition, the goal is to set the threshold limiting value. Such a problem was dealt in [12] using fully Bayesian approach. Moreover, a similar task regarding the $NO_2$ concentration is discussed in [16].

Application 5: Device calibration

Later on in Section 6.1, we concentrate in more details on the *calibration* task and *exemplify the proposed methodology* through analysis of data from a calibrated device and a casual device (needs to be calibrated) in order to demonstrate practical efficiency of our detection method.

Application 6: Randomly spaced time series

Randomly spaced time series can come from a situation when the observation times are driven by the series itself. For instance, cumulative counts of occurrences of a disease in a given area [26] or aggregated claim amounts for a specific line of business in *non-life insurance* as illustrated in Section 6.2.

Besides that, there are many other applications of the changepoint within the linear relations framework in, for instance, glaciology [27], empirical economics [9], dietary assessment [11], image forensics [28], or healthcare industry [29–31].

## 3. Spectral Weak Invariance Principle

A theoretical device is going to be developed in order to construct the changepoint tests. The *smallest eigenvalue* of $\boldsymbol{\Sigma}^{-1}[\boldsymbol{X},\boldsymbol{Y}]^\top[\boldsymbol{X},\boldsymbol{Y}]$—the squared smallest singular value of the matrix $[\boldsymbol{X},\boldsymbol{Y}]\boldsymbol{\Sigma}^{-1/2}$, i.e., the data matrix $[\boldsymbol{X},\boldsymbol{Y}]$ multiplied by the inverse of a matrix square root from the error variance structure (cf. subsequent Assumption $\mathcal{E}$)—plays a key role. We proceed to the assumptions that are needed for deriving forthcoming asymptotic results. Henceforth, $\xrightarrow{\mathsf{P}}$ denotes convergence in probability, $\xrightarrow{\mathsf{D}}$ convergence in distribution, $\xrightarrow[n\to\infty]{\mathsf{D}[0,1]}$ weak convergence in the Skorokhod topology $\mathsf{D}[0,1]$ of càdlàg functions on $[0,1]$, and $[x]$ denotes the integer part of the real number $x$.

### 3.1. Assumptions

Firstly, a *design assumption* on the unobservable regressors is needed.

**Assumption $\mathcal{D}$.** $\boldsymbol{\Delta}_t := \lim_{n\to\infty} n^{-1}\boldsymbol{Z}_{[nt]}^\top\boldsymbol{Z}_{[nt]}$, $\boldsymbol{\Delta}_{-t} := \lim_{n\to\infty} n^{-1}\boldsymbol{Z}_{-[nt]}^\top\boldsymbol{Z}_{-[nt]}$ *for every* $t \in (0,1)$, *and* $\boldsymbol{\Delta} := \lim_{n\to\infty} n^{-1}\boldsymbol{Z}^\top\boldsymbol{Z}$ *are positive definite.*

It basically says that the error-free design points do not concentrate to close to each other (i.e., strict positive definiteness) and, simultaneously, they do not spread-out too far (i.e., existence of limits). For example in one-dimensional case (i.e., $p = 1$), a simple equidistant design, where $Z_{i,1} = i/(n+1)$, provides $\boldsymbol{\Delta}_t = t^3/3$ and $\boldsymbol{\Delta} = 1/3$.

Prior to postulating an *errors' assumption*, we summarize the notion of *strong mixing* ($\alpha$-mixing) dependence in more detail, which will be imposed on the model's errors. Suppose that $\{\xi_n\}_{n=1}^\infty$ is a sequence of random elements on a probability space $(\Omega,\mathcal{F},\mathsf{P})$. For sub-$\sigma$-fields $\mathcal{A},\mathcal{B}\subseteq\mathcal{F}$, let $\alpha(\mathcal{A}|\mathcal{B}) := \sup_{A\in\mathcal{A},B\in\mathcal{B}}|\mathsf{P}(A\cap B)-\mathsf{P}(A)\mathsf{P}(B)|$. Intuitively, $\alpha(\cdot|\cdot)$ measures the dependence of the events in $\mathcal{B}$ on those in $\mathcal{A}$. There are many ways in which one can describe weak dependence or, in other words, *asymptotic independence* of random variables, see [32]. Considering a filtration $\mathcal{F}_m^n = \sigma\{\xi_i \in \mathcal{F}, m \le i \le n\}$, sequence $\{\xi_n\}_{n=1}^\infty$ of random variables is said to be *strong mixing* ($\alpha$-mixing) if $\alpha(\xi_\circ,n) = \sup_{k\in\mathbb{N}}\alpha(\mathcal{F}_1^k|\mathcal{F}_{k+n}^\infty) \to 0$ as $n\to\infty$. A class of $m$-dependent processes was comprehensively analyzed in [33]. They are $\alpha$-mixing, since they are finite order ARMA processes with innovations satisfying *Doeblin's condition* ([34], p. 168). Finite order processes, which do

not satisfy Doeblin's condition, can be shown to be $\alpha$-mixing ([35], pp. 312–313). General conditions under which stationary Markov processes are $\alpha$-mixing are provided in [36]. Since functions of mixing processes are themselves mixing [32], time-varying functions of any of the processes just mentioned are mixing as well. This means that the class of the $\alpha$-mixing processes is sufficiently large for the further practical applications and that is why we chose such a mixing condition.

**Assumption $\mathcal{E}$.** *$\{[\boldsymbol{\Theta}_{n,\bullet}, \varepsilon_n]\}_{n=1}^{\infty}$ is a sequence of $\alpha$-mixing absolutely continuous random vectors with zero mean and a variance matrix $\sigma^2\boldsymbol{\Sigma}$ with an unknown $\sigma^2 > 0$ and a known positive definite $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\Theta}} & \boldsymbol{\Sigma}_{\boldsymbol{\Theta},\varepsilon} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Theta},\varepsilon}^{\top} & 1 \end{bmatrix}$ such that $\alpha([\boldsymbol{\Theta}_{\circ,\bullet}, \varepsilon_{\circ}], n) = \mathcal{O}(n^{-1-\varpi})$ as $n \to \infty$ for some $\varpi > 0$, $\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty$, $\sup_{n\in\mathbb{N}} \mathsf{E}\,|\Theta_{n,j}|^{4+\omega} < \infty$, $j \in \{1, \ldots, p\}$, and $\sup_{n\in\mathbb{N}} \mathsf{E}\,|\varepsilon_n|^{4+\omega} < \infty$ for some $\omega > 0$ such that $\omega\varpi > 2$.*

Let us emphasize that the sequence of the errors *do not have to be stationary*. The assumption of an unknown $\sigma^2$ and a known $\boldsymbol{\Sigma}$ implies that we know the ratio of any pair of covariances in advance. In the simplest situation, a homoscedastic covariance structure of the within-individual errors $[\boldsymbol{\Theta}_{n,\bullet}, \varepsilon_n]$ can be assumed (i.e., $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+1}$), if prior experience or essence of the analyzed problem allow for that. On the other hand, if the covariance matrix $\boldsymbol{\Sigma}$ is unknown, it can be estimated when possessing *replicate measurements* or *validation data* as commented in [16]. There are various approaches proposed to serve this purpose. In order to mention at least some of them, we refer to [22,37,38], or [39]. On the top of that, we have to bear in mind that $\boldsymbol{\Sigma}$ cannot be completely unspecified. If $\boldsymbol{\Sigma}$ is unrestricted, no strongly consistent estimator for $\boldsymbol{\beta}$ can exist even under normally distributed errors [40].

Relaxations of Assumption $\mathcal{E}$ with respect to the known and constant $\boldsymbol{\Sigma}$ are provided and discussed in Section 4.4—*unknown* covariance matrix—and in Section 4.5—*changing* covariance structure.

Furthermore, a *variance assumption* for the misfit disturbances is stated. It can be considered to be a typical assumption for the *long-run variance* of residuals. Let us denote $\boldsymbol{\Sigma}^{-1/2} = \begin{bmatrix} \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} & \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} \\ \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^{\top} & \bar{\Sigma}_{\varepsilon} \end{bmatrix}$ a symmetric square root of $\boldsymbol{\Sigma}^{-1}$, where $\bar{\Sigma}_{\varepsilon} \in \mathbb{R}$ is a scalar.

**Assumption $\mathcal{V}$.** *There exist $\phi := \bar{\Sigma}_{\varepsilon} - \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^{\top}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \boldsymbol{\beta}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^{\top})^{-1}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} + \boldsymbol{\beta}\bar{\Sigma}_{\varepsilon}) \neq 0$ and $v := \lim_{n\to\infty} n^{-1}\,\mathsf{Var}\,\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 > 0$.*

Let us remark that $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} = \boldsymbol{0}$ for the uncorrelated error structure and, then, $\phi = \bar{\Sigma}_{\varepsilon}$.

### 3.2. Swip

Finally, the spectral weak invariance principle (SWIP) [41] for the smallest eigenvalues is provided. Let us denote $\lambda_i := \lambda_{min}(\boldsymbol{\Sigma}^{-1}[\boldsymbol{X}_i, \boldsymbol{Y}_i]^{\top}[\boldsymbol{X}_i, \boldsymbol{Y}_i])$ for $2 \leq i \leq n$, $\lambda_0 := \lambda_1 := 0$ and $\widetilde{\lambda}_i := \lambda_{min}(\boldsymbol{\Sigma}^{-1}[\boldsymbol{X}_{-i}, \boldsymbol{Y}_{-i}]^{\top}[\boldsymbol{X}_{-i}, \boldsymbol{Y}_{-i}])$ for $0 \leq i \leq n-2$, $\widetilde{\lambda}_n := \widetilde{\lambda}_{n-1} := 0$. Please note that $\lambda_n \equiv \widetilde{\lambda}_0$.

**Proposition 1** (SWIP)**.** *Let $\mathcal{M}$ and $\mathcal{H}_0$ hold. Under Assumptions $\mathcal{D}$, $\mathcal{E}$, and $\mathcal{V}$,*

$$\left\{ \frac{1}{\sqrt{n}} \left( \lambda_{[nt]} - [nt]\sigma^2 \right) \right\}_{t\in[0,1]} \xrightarrow[n\to\infty]{\mathsf{D}[0,1]} \left\{ \frac{\phi^2 v}{1 + \|\boldsymbol{\alpha}\|_2^2} \mathcal{W}(t) \right\}_{t\in[0,1]}$$

*and*

$$\left\{ \frac{1}{\sqrt{n}} \left( \widetilde{\lambda}_{[n(1-t)]} - [n(1-t)]\sigma^2 \right) \right\}_{t\in[0,1]} \xrightarrow[n\to\infty]{\mathsf{D}[0,1]} \left\{ \frac{\phi^2 v}{1 + \|\boldsymbol{\alpha}\|_2^2} \widetilde{\mathcal{W}}(t) \right\}_{t\in[0,1]},$$

*where* $\{\mathcal{W}(t)\}_{t\in[0,1]}$ *is a standard Wiener process ,* $\widetilde{\mathcal{W}}(t) = \mathcal{W}(1) - \mathcal{W}(t)$*, and* $\boldsymbol{\alpha} = (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \boldsymbol{\beta}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^{\top})^{-1}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + \boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}})$.

### 3.3. Extension to Multivariate Case

Suppose that $\boldsymbol{\beta} \in \mathbb{R}^{p\times q}$, $\boldsymbol{Y} \in \mathbb{R}^{n\times q}$, and $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times q}$, $q \geq 1$. Let the singular value decomposition (SVD) of the partial transformed data be

$$[\boldsymbol{X}_{[nt]}, \boldsymbol{Y}_{[nt]}]\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{U}(t)\boldsymbol{\Gamma}(t)\boldsymbol{V}^{\top}(t) = \sum_{i=1}^{p+q} \varsigma(t)^{(i)}\boldsymbol{u}(t)^{(i)}\boldsymbol{v}(t)^{(i)\top},$$

where $\boldsymbol{u}(t)^{(i)}$'s are the left-singular vectors, $\boldsymbol{v}(t)^{(i)}$'s are the right-singular vectors, and $\varsigma(t)^{(i)}$'s are the singular values in the non-increasing order. One may replace $\lambda_{[nt]}$ by

$$\Lambda_{[nt]} := \sum_{j=1}^{q} \left(\varsigma(t)^{(p+j)}\right)^2$$

in Proposition 1 (and analogously for $\widetilde{\lambda}_{[n(1-t)]}$). Then, the SWIP can be derived again (see the proof of Proposition 1), provided adequately extended assumptions on the errors $\{\boldsymbol{\varepsilon}_{n,1}\}_{n=1}^{\infty}, \ldots, \{\boldsymbol{\varepsilon}_{n,q}\}_{n=1}^{\infty}$ instead of the original ones $\{\varepsilon_n\}_{n=1}^{\infty}$. However, the consequent proofs would become more technical.

## 4. Nuisance-Parameter-Free Detection

Consistent estimation of $\boldsymbol{\beta}$ can be performed via the generalized TLS approach [19,42]. The optimizing problem

$$[\boldsymbol{b}, \hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\varepsilon}}] := \underset{[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]\in\mathbb{R}^{n\times(p+1)},\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \left\|[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\boldsymbol{\Sigma}^{-1/2}\right\|_F \quad \text{s.t.} \quad \boldsymbol{Y} - \boldsymbol{\varepsilon} = (\boldsymbol{X} - \boldsymbol{\Theta})\boldsymbol{\beta},$$

where $\|\cdot\|_F$ stands for the Frobenius matrix norm, has a solution consisting of the estimator

$$\boldsymbol{b} = (\boldsymbol{X}^{\top}\boldsymbol{X} - \lambda_n\boldsymbol{\Sigma}_{\boldsymbol{\Theta}})^{-1}(\boldsymbol{X}^{\top}\boldsymbol{Y} - \lambda_n\boldsymbol{\Sigma}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}) \tag{1}$$

and the fitted errors $[\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\varepsilon}}]$ such that

$$\left\|[\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\varepsilon}}]\boldsymbol{\Sigma}^{-1/2}\right\|_F^2 = \lambda_n. \tag{2}$$

We construct the changepoint test statistics based on property (2).

### 4.1. Changepoint Test Statistics

Let us think of two TLS estimates of $\boldsymbol{\beta}$: The first one based on the first $i$ data lines $[\boldsymbol{X}_i, \boldsymbol{Y}_i]$ and the second one based on the first $k$ data lines $[\boldsymbol{X}_k, \boldsymbol{Y}_k]$ such that $1 \leq i \leq k \leq n$. Under the null $\mathcal{H}_0$, these two TLS estimates should be close to each other. On the other hand, under the alternative $\mathcal{H}_A$ such that $\tau \in \{i, \ldots, k\}$, they should be somehow different. A similar conclusion can be made for the *goodness-of-fit* statistics coming from (2). It means that

$$\lambda_i - \frac{i}{k}\lambda_k$$

should be reasonably small under the null $\mathcal{H}_0$. Under the alternative $\mathcal{H}_A$ such that $\tau \in \{i, \ldots, k\}$, it should be relatively large. For the multivariate case described in previous Section 3.3, one has to replace $\lambda_k$ by $\Lambda_k = \sum_{j=1}^{q}\left(\varsigma(k/n)^{(p+j)}\right)^2$.

We rely on *self-normalized test statistics* introduced in [43], because the unknown quantity $\phi^2 v/(1+\|\boldsymbol{\alpha}\|_2^2)$ from Proposition 1 cancels out in the test statistics. Our *supremum-type self-normalized test statistic* based on the *goodness-of-fit* is defined as

$$\mathscr{S}_n := \max_{1\le k<n} \frac{\left|\lambda_k - \frac{k}{n}\lambda_n\right|}{\max_{1\le i<k}\left|\lambda_i - \frac{i}{k}\lambda_k\right| + \max_{k<i\le n}\left|\widetilde{\lambda}_i - \frac{n-i}{n-k}\widetilde{\lambda}_k\right|} \tag{3}$$

and the *integral-type self-normalized test statistic* is defined as

$$\mathscr{T}_n := \sum_{k=1}^{n-1} \frac{\left(\lambda_k - \frac{k}{n}\lambda_n\right)^2}{\sum_{i=1}^{k-1}\left(\lambda_i - \frac{i}{k}\lambda_k\right)^2 + \sum_{i=k+1}^{n}\left(\widetilde{\lambda}_i - \frac{n-i}{n-k}\widetilde{\lambda}_k\right)^2}. \tag{4}$$

Let us note that evaluations of the above defined test statistics require just several singular value decompositions, which is reasonably *quick*. Our new test statistics involve *neither nuisance parameters nor tuning constants* (i.e., nuisance-parameter-free) and will work for non-stationary and weakly dependent data. On the top of that, *no boundary issue* is present meaning that the tests can detect the change close to the beginning or to the end of the studied regime. Furthermore, the power at the boundaries can be improved by introducing weight functions, see, e.g., [44].

Under the null hypothesis and the technical assumptions from Section 3.1, the test statistics defined in (3) and (4) converge to *non-degenerate limit distributions* (their quantiles are in Section 4.2).

**Theorem 1** (Under the null)**.** *Let $\mathcal{M}$ and $\mathcal{H}_0$ hold. Under Assumptions $\mathcal{D}$, $\mathcal{E}$, and $\mathcal{V}$,*

$$\mathscr{S}_n \xrightarrow[n\to\infty]{\mathsf{D}} \sup_{t\in[0,1]} \frac{\left|\mathcal{W}(t) - t\mathcal{W}(1)\right|}{\sup_{s\in[0,t]}\left|\mathcal{W}(s) - \frac{s}{t}\mathcal{W}(t)\right| + \sup_{s\in[t,1]}\left|\widetilde{\mathcal{W}}(s) - \frac{1-s}{1-t}\widetilde{\mathcal{W}}(t)\right|} \tag{5}$$

*and*

$$\mathscr{T}_n \xrightarrow[n\to\infty]{\mathsf{D}} \int_0^1 \frac{\left\{\mathcal{W}(t) - t\mathcal{W}(1)\right\}^2}{\int_0^t\left\{\mathcal{W}(s) - \frac{s}{t}\mathcal{W}(t)\right\}^2 ds + \int_t^1\left\{\widetilde{\mathcal{W}}(s) - \frac{1-s}{1-t}\widetilde{\mathcal{W}}(t)\right\}^2 ds} dt, \tag{6}$$

*where $\{\mathcal{W}(t)\}_{t\in[0,1]}$ is a standard Wiener process and $\widetilde{\mathcal{W}}(t) = \mathcal{W}(1) - \mathcal{W}(t)$.*

The null hypothesis is rejected at significance level $\alpha$ for large values of $\mathscr{S}_n$ and $\mathscr{T}_n$. The critical values can be obtained as the $(1-\alpha)$-quantiles of the asymptotic distributions from (5) and (6). To describe limit behavior of the test statistics under the alternative, an additional *changepoint assumption* is required.

**Assumption $\mathcal{C}$.** *For some $\zeta \in (0,1)$, as $n \to \infty$,*

$$\|\boldsymbol{\delta}\|_2 \to 0 \quad \textit{and} \quad (\eta\kappa - \boldsymbol{\varphi}^\top\boldsymbol{\varphi})\sqrt{n} \to \infty, \tag{7}$$

*where*

$$\begin{aligned}
\kappa &:= (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top + \bar{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_\zeta(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + \boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}}) + (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top + \bar{\Sigma}_{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\beta}}^\top)\boldsymbol{\Delta}_{-\zeta}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + \tilde{\boldsymbol{\beta}}\bar{\Sigma}_{\boldsymbol{\varepsilon}}),\\
\boldsymbol{\varphi} &:= (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \bar{\Sigma}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_\zeta(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + \boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}}) + (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \bar{\Sigma}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\beta}}^\top)\boldsymbol{\Delta}_{-\zeta}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + \tilde{\boldsymbol{\beta}}\bar{\Sigma}_{\boldsymbol{\varepsilon}}),\\
\eta &:= \lambda_{min}((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \boldsymbol{\beta}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top) + \sigma^2\boldsymbol{I}_p) - \sigma^2, \textit{ and } \tilde{\boldsymbol{\beta}} := \boldsymbol{\beta} + \boldsymbol{\delta}.
\end{aligned}$$

This assumption may be considered to be a changepoint *detectability requirement* for local alternatives, because it manages the relationship between the size of the change, the location of the change, and the noisiness of the data in order to be able to detect the changepoint. In case of uncorrelated error structure, the previous formulae become simpler due to $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} = \mathbf{0}$. Assumption $\mathcal{C}$ is automatically fulfilled, for instance, for an arbitrary

$\delta \to 0$ and the one-dimensional equidistant design points $Z_i$'s on $(0,1)$ with homoscedastic error structure, because then $\eta\kappa - \boldsymbol{\varphi}^\top \boldsymbol{\varphi} = \beta^2\{\zeta^3 + (1-\zeta)^3\}\{1 - \zeta^3 - (1-\zeta)^3\}/9 + O(\delta)$ as $\delta \to 0$. Furthermore, let us remark that $\boldsymbol{\vartheta} := \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \boldsymbol{\beta}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^\top$ has full rank under Assumption $\mathcal{V}$.

A practical purpose should not lie in the detection of changes that are eminent or evident, but rather concealed or hidden. The tests based on $\mathscr{S}_n$ and $\mathscr{T}_n$ are shown to be *consistent*, as the test statistics converge to infinity under some *local* alternatives, provided that the size of the change does not convergence to zero too fast, cf. Assumption $\mathcal{C}$ where $\kappa$ and $\boldsymbol{\varphi}$ depend on $\delta$.

**Theorem 2** (Under local alternatives)**.** *Let $\mathcal{M}$ and $\mathcal{H}_A$ hold such that $\tau = [n\zeta]$ for some $\zeta \in (0,1)$. Under Assumptions $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$, and $\mathcal{V}$,*

$$\mathscr{S}_n \xrightarrow[n\to\infty]{\mathsf{P}} \infty \xleftarrow[n\to\infty]{\mathsf{P}} \mathscr{T}_n. \tag{8}$$

Assumption $\mathcal{C}$ can be sharpened as remarked below with the corresponding proof in the Appendix A.

**Remark 1.** *The second part of relation (7) can be replaced by*

$$\sqrt{n}\Big\{\kappa + \eta - \sqrt{(\kappa + 2\sigma^2 + \eta)^2 - 4(\kappa + \sigma^2 - \boldsymbol{\varphi}^\top(\boldsymbol{\vartheta}^\top \boldsymbol{\Delta}\boldsymbol{\vartheta} + \sigma^2 \boldsymbol{I}_p)^{-1}\boldsymbol{\varphi})(\sigma^2 + \eta)}\Big\} \to \infty \tag{9}$$

*and the assertion of Theorem 2 still holds.*

Basically, Theorem 2 discloses that in presence of the structural change in linear relations, the test statistics *explode above all bounds*. Hence, the asymptotic distributions from Theorem 1 can be used to construct the tests, although explicit forms of those distributions stated in (5) and (6) are unknown.

### 4.2. Asymptotic Critical Values

The critical values may be determined by simulations from the limit distributions $\mathscr{S}_n$ and $\mathscr{T}_n$ from Theorem 1. Theorem 2 ensures that we reject the null hypothesis for large values of the test statistics. We have simulated the asymptotic distributions (5) and (6) by *discretizing* the standard Wiener process and using the relationship of a random walk to the standard Wiener process. We considered 1000 as the number of discretization points within $[0,1]$ interval and the number of simulation runs equals to $100,000$. In Table 1, we present several critical values for the test statistics $\mathscr{S}_n$ and $\mathscr{T}_n$.

**Table 1.** Simulated asymptotic critical values for $\mathscr{S}_n$ and $\mathscr{T}_n$.

| $100(1-\alpha)\%$ | 90% | 95% | 97.5% | 99% | 99.5% |
|---|---|---|---|---|---|
| $\mathscr{S}$-based | 1.209008 | 1.393566 | 1.571462 | 1.782524 | 1.966223 |
| $\mathscr{T}$-based | 5.700222 | 7.165705 | 8.807070 | 10.597625 | 11.755233 |

### 4.3. Changepoint Estimator

If a change is *detected*, it is of interest to estimate the changepoint. It is sensible to use

$$\hat{\tau}_n := \underset{1\le k\le n-1}{\operatorname{argmax}} \frac{\left|\lambda_k - \frac{k}{n}\lambda_n\right| + \left|\widetilde{\lambda}_k - \frac{n-k}{n}\widetilde{\lambda}_0\right|}{\max_{1\le i<k}\left|\lambda_i - \frac{i}{k}\lambda_k\right| + \max_{k<i\le n}\left|\widetilde{\lambda}_i - \frac{n-i}{n-k}\widetilde{\lambda}_k\right|}$$

as a *changepoint estimator*. Our next theorem shows that under the alternative, the changepoint $\tau$ is consistently estimated by the estimator $\hat{\tau}_n$ with the corresponding optimal convergence rate.

**Corollary 1** (Consistency)**.** *Let the assumptions of Theorem 2 hold. If*

$$\psi_n := \sqrt{n} \inf_{t\in(\zeta,1)} \{\eta(t)\kappa(t) - \boldsymbol{\varphi}(t)^\top \boldsymbol{\varphi}(t)\} \xrightarrow{n\to\infty} \infty; \tag{10}$$

$$\tilde{\psi}_n := \sqrt{n} \inf_{t\in(0,\zeta)} \{\tilde{\eta}(t)\tilde{\kappa}(t) - \tilde{\boldsymbol{\varphi}}(t)^\top \tilde{\boldsymbol{\varphi}}(t)\} \xrightarrow{n\to\infty} \infty, \tag{11}$$

*where $\kappa(t)$, $\tilde{\kappa}(t)$, $\boldsymbol{\varphi}(t)$, $\tilde{\boldsymbol{\varphi}}(t)$, $\eta(t)$, and $\tilde{\eta}(t)$ are defined in the proof, then*

$$\hat{\tau}_n/n - \zeta = O_{\mathsf{P}}\big((\psi_n \wedge \tilde{\psi}_n)^{-1}\big), \quad n \to \infty. \tag{12}$$

Conditions (10) and (11) serve as a uniform intermediary between the size of the change, the location of the change, the sample size, and the heteroscedasticity of the disturbances (cf. the proof of Corollary 1 in the Appendix A) for assuring changepoint estimator's consistency. These assumptions are again automatically fulfilled for the case discussed below Assumption $\mathcal{C}$. Other concepts of the changepoint estimator's construction relying on the ratio type statistics [45] or on the different dependency structures [46] can be adopted as well.

### 4.4. Unknown Covariance Matrix

If the true covariance matrix of the errors is unknown, possessing a surrogate covariance matrix, which is not far away from the original one, can help to overcome such an issue.

**Assumption** $\mathcal{E}1$**.** *The same as Assumption $\mathcal{E}$, except that the matrix $\boldsymbol{\Sigma}$ is* unknown*. Moreover, there exists a known positive definite $\underline{\boldsymbol{\Sigma}}$ such that* $\lim_{n\to\infty} \|\underline{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\underline{\boldsymbol{\Sigma}}^{-1/2} - \boldsymbol{I}_{p+1}\|_2 = 0$.

Here, $\|\cdot\|_2$ is the induced matrix norm by the Euclidean vector norm. Please note that $\|\boldsymbol{M}\|_2 \le \|\boldsymbol{M}\|_F$ and $\|\boldsymbol{M}\|_2^2 = \lambda_{max}(\boldsymbol{M}^\top\boldsymbol{M})$, where $\lambda_{max}(\cdot)$ stands for the largest eigenvalue. For example, if $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \frac{1}{\log n} \\ \frac{1}{\log n} & 1 \end{bmatrix}$ and $\underline{\boldsymbol{\Sigma}} = \boldsymbol{I}_2$, then $\|\underline{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\underline{\boldsymbol{\Sigma}}^{-1/2} - \boldsymbol{I}_2\|_2 = 1/\log n \to 0$.

To practically achieve Assumption $\mathcal{E}1$ without any prior knowledge about the correlation structure, it is reasonable to re-scale (e.g., normalize or standardize) the original data, to choose $\underline{\boldsymbol{\Sigma}} = \boldsymbol{I}_{p+1}$, and, then, to suppose that we deal with transformed errors with common variance and covariances vanishing sufficiently fast to zero as the sample size increases.

Let us denote $\underline{\lambda}_i := \lambda_{min}(\underline{\boldsymbol{\Sigma}}^{-1}[\boldsymbol{X}_i, \boldsymbol{Y}_i]^\top[\boldsymbol{X}_i, \boldsymbol{Y}_i])$ for $2 \le i \le n$, $\underline{\lambda}_0 := \underline{\lambda}_1 := 0$ and $\widetilde{\underline{\lambda}}_i := \lambda_{min}(\underline{\boldsymbol{\Sigma}}^{-1}[\boldsymbol{X}_{-i}, \boldsymbol{Y}_{-i}]^\top[\boldsymbol{X}_{-i}, \boldsymbol{Y}_{-i}])$ for $0 \le i \le n-2$, $\widetilde{\underline{\lambda}}_n := \widetilde{\underline{\lambda}}_{n-1} := 0$. Please note that $\underline{\lambda}_n \equiv \widetilde{\underline{\lambda}}_0$.

**Lemma 1.** *If Assumption $\mathcal{E}1$ holds instead of Assumption $\mathcal{E}$, then Proposition 1, Theorem 1, Theorem 2, and Corollary 1 are valid when $\lambda_i$'s and $\widetilde{\lambda}_i$'s are replaced by $\underline{\lambda}_i$'s and $\widetilde{\underline{\lambda}}_i$'s, respectively.*

### 4.5. Simultaneously Changing Relation and Covariance Structure

Practically speaking, it is plausible that the covariance structure could disrupt alongside the abruptly changed parameter of the linear relations. Suppose that the covariance matrix of the errors is $\sigma^2\underline{\boldsymbol{\Sigma}}$ up to the changepoint $\tau$, such that $\underline{\boldsymbol{\Sigma}} \equiv \boldsymbol{\Sigma}$, and it becomes $\sigma^2\underline{\boldsymbol{\Sigma}}$ after the changepoint $\tau$, where the bottom-right element of $\underline{\boldsymbol{\Sigma}}$ does not have to be equal 1. Now, $\underline{\boldsymbol{\Sigma}}$ and $\underline{\boldsymbol{\Sigma}}$ are considered to be unknown.

Let us construct a transformed data matrix

$$[\underset{\sim}{\boldsymbol{X}}, \underset{\sim}{\boldsymbol{Y}}] := \begin{bmatrix} [\boldsymbol{X}_\tau, \boldsymbol{Y}_\tau]\underline{\boldsymbol{\Sigma}}^{-1/2} \\ [\boldsymbol{X}_{-\tau}, \boldsymbol{Y}_{-\tau}]\underline{\boldsymbol{\Sigma}}^{-1/2} \end{bmatrix}.$$

In the analogous manner, we define $[\underset{\sim}{\boldsymbol{\Theta}}, \underset{\sim}{\boldsymbol{\varepsilon}}]$. It holds that $\mathsf{Var}[\underset{\sim}{\boldsymbol{\Theta}}_{i,\bullet}, \underset{\sim}{\varepsilon}_i] = \sigma^2 \boldsymbol{I}_{p+1}$ for $i = 1, \ldots, n$. Consequently, put $\underset{\sim}{\boldsymbol{Z}} := \underset{\sim}{\boldsymbol{X}} - \underset{\sim}{\boldsymbol{\Theta}}$. The EIV model $\mathcal{M}$–$\mathcal{H}_0$ implies $\left([\boldsymbol{X}, \boldsymbol{Y}] - [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\right) \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = \mathbf{0}$. With respect to the alternative $\mathcal{H}_A$, we define $\underset{\sim}{\boldsymbol{\beta}}$ such that $\left([\underset{\sim}{\boldsymbol{X}}_\tau, \underset{\sim}{\boldsymbol{Y}}_\tau] - [\underset{\sim}{\boldsymbol{\Theta}}_\tau, \underset{\sim}{\boldsymbol{\varepsilon}}_\tau]\right) \begin{bmatrix} \underset{\sim}{\boldsymbol{\beta}} \\ -1 \end{bmatrix} = \mathbf{0}$. Thus,

$$\underset{\sim}{\boldsymbol{\beta}} = \left(\tilde{\underset{\leftarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta}} \boldsymbol{\beta} - \tilde{\underset{\leftarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}\right) / \left(\tilde{\underset{\leftarrow}{\Sigma}}_{\varepsilon} - \tilde{\underset{\leftarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}^\top \boldsymbol{\beta}\right), \tag{13}$$

where $\underset{\leftarrow}{\boldsymbol{\Sigma}}^{1/2} := \begin{bmatrix} \tilde{\underset{\leftarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta}} & \tilde{\underset{\leftarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon} \\ \tilde{\underset{\leftarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}^\top & \tilde{\underset{\leftarrow}{\Sigma}}_{\varepsilon} \end{bmatrix}$ is a symmetric square root of $\underset{\leftarrow}{\boldsymbol{\Sigma}}$. Similarly, we define $\underset{\sim}{\boldsymbol{\delta}}$ such that $\left([\underset{\sim}{\boldsymbol{X}}_{-\tau}, \underset{\sim}{\boldsymbol{Y}}_{-\tau}] - [\underset{\sim}{\boldsymbol{\Theta}}_{-\tau}, \underset{\sim}{\boldsymbol{\varepsilon}}_{-\tau}]\right) \begin{bmatrix} \underset{\sim}{\boldsymbol{\beta}} + \underset{\sim}{\boldsymbol{\delta}} \\ -1 \end{bmatrix} = \mathbf{0}$. Hence,

$$\underset{\sim}{\boldsymbol{\delta}} = \left(\tilde{\underset{\rightarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta}} (\boldsymbol{\beta} + \boldsymbol{\delta}) - \tilde{\underset{\rightarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}\right) / \left(\tilde{\underset{\rightarrow}{\Sigma}}_{\varepsilon} - \tilde{\underset{\rightarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}^\top (\boldsymbol{\beta} + \boldsymbol{\delta})\right) - \underset{\sim}{\boldsymbol{\beta}}, \tag{14}$$

where $\underset{\rightarrow}{\boldsymbol{\Sigma}}^{1/2} := \begin{bmatrix} \tilde{\underset{\rightarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta}} & \tilde{\underset{\rightarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon} \\ \tilde{\underset{\rightarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}^\top & \tilde{\underset{\rightarrow}{\Sigma}}_{\varepsilon} \end{bmatrix}$ is a symmetric square root of $\underset{\rightarrow}{\boldsymbol{\Sigma}}$. A technical requirement assuring that $\underset{\sim}{\boldsymbol{\beta}}$ and $\underset{\sim}{\boldsymbol{\delta}}$ are well-defined is needed, cf. (13) and (14).

**Assumption** $\mathcal{B}$. $\tilde{\underset{\leftarrow}{\Sigma}}_{\varepsilon} - \tilde{\underset{\leftarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}^\top \boldsymbol{\beta} \neq 0 \neq \tilde{\underset{\rightarrow}{\Sigma}}_{\varepsilon} - \tilde{\underset{\rightarrow}{\boldsymbol{\Sigma}}}_{\boldsymbol{\Theta},\varepsilon}^\top (\boldsymbol{\beta} + \boldsymbol{\delta})$.

This assumption basically implies that the changepoint EIV problem can be correctly postulated for the transformed data $[\underset{\sim}{\boldsymbol{X}}, \underset{\sim}{\boldsymbol{Y}}]$.

Furthermore, let us denote $\underset{\sim}{\lambda}_i := \lambda_{min}([\underset{\sim}{\boldsymbol{X}}_i, \underset{\sim}{\boldsymbol{Y}}_i]^\top [\underset{\sim}{\boldsymbol{X}}_i, \underset{\sim}{\boldsymbol{Y}}_i])$ for $2 \leq i \leq n$, $\underset{\sim}{\lambda}_0 := \underset{\sim}{\lambda}_1 := 0$ and $\tilde{\underset{\sim}{\lambda}}_i := \lambda_{min}([\underset{\sim}{\boldsymbol{X}}_{-i}, \underset{\sim}{\boldsymbol{Y}}_{-i}]^\top [\underset{\sim}{\boldsymbol{X}}_{-i}, \underset{\sim}{\boldsymbol{Y}}_{-i}])$ for $0 \leq i \leq n-2$, $\tilde{\underset{\sim}{\lambda}}_n := \tilde{\underset{\sim}{\lambda}}_{n-1} := 0$. Please note that $\underset{\sim}{\lambda}_n \equiv \tilde{\underset{\sim}{\lambda}}_0$ and all these $\underset{\sim}{\lambda}$'s are unknown. Clearly, if one appropriately changes Assumptions $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$, and $\mathcal{V}$ for the 'under-tilde' versions of the involved entities, the corresponding Proposition 1, Theorem 1, Theorem 2, and Corollary 1 come into force. On one hand, Assumptions $\mathcal{C}$, $\mathcal{E}$, and $\mathcal{V}$ become even simpler due to the homoscedastic covariance matrix of $[\underset{\sim}{\boldsymbol{\Theta}}_{i,\bullet}, \underset{\sim}{\varepsilon}_i]$. On the other hand, the transformed data $[\underset{\sim}{\boldsymbol{X}}, \underset{\sim}{\boldsymbol{Y}}]$ are simply not observable.

Since a local alternative for the change in the parameter $\boldsymbol{\beta}$ is considered, a local alternative for the change of the unknown covariance matrix is assumed as well. Besides that, a known covariance matrix close in some sense to the true changing covariance matrix is required.

**Assumption** $\mathcal{E}2$. *The same as Assumption $\mathcal{E}$, except the fact that $\boldsymbol{Z}$ is replaced by $\underset{\sim}{\boldsymbol{Z}}$, $\mathsf{Var}[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i] = \sigma^2 \underset{\leftarrow}{\boldsymbol{\Sigma}} \equiv \sigma^2 \boldsymbol{\Sigma}$ under the null hypothesis $\mathcal{H}_0$, and $\mathsf{Var}[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i] = \sigma^2 (\underset{\leftarrow}{\boldsymbol{\Sigma}}\, \mathbb{1}\{i \leq \tau\} + \underset{\rightarrow}{\boldsymbol{\Sigma}}\, \mathbb{1}\{i > \tau\})$ under the alternative $\mathcal{H}_A$ for $i = 1, \ldots, n$. The matrices $\underset{\leftarrow}{\boldsymbol{\Sigma}}$, $\underset{\rightarrow}{\boldsymbol{\Sigma}}$ are unknown positive definite and $\sigma^2 > 0$ is also unknown. Moreover, there exists a known positive definite $\underline{\boldsymbol{\Sigma}}$ such that $\lim_{n\to\infty} \|\underline{\boldsymbol{\Sigma}}^{-1/2} \underset{\leftarrow}{\boldsymbol{\Sigma}} \underline{\boldsymbol{\Sigma}}^{-1/2} - \boldsymbol{I}_{p+1}\|_2 = 0 = \lim_{n\to\infty} \|\underline{\boldsymbol{\Sigma}}^{-1/2} \underset{\rightarrow}{\boldsymbol{\Sigma}} \underline{\boldsymbol{\Sigma}}^{-1/2} - \boldsymbol{I}_{p+1}\|_2$.*

One wishes to work with the observable data $[\boldsymbol{X}, \boldsymbol{Y}] \underline{\boldsymbol{\Sigma}}^{-1/2}$, to calculate the smallest eigenvalues from $\underline{\boldsymbol{\Sigma}}^{-1} [\boldsymbol{X}_i, \boldsymbol{Y}_i]^\top [\boldsymbol{X}_i, \boldsymbol{Y}_i]$, and consequently to plug-in them into the original test statistics and changepoint estimator preserving the derived properties.

**Theorem 3** (Under local alternatives of relation and covariance)**.** *Let $\mathcal{M}$ and $\mathcal{H}_A$ hold such that $\tau = [n\zeta]$ for some $\zeta \in (0, 1)$. Suppose that Assumption $\mathcal{E}2$ is satisfied instead of Assumption $\mathcal{E}$. Moreover, suppose that Assumptions $\mathcal{C}$, $\mathcal{D}$, and $\mathcal{V}$ are fulfilled for $\underset{\sim}{\boldsymbol{Z}}$ and the covariance matrix $\sigma^2 \boldsymbol{I}_{p+1}$ instead of $\boldsymbol{Z}$ and $\sigma^2 \boldsymbol{\Sigma}$, respectively. If Assumption $\mathcal{B}$ holds and $\underset{\sim}{\boldsymbol{\delta}} \neq \mathbf{0}$, then Theorem 2 and Corollary 1 are valid when $\lambda_i$'s and $\tilde{\lambda}_i$'s are replaced by $\underset{\sim}{\lambda}_i$'s and $\tilde{\underset{\sim}{\lambda}}_i$'s, respectively.*

## 5. Simulation Study

We are interested in the performance of the tests based on the self-normalized test statistics $\mathscr{S}_n$ and $\mathscr{T}_n$ that are completely nuisance-parameter-free. We focused on the comparison of the *accuracy of critical values* obtained by the simulation from the limit distributions.

In Figures 1–4, one may see *size-power plots* considering the test statistics $\mathscr{S}_n$ and $\mathscr{T}_n$ under the null hypothesis and under the alternative. Figures 1 and 2 correspond to one input covariate (i.e., $p = 1$) with choices of $\beta = 1$ and $Z_{i,1} = 100i/(n+1)$. A case with four error-prone regressors (i.e., $p = 4$) is illustrated in Figures 3 and 4 for choices of $\boldsymbol{\beta} = [1,1,1,1]^\top$ and $Z_{i,\bullet} = 100 \times [i/(n+1), (i/(n+1))^{3/2}, (i/(n+1))^{4/3}, (i/(n+1))^{5/4}]$. Next, $n \in \{200, 1000\}$ and $\tau \in \{n/4, n/2\}$. The size of change is $\delta \in \{0.1, 0.5\}$ for $p = 1$ and $\boldsymbol{\delta} \in \{[0.1, 0.1, 0.1, 0.1]^\top, [0.5, 0.5, 0.5, 0.5]^\top\}$ for $p = 4$. Especially smaller values of the break should represent the situations under the local alternatives. In Figures 1 and 3, the empirical rejection frequency under the null hypothesis (actual $\alpha$-errors) is plotted against the theoretical size (theoretical $\alpha$-errors with $\alpha \in \{1\%, 5\%, 10\%\}$), illustrating the size of the tests. The ideal situation under the null hypothesis is depicted by the straight diagonal dotted line. The empirical rejection frequencies (1−errors of the second type) under the alternative (with different changepoints and values of the change) are shown in Figures 2 and 4, illustrating the power of the tests. Under the alternative, the desired situation would be a steep function with values close to 1. For more details on the size-power plots we may refer, e.g., to [47]. The standard deviation of the random disturbances was set to $\sigma \in \{0.5, 1.0\}$ and the random error terms $\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}$, and $\{\varepsilon_n\}_{n=1}^{\infty}$ were independently simulated as three time series:

- IID … independent and identically distributed random variables;
- AR(1) … autoregressive (AR) process of order one with a coefficient of autoregression equal 0.5;
- ARCH(1) … autoregressive conditional heteroscedasticity (ARCH) process with the second coefficient equal 0.5.

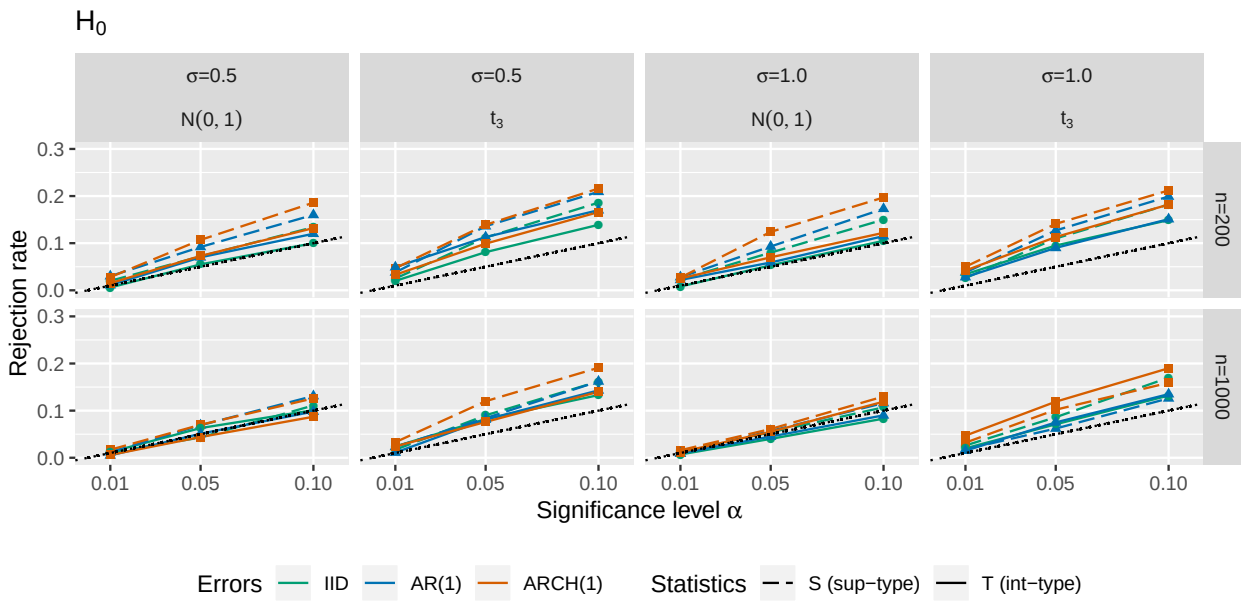


**Figure 1.** Size-power plots for $\mathscr{S}_n$ and $\mathscr{T}_n$ under $\mathcal{H}_0$ ($p = 1$).

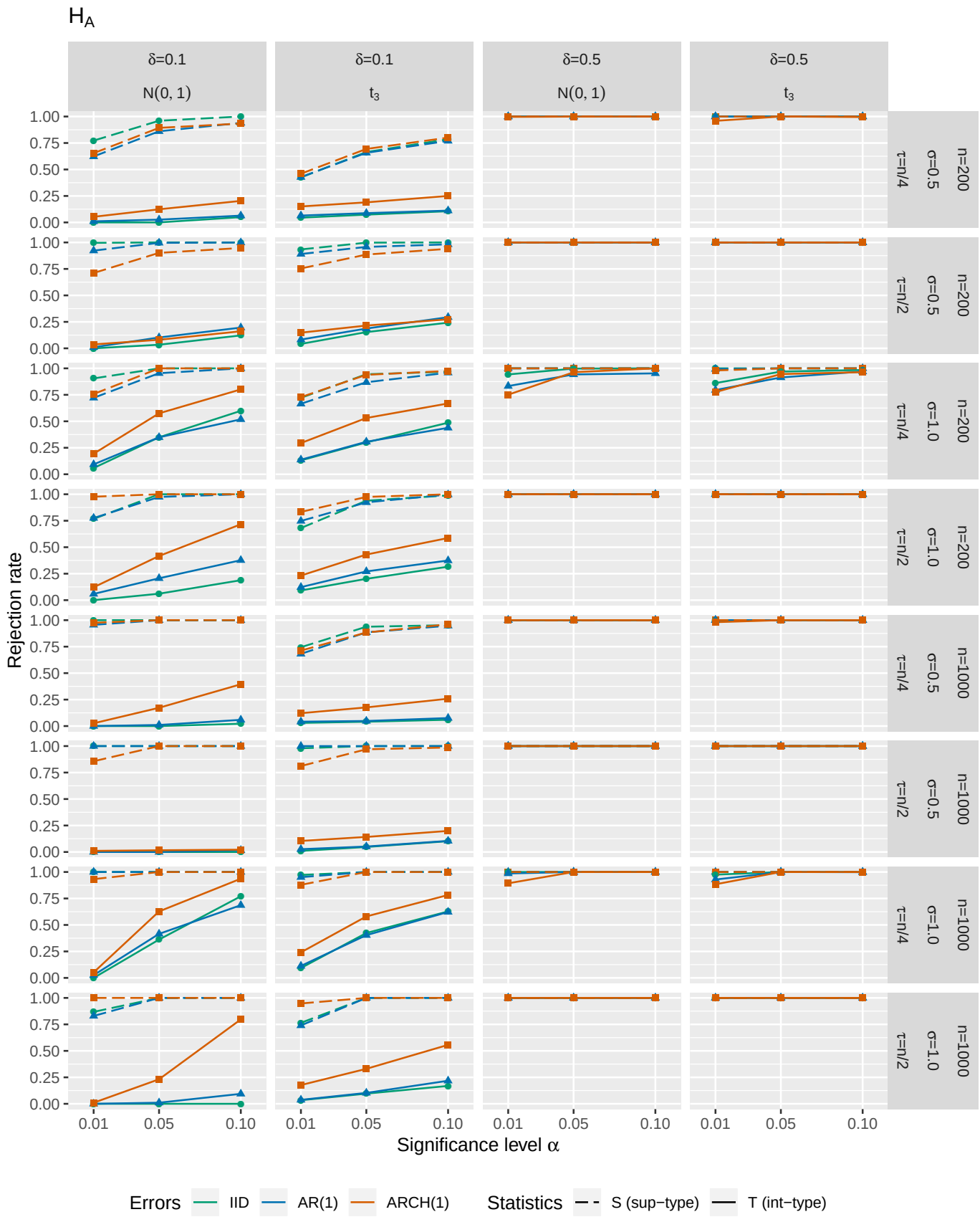


**Figure 2.** Size-power plots for $\mathscr{S}_n$ and $\mathscr{T}_n$ under $\mathcal{H}_A$ ($p = 1$).

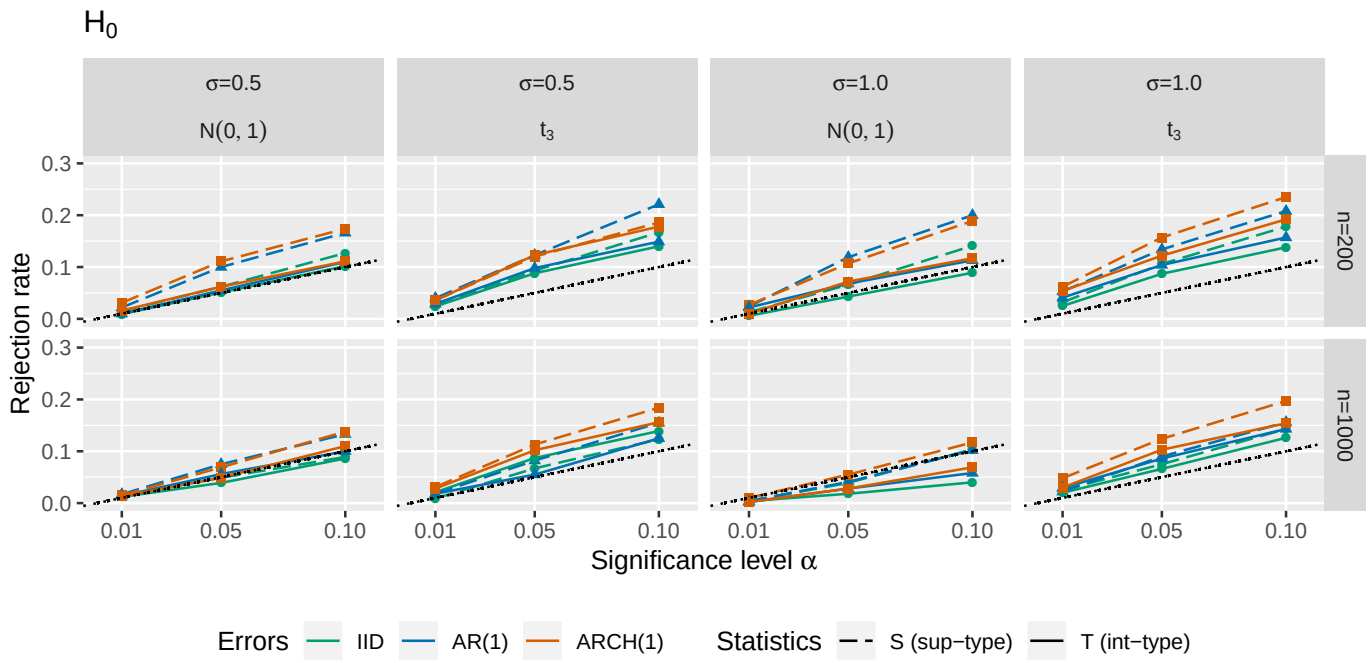


**Figure 3.** Size-power plots for $\mathscr{S}_n$ and $\mathscr{T}_n$ under $\mathcal{H}_0$ ($p = 4$).

The standard normal distribution and the Student $t$-distribution with 3 degrees of freedom are used for generating the innovations of the models' errors. All of the time series are standardized such that they have variance equal $\sigma^2$. Let us remark that the setup of Student $t_3$-distribution does not satisfy Assumption $\mathcal{E}$. However, it can be considered to be a misspecified model and one would like to inspect performance of our procedures on such a model that violates our assumptions. In the simulations of the rejection rates, we used $10,000$ repetitions.

In all of the subfigures of Figures 1 and 3 depicting a situation under the null hypothesis, we may see that comparing the accuracy of $\alpha$-levels (sizes) for different self-normalized test statistics, the integral-type ($\mathscr{T}$-based) method seems to keep the theoretical significance level more firmly than the supremum-type ($\mathscr{S}$-based) method. Comparing the case of $\mathsf{N}(0,1)$ innovations with the case of $t_3$ innovations, the rejection rates under the null tend to be slightly higher for the $t_3$-distribution. Despite the fact that the $t_3$-distributed errors violate Assumption $\mathcal{E}$, the performance of our tests is still surprisingly satisfactory in such case. As expected, the accuracy of the critical values tends to be better for larger $n$ and smaller $p$. The more complicated dependence structure of errors is assumed, the worse performance of the tests is obtained. Furthermore, the less volatile errors are set, the better tests' sizes are attained.

The $\mathscr{T}$-method performs better under the null. However under the alternative, the $\mathscr{S}$-method has a tendency to have slightly higher power than the $\mathscr{T}$-method (see Figures 2 and 4). We may also conclude that under $\mathcal{H}_A$ with less volatile errors, the power of the test increases. The power decreases when the changepoint is closer to the beginning or the end of the input-output data. The heavier tails ($t_3$ against $\mathsf{N}(0,1)$) give worse results in general for both test statistics. Moreover, 'more dependent' scenarios reveal worsening of the test statistics' performance. Furthermore, the smaller size of the change is considered, the lower power of the test is achieved. In addition, again, the power gets higher for smaller $p$ and larger $n$.

Afterwards, a simulation experiment is performed to study the *finite sample* properties of the changepoint estimator for a change in the linear relations' parameter. We numerically present only the case of $p = 1$. In particular, the interest lies in the *empirical distributions*

of the proposed estimator visualized via boxplots, see Figure 5 (bars correspond to the interquartile range). The simulation setup is kept the same as described above.

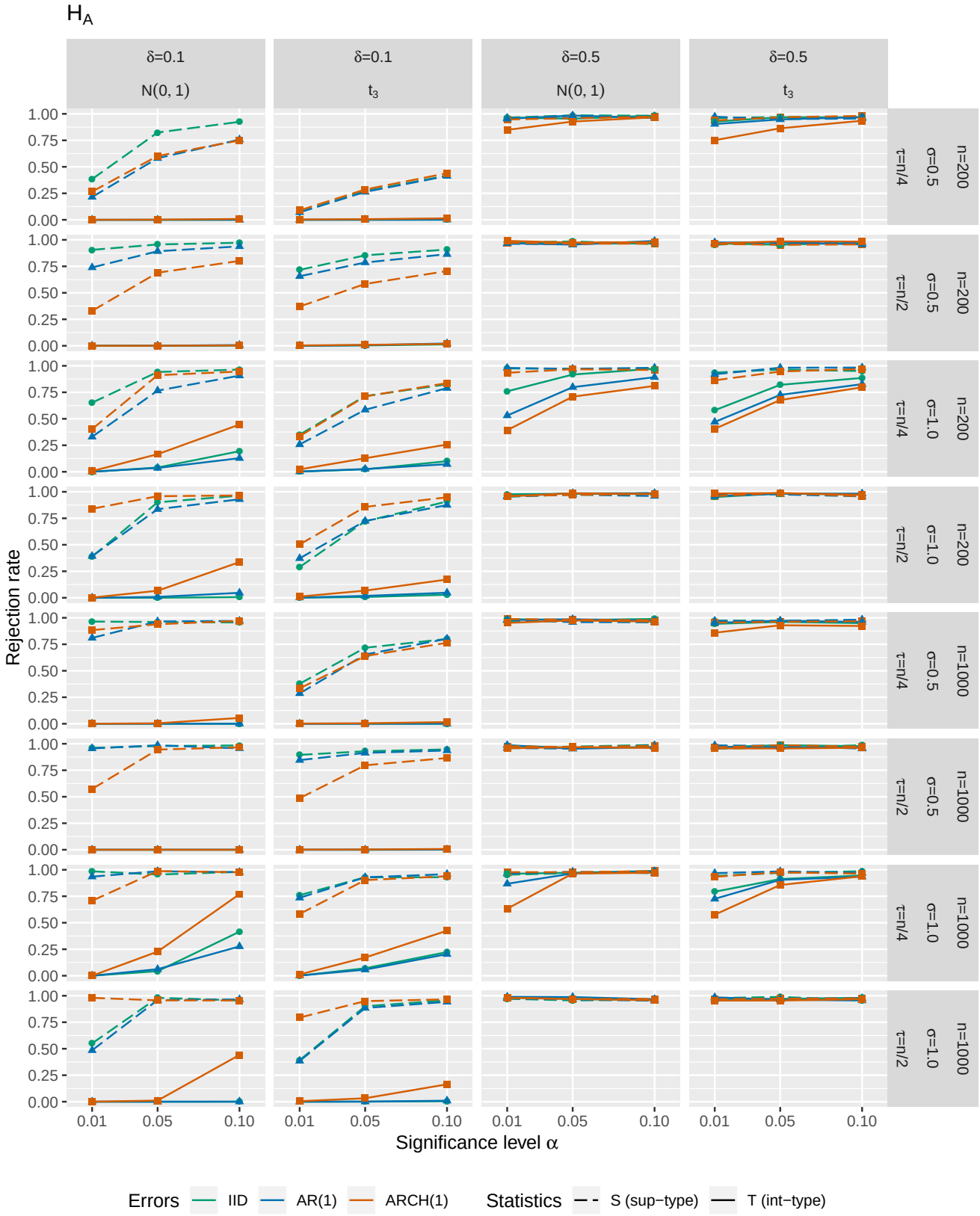


**Figure 4.** Size-power plots for $\mathscr{S}_n$ and $\mathscr{T}_n$ under $\mathcal{H}_A$ ($p = 4$).

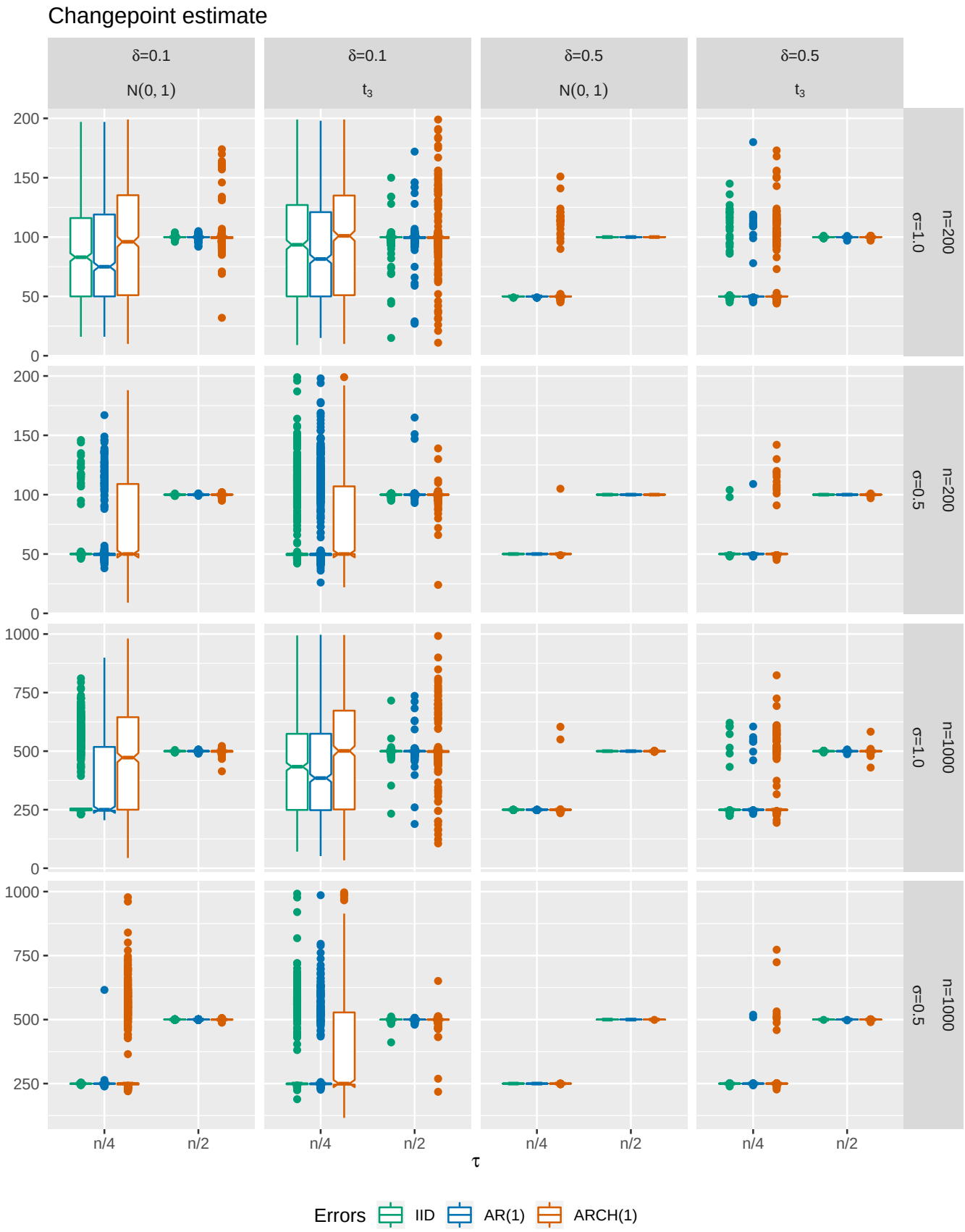


**Figure 5.** Boxplots of the estimated changepoint $\hat{\tau}_n$ ($p = 1$).

It can be concluded that the precision of our changepoint estimate is satisfactory even for relatively small sample sizes regardless of the errors' structure. Less volatile model errors provide more precise changepoint estimate. The less complicated dependence structure is assumed, the higher accuracy of the estimator is obtained. Furthermore, the disturbances with heavier tails yield less precise estimates than innovations with light tails. One may notice that higher precision is obtained when the changepoint is closer to the middle of the data. It is also clear that the precision of $\hat{\tau}_n$ improves markedly as the size of change increases.

## 6. Applications

### *6.1. Device Calibration*

A company has two industrial devices, where the first one is calibrated according to some institute of standards and the second one is just a casual device. We want to test whether the second device is calibrated according to the first one. In this *calibration problem*, it means to know whether the second device has approximately the *same performance up to some unknown multiplication constant* as the first one. Consequently, other devices of the same type are needed to be calibrated as well. For some reasons, e.g., economic or logistic, it is only possible to calibrate one device by the official authorities.

Our data set, provided by a Czech steelmaker, contains 100 couples of speed values of two hammer rams (see Figure 6), where the first forging hammer is calibrated. We set the same power level on both hammers and measure the speed of each hammer ram repeatedly changing only the power level. Our measurements of the speed are encumbered with errors of the same variability in both cases, because we use the same device for measuring the speed and both forging hammers are of the same type. Since the power set for the forging hammer is directly proportional to the speed of the hammer ram, our goal is to test whether the *ratio of two hammer rams' speeds is kept constant* over changing the power level or not. Therefore, our changepoint in the EIV model is very suitable for this setup—a linear dependence and errors in both measured speeds (with the same variance).

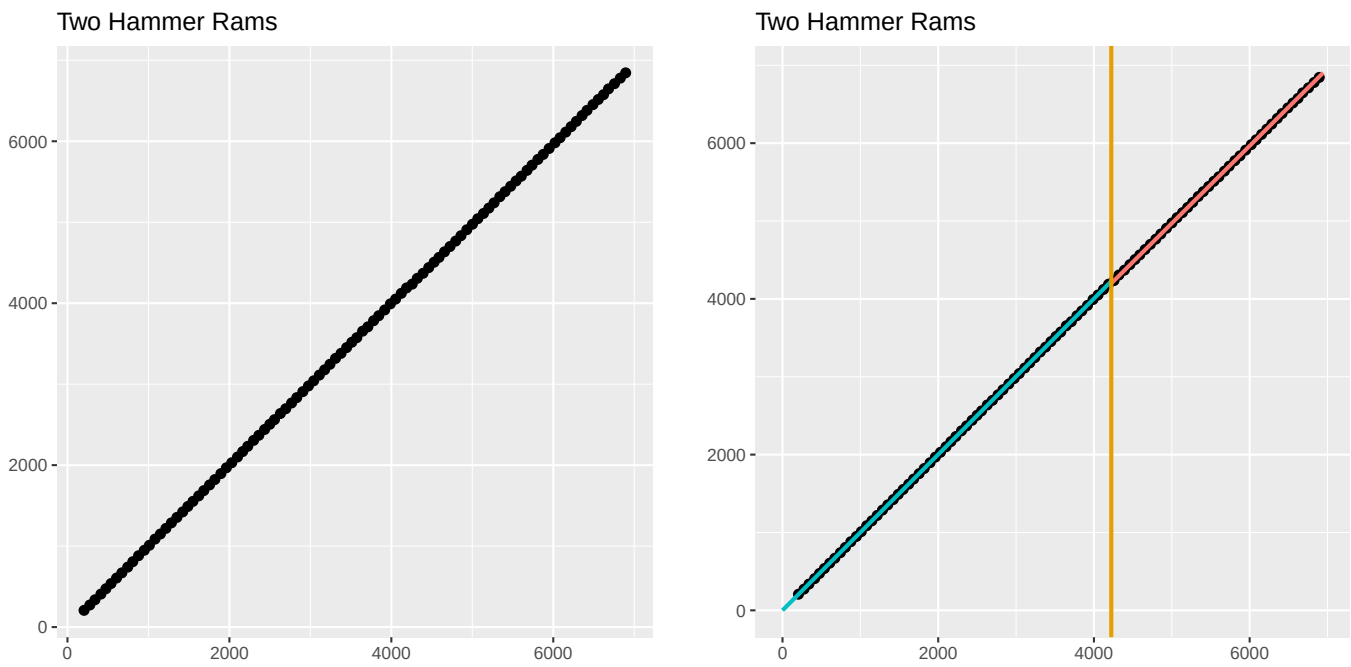


**Figure 6.** Speeds of two hammer rams, where the first one displayed on the x-axis is calibrated—without (**left**) and with (**right**) the estimated trend. The changepoint estimate corresponding to the technical issues is depicted by the vertical line.

Both our changepoint tests—$\mathscr{S}_n = 28.42$ and $\mathscr{T}_n = 127.53$—reject the null hypothesis of a constant linear coefficient between two hammer rams' speed values at the significance level of $\alpha = 5\%$ (cf. Table 1), indicating a changed performance of the second non-calibrated hammer ram, although there is no visible change in trend present—the left part of Figure 6.

As an estimate for our change, we obtain $\hat{\tau}_n = 60$ (depicted by a vertical line in Figure 6—right), which corresponds to the 60th measurement of pair of speeds. After this particular measurement, we have background information that a technical issue appeared to the second hammer ram—one of its oil tubes started to leak. Our procedure is indeed capable of detecting and, consequently, estimating the (hardly visible, but present) changepoint in the ratio of the hammer rams' speeds. In addition, this is done fully automatically without expert knowledge about the oil tube issue and also without setting tuning parameters. Moreover, the estimated ratio via the TLS approach before the change is 1.000891 (the slope of the green line in Figure 6), which basically says that the hammer rams work approximately in the same way. However, the estimated ratio via the TLS approach after the change is 0.9941487 (the slope of the red line in Figure 6), which is significantly different from constant 1 (see a formal statistical test from [22]). In contrast, a significant change in trend is not detected using the ratio type changepoint test for the regression parameter [48]. A possible reason could be that such a detection approach misleadingly considers the covariates as error-free (i.e., measured precisely).

Other calibration examples, where our methodology is applicable, can be found in, e.g., [37] or [38].

### 6.2. Randomly Spaced Time Series in Insurance

An insurance company records developments of the paid claim amounts coming from car accidents over time. Two types of claims with respect to car insurance are considered: *material damage* and *bodily injury*. The interest lies in *disorder recognition* between aggregated claim amounts paid for these two types of claims comprising two lines of insurance business.

Two claims (material damage and bodily injury) can come of one car accident and the corresponding claim amounts are expected to be related. We possess a data set from the Guarantee Fund of the Czech Insurers' Bureau for car insurance. The aggregated claim amounts from the bodily injuries are monitored every time after the aggregated claim payments from the material damages exceed a threshold of CZK 2.5M ($\approx$ €100,000). This is performed by the bureau from the beginning of 2018 up to the end of 2019. The observation times ($n = 96$) of the cumulative bodily injury claim amounts are random and irregularly spaced, because these time points are driven by a series of the cumulative material damage claim amounts.

*Identification of structural breaks* in the claim amounts and their development helps us to tie in specific legal, economic, or natural changes to the time epochs in which they occurred [49–51]. Our task is to test whether there is a *common linear time trend* of the cumulative bodily injury claim amounts or this trend has abruptly changed inside the monitoring regime [52–55]. Both our changepoint tests—$\mathscr{S}_n = 4.55$ and $\mathscr{T}_n = 22.75$—reject the null hypothesis of a constant linear trend at the significance level of $\alpha = 5\%$ (cf. Table 1), indicating a changed payment development of the bodily injury claims.

As an estimate for the change, we obtain $\hat{\tau}_n = 63$ (depicted by a vertical line in Figure 7), which corresponds to 18 April 2019. The bureau for car insurance set some regulatory procedures forcing the insurance companies to change their reporting policy since the second quarter of 2019. This can result in a change of the bodily injury claim payments' development and our detection procedures automatically reveal such *regime switching*. Additionally to that, Easter 2019 could also contribute to the change in the claim amounts. For example, Easter customs in the Central Europe and occasional drivers after the winter period cause more car accidents in general.

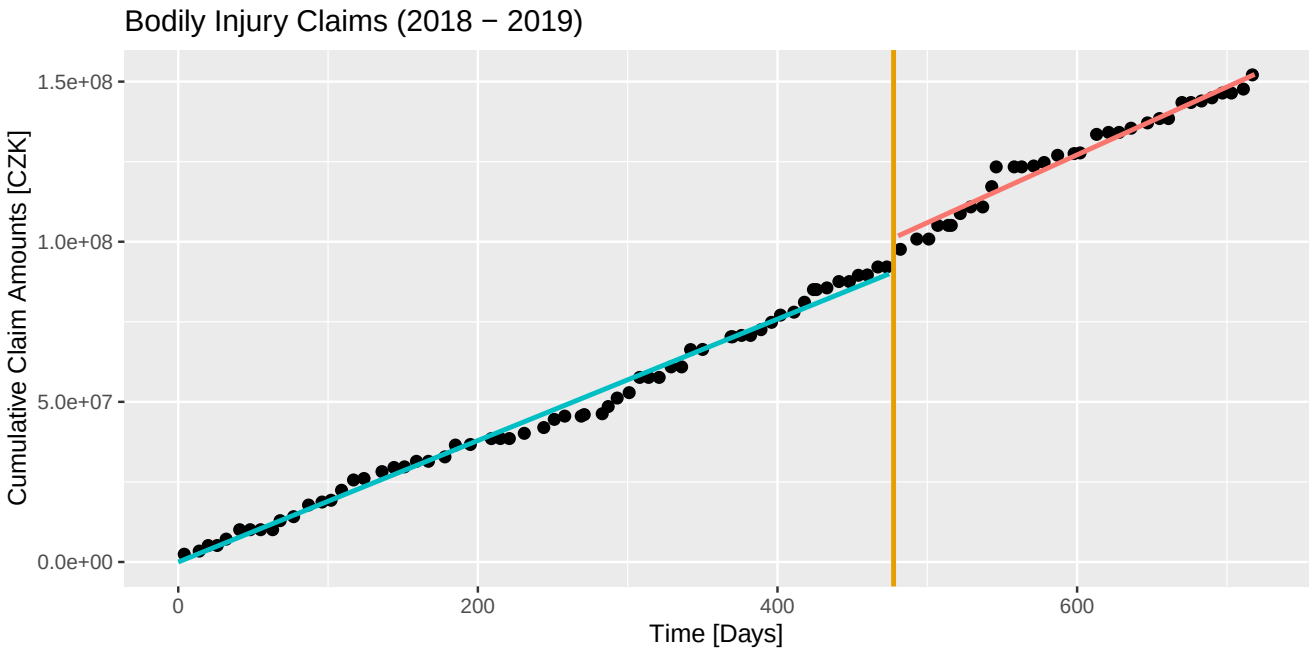


**Figure 7.** Cumulative claim amounts (y-axis) at unequally spaced non-deterministic time points (x-axis). The detected changepoint of the time trend is depicted by the vertical line.

The data were firstly re-scaled in order to reach a common empirical variance yielding satisfiable Assumption $\mathcal{D}$ and, then, inversely transformed back. The estimated linear time trend via the TLS approach before the change is 189,599.2 (the slope of the green line in Figure 7) and after the change becomes 211,885.2 (the slope of the red line in Figure 7). In contrast, a significant change in trend is not detected using the ratio type changepoint test for the regression parameter [48]. Again, a possible explanation could be that such a traditional detection approach misleadingly considers the time points as deterministic and error-free.

## 7. Conclusions

Our changepoint problem in linear relations is linearly defined, but comes with a highly non-linear solution and inference. We have proposed two tests for changepoints with desirable theoretical properties: The asymptotic size of the tests is guaranteed by a limit theorem even under non-stationarity and weak dependency; the tests and the related changepoint estimator are consistent. We are not aware of any similar results even for independent and identically distributed errors. By combining self-normalization and the proposed spectral weak invariance principle, there are neither tuning constants nor nuisance parameters involved in the whole testing procedure. Unknown or changing covariance structure of the errors is allowed. Therefore, the detection methods are completely data-driven, which makes this framework effortlessly applicable as demonstrated. In our simulations, the tests show reliable performance. Practical implementations of the developed detection techniques are demonstrated on two problems from calibration and insurance.

**Funding:** This research was supported by the Czech Science Foundation project GAČR No. 18-01781Y.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets presented in this study are available on request from the author. The data are not publicly available due to commercial issues.

**Acknowledgments:** The author would like to thank two anonymous referees for careful reading of the paper and for providing suggestions that improved this paper.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A. Proofs

**Proof of Proposition 1.** Let the singular value decomposition of the transformed 'partial' data matrix be

$$[\boldsymbol{X}_{[nt]}, \boldsymbol{Y}_{[nt]}]\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{U}(t)\boldsymbol{\Gamma}(t)\boldsymbol{V}(t)^\top = \sum_{i=1}^{p+1} \varsigma(t)^{(i)}\boldsymbol{u}(t)^{(i)}\boldsymbol{v}(t)^{(i)\top}$$

for some $t \in (0,1]$. Please note that we are in a situation of no change in the parameter $\boldsymbol{\beta}$. Bearing in mind Assumptions $\mathcal{D}$ and $\mathcal{E}$, Lemma 2.1 in [1] and Theorem 3.1 in [20] provide that $0 \neq v_{p+1}(t)^{(p+1)}$ (i.e., the last element of the last right-singular vector $\boldsymbol{v}(t)^{(p+1)}$ corresponding to the smallest singular value) with probability tending to one as $n$ increases. According to proof of Lemma 4.2 in [1], one gets

$$\frac{1}{\sqrt{n}}\left(\lambda_{[nt]} - [nt]\sigma^2\right) = \left(v_{p+1}(t)^{(p+1)}\right)^2 [\boldsymbol{a}_t^\top, -1]\left\{\frac{1}{\sqrt{n}}(\boldsymbol{D}_t - \mathsf{E}\,\boldsymbol{D}_t)\right\}\begin{bmatrix}\boldsymbol{a}_t\\ -1\end{bmatrix} \tag{A1}$$

$$+ \left(v_{p+1}(t)^{(p+1)}\right)^2 \sqrt{n}[\boldsymbol{a}_t^\top, -1]\boldsymbol{\Sigma}^{-1/2}\begin{bmatrix}\boldsymbol{I}_p\\ \boldsymbol{\beta}^\top\end{bmatrix}\frac{1}{n}\boldsymbol{Z}_{[nt]}^\top\boldsymbol{Z}_{[nt]}[\boldsymbol{I}_p, \boldsymbol{\beta}]\boldsymbol{\Sigma}^{-1/2}\begin{bmatrix}\boldsymbol{a}_t\\ -1\end{bmatrix}, \tag{A2}$$

where $\boldsymbol{a}_t := (\tilde{\boldsymbol{X}}_{[nt]}^\top\tilde{\boldsymbol{X}}_{[nt]} - \lambda_{[nt]}\boldsymbol{I}_p)^{-1}\tilde{\boldsymbol{X}}_{[nt]}^\top\tilde{\boldsymbol{Y}}_{[nt]}$ is the TLS estimator for the transformed data $[\tilde{\boldsymbol{X}}_{[nt]}, \tilde{\boldsymbol{Y}}_{[nt]}] := [\boldsymbol{X}_{[nt]}, \boldsymbol{Y}_{[nt]}]\boldsymbol{\Sigma}^{-1/2}$ and, additionally, $\boldsymbol{D}_t := [\tilde{\boldsymbol{X}}_{[nt]}, \tilde{\boldsymbol{Y}}_{[nt]}]^\top[\tilde{\boldsymbol{X}}_{[nt]}, \tilde{\boldsymbol{Y}}_{[nt]}]$. With respect to [20], we have

$$\left(v_{p+1}(t)^{(p+1)}\right)^2 = 1 - \left\|[v_1(t)^{(p+1)}, \ldots, v_p(t)^{(p+1)}]^\top\right\|_2^2 \to \frac{1}{1 + \|\boldsymbol{\alpha}\|_2^2}$$

almost surely as $n \to \infty$. Moreover, $\sqrt{n}(\boldsymbol{a}_t - \boldsymbol{\alpha}) = O_{\mathsf{P}}(1)$ as $n \to \infty$ by [56]. The strong law of large numbers for $\alpha$-mixing by [57] together with Theorem 3.1 by [20] lead to $\boldsymbol{a}_t - \boldsymbol{\alpha} = o(1)$ almost surely. Since Assumption $\mathcal{D}$ holds, the expression in (A2) is $o_{\mathsf{P}}(1)$. Furthermore, the expression on the right hand side of (A1) is $o(1)$ away from

$$\frac{1}{1 + \|\boldsymbol{\alpha}\|_2^2}[\boldsymbol{\alpha}^\top, -1]\left\{\frac{1}{\sqrt{n}}(\boldsymbol{D}_t - \mathsf{E}\,\boldsymbol{D}_t)\right\}\begin{bmatrix}\boldsymbol{\alpha}\\ -1\end{bmatrix} \tag{A3}$$

as $n \to \infty$. Hence, the process from the left hand side of (A1) in $\mathsf{D}[0,1]$ has approximately the same distribution as the process (A3).

Please note that

$$[\boldsymbol{\alpha}^\top, -1]\boldsymbol{D}_t\begin{bmatrix}\boldsymbol{\alpha}\\ -1\end{bmatrix} = (\bar{\Sigma}_\varepsilon - \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^\top\boldsymbol{\alpha})^2\|\boldsymbol{Y}_{[nt]} - \boldsymbol{X}_{[nt]}\boldsymbol{\beta}\|_2^2.$$

Using the functional central limit theorem for $\alpha$-mixing by [58] or Corollary 3.2.1 in [59] in an analogous fashion as in the proof of Theorem 2.3 by [56], one gets

$$\left\{[\boldsymbol{\alpha}^\top, -1]\left\{\frac{1}{\sqrt{n}}(\boldsymbol{D}_t - \mathsf{E}\,\boldsymbol{D}_t)\right\}\begin{bmatrix}\boldsymbol{\alpha}\\ -1\end{bmatrix}\right\}_{t\in[0,1]} \xrightarrow[n\to\infty]{\mathsf{D}[0,1]} \{\phi^2 v\mathcal{W}(t)\}_{t\in[0,1]}$$

due to Assumption $\mathcal{V}$.

Similarly for $\left\{\frac{1}{\sqrt{n}}\left(\tilde{\lambda}_{[n(1-t)]} - [n(1-t)]\sigma^2\right)\right\}_{t\in[0,1]}$ and $\{\widetilde{\mathcal{W}}(t)\}_{t\in[0,1]}$. □

**Proof of Theorem 1.** The SWIP (Proposition 1) and Lemma 1 from [60] in combination with the continuous mapping device complete the proof. □

**Proof of Theorem 2.** Under $\mathcal{H}_A$, let us find a lower bound for the smallest eigenvalue of the positive semi-definite matrix

$$\frac{1}{n}\boldsymbol{\Sigma}^{-1/2}[\boldsymbol{X},\boldsymbol{Y}]^\top[\boldsymbol{X},\boldsymbol{Y}]\boldsymbol{\Sigma}^{-1/2} = \frac{1}{n}\begin{bmatrix}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}} & \tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{Y}}\\ \tilde{\boldsymbol{Y}}^\top\tilde{\boldsymbol{X}} & \tilde{\boldsymbol{Y}}^\top\tilde{\boldsymbol{Y}}\end{bmatrix} =: \begin{bmatrix}\boldsymbol{A} & \boldsymbol{c}\\ \boldsymbol{c}^\top & d\end{bmatrix}, \tag{A4}$$

where $[\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}] := [\boldsymbol{X},\boldsymbol{Y}]\boldsymbol{\Sigma}^{-1/2}$. With respect to Theorem 1 in [61], we get

$$\lambda_{min}\left(\begin{bmatrix}\boldsymbol{A} & \boldsymbol{c}\\ \boldsymbol{c}^\top & d\end{bmatrix}\right) \geq \frac{d+\ell}{2} - \sqrt{\frac{(d-\ell)^2}{4} + \boldsymbol{c}^\top\boldsymbol{c}}, \tag{A5}$$

where $\ell$ is any lower bound on the smallest eigenvalue of the matrix $\boldsymbol{A}$. Recall that Assumption $\mathcal{E}$ and the proof of Theorem 3.1 by [20] provide

$$\frac{1}{n}\tilde{\boldsymbol{\varepsilon}}^\top\tilde{\boldsymbol{\varepsilon}} \to \sigma^2,\ \frac{1}{n}\tilde{\boldsymbol{\Theta}}^\top\tilde{\boldsymbol{\varepsilon}} \to \boldsymbol{0},\ \frac{1}{n}\tilde{\boldsymbol{\Theta}}^\top\tilde{\boldsymbol{\Theta}} \to \sigma^2\boldsymbol{I}_p,\ \frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{\varepsilon}} \to 0,\ \frac{1}{n}\tilde{\boldsymbol{Z}}^\top\tilde{\boldsymbol{\Theta}} \to \boldsymbol{0} \tag{A6}$$

almost surely as $n \to \infty$, where $[\tilde{\boldsymbol{\Theta}},\tilde{\boldsymbol{\varepsilon}}] := [\boldsymbol{\Theta},\boldsymbol{\varepsilon}]\boldsymbol{\Sigma}^{-1/2}$ and

$$\tilde{\boldsymbol{Z}} := \boldsymbol{Z}[\boldsymbol{I}_p,\boldsymbol{\beta}]\begin{bmatrix}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}\\ \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top\end{bmatrix}.$$

By Assumptions $\mathcal{C}$ and $\mathcal{D}$, one can obtain

$$\begin{aligned}\lambda(\boldsymbol{A})_{min} &= \lambda_{min}\left(\frac{1}{n}(\tilde{\boldsymbol{Z}}+\tilde{\boldsymbol{\Theta}})^\top(\tilde{\boldsymbol{Z}}+\tilde{\boldsymbol{\Theta}})\right)\\ &\to \lambda_{min}\left([\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}},\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}]\begin{bmatrix}\boldsymbol{I}_p\\ \boldsymbol{\beta}^\top\end{bmatrix}\boldsymbol{\Delta}[\boldsymbol{I}_p,\boldsymbol{\beta}]\begin{bmatrix}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}\\ \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top\end{bmatrix} + \sigma^2\boldsymbol{I}_p\right) = \sigma^2+\eta\end{aligned} \tag{A7}$$

almost surely as $n \to \infty$. Relation (A7) immediately provides a limit of a candidate for $\ell$. Now, (A4) and (A5) lead to

$$\begin{aligned}\liminf_{n\to\infty}\lambda_{min}\left(\frac{1}{n}\boldsymbol{\Sigma}^{-1/2}[\boldsymbol{X},\boldsymbol{Y}]^\top[\boldsymbol{X},\boldsymbol{Y}]\boldsymbol{\Sigma}^{-1/2}\right) &\geq \frac{\lim_{n\to\infty}\frac{1}{n}\tilde{\boldsymbol{Y}}^\top\tilde{\boldsymbol{Y}}+\sigma^2+\eta}{2}\\ &\quad - \sqrt{\frac{\left(\lim_{n\to\infty}\frac{1}{n}\tilde{\boldsymbol{Y}}^\top\tilde{\boldsymbol{Y}}-\sigma^2-\eta\right)^2}{4} + \lim_{n\to\infty}\left\|\frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{Y}}\right\|_2^2}.\end{aligned} \tag{A8}$$

Assumptions $\mathcal{C}$, $\mathcal{D}$, and relations (A6) yield

$$\begin{aligned}\frac{1}{n}\tilde{\boldsymbol{Y}}^\top\tilde{\boldsymbol{Y}} &= \frac{1}{n}\tilde{\boldsymbol{Y}}_\tau^\top\tilde{\boldsymbol{Y}}_\tau + \frac{1}{n}\tilde{\boldsymbol{Y}}_{-\tau}^\top\tilde{\boldsymbol{Y}}_{-\tau} = (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top + \bar{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_\zeta(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + \boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}})\\ &\quad + \sigma^2 + (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top + \bar{\Sigma}_{\boldsymbol{\varepsilon}}(\boldsymbol{\beta}+\boldsymbol{\delta})^\top)\boldsymbol{\Delta}_{-\zeta}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + (\boldsymbol{\beta}+\boldsymbol{\delta})\bar{\Sigma}_{\boldsymbol{\varepsilon}}) + o(1) = \kappa + \sigma^2 + o(1)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{Y}} &= \frac{1}{n}\tilde{\boldsymbol{X}}_\tau^\top\tilde{\boldsymbol{Y}}_\tau + \frac{1}{n}\tilde{\boldsymbol{X}}_{-\tau}^\top\tilde{\boldsymbol{Y}}_{-\tau} = (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_\zeta(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + \boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}})\\ &\quad + (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}(\boldsymbol{\beta}+\boldsymbol{\delta})^\top)\boldsymbol{\Delta}_{-\zeta}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}} + (\boldsymbol{\beta}+\boldsymbol{\delta})\bar{\Sigma}_{\boldsymbol{\varepsilon}}) + o(1) = \boldsymbol{\varphi} + o(1)\end{aligned}$$

almost surely as $n \to \infty$. Thus,

$$\frac{\frac{1}{n}\tilde{\boldsymbol{Y}}^\top\tilde{\boldsymbol{Y}}+\sigma^2+\eta}{2}-\sqrt{\frac{\left(\frac{1}{n}\tilde{\boldsymbol{Y}}^\top\tilde{\boldsymbol{Y}}-\sigma^2-\eta\right)^2}{4}+\left\|\frac{1}{n}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{Y}}\right\|_2^2}$$
$$=\frac{\kappa+\eta-\sqrt{(\kappa-\eta)^2+4\boldsymbol{\varphi}_2^\top\boldsymbol{\varphi}_2}}{2}+\sigma^2+o(1) \quad \text{(A9)}$$

almost surely as $n\to\infty$. Hence, combining (A8) and (A9) ends up with

$$\liminf_{n\to\infty}\lambda_{min}\left(\frac{1}{n}[\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}]^\top[\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}]\right)-\sigma^2\geq\lim_{n\to\infty}\frac{2\{\eta\kappa-\boldsymbol{\varphi}^\top\boldsymbol{\varphi}\}}{\kappa+\eta+\sqrt{(\kappa-\eta)^2+4\boldsymbol{\varphi}^\top\boldsymbol{\varphi}}}.$$

Then,

$$\frac{1}{\sqrt{n}}|\lambda_n-n\sigma^2|\xrightarrow[n\to\infty]{\text{a.s.}}\infty \quad \text{(A10)}$$

by Assumption $\mathcal{C}$.

With respect to Assumptions $\mathcal{D}$, $\mathcal{E}$, $\mathcal{V}$ and according to the underlying proof of Theorem 1, $\frac{1}{\sqrt{n}}\max_{1\leq i<\tau}\left|\lambda_i-\frac{i}{\tau}\lambda_\tau\right|$ and $\frac{1}{\sqrt{n}}\max_{\tau<i\leq n}\left|\widetilde{\lambda}_i-\frac{n-i}{n-\tau}\widetilde{\lambda}_\tau\right|$ are $O_{\mathsf{P}}(1)$ as $n\to\infty$. Moreover, $\frac{1}{\sqrt{n}}\left|\lambda_\tau-\tau\sigma^2\right|=O_{\mathsf{P}}(1)$ as $n\to\infty$ due to Proposition 1.

Please note that there are no changes in the linear parameter corresponding to the first $\tau$ observations as well as to the last (remaining) $n-\tau$ observations. Let $k=\tau$. Thus, under $\mathcal{H}_A$,

$$\begin{aligned}\mathscr{S}_n&\geq\frac{\left|\lambda_\tau-\frac{\tau}{n}\lambda_n\right|}{\max_{1\leq i<\tau}\left|\lambda_i-\frac{i}{\tau}\lambda_\tau\right|+\max_{\tau<i\leq n}\left|\widetilde{\lambda}_i-\frac{n-i}{n-\tau}\widetilde{\lambda}_\tau\right|}\\&\geq\frac{\frac{1}{\sqrt{n}}\left|\left|\lambda_\tau-\tau\sigma^2\right|-\frac{\tau}{n}\left|n\sigma^2-\lambda_n\right|\right|}{\frac{1}{\sqrt{n}}\max_{1\leq i<\tau}\left|\lambda_i-\frac{i}{\tau}\lambda_\tau\right|+\frac{1}{\sqrt{n}}\max_{\tau<i\leq n}\left|\widetilde{\lambda}_i-\frac{n-i}{n-\tau}\widetilde{\lambda}_\tau\right|}\xrightarrow[n\to\infty]{\mathsf{P}}\infty,\end{aligned}$$

because of (A10).

Furthermore, again under $\mathcal{H}_A$,

$$\begin{aligned}\mathscr{T}_n&\geq\frac{\left(\lambda_\tau-\frac{\tau}{n}\lambda_n\right)^2}{\sum_{i=1}^{\tau-1}\left(\lambda_i-\frac{i}{\tau}\lambda_\tau\right)^2+\sum_{i=\tau+1}^{n}\left(\widetilde{\lambda}_i-\frac{n-i}{n-\tau}\widetilde{\lambda}_\tau\right)^2}\\&\geq\frac{\frac{1}{n}\left(\left|\lambda_\tau-\tau\sigma^2\right|-\frac{\tau}{n}\left|n\sigma^2-\lambda_n\right|\right)^2}{\frac{1}{n}\sum_{i=1}^{\tau-1}\left(\lambda_i-\frac{i}{\tau}\lambda_\tau\right)^2+\frac{1}{n}\sum_{i=\tau+1}^{n}\left(\widetilde{\lambda}_i-\frac{n-i}{n-\tau}\widetilde{\lambda}_\tau\right)^2}\xrightarrow[n\to\infty]{\mathsf{P}}\infty,\end{aligned}$$

because of similar arguments as in the case of $\mathscr{S}_n$. □

**Proof of Remark 1.** It is sufficient to replace Theorem 1 by [61] with Theorem 3.1 by [62] in the proof of Theorem 2. □

**Proof of Corollary 1.** First of all, let us define $\tilde{\boldsymbol{\beta}}:=\boldsymbol{\beta}+\boldsymbol{\delta}$,

$$\begin{aligned}\kappa(t)&:=(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top+\bar{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_\zeta(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}})+(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top+\bar{\Sigma}_{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\beta}}^\top)(\boldsymbol{\Delta}_t-\boldsymbol{\Delta}_\zeta)(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\tilde{\boldsymbol{\beta}}\bar{\Sigma}_{\boldsymbol{\varepsilon}}),\\\tilde{\kappa}(t)&:=(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top+\bar{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_{-\zeta}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}})+(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top+\bar{\Sigma}_{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\beta}}^\top)(\boldsymbol{\Delta}_{-t}-\boldsymbol{\Delta}_{-\zeta})(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\tilde{\boldsymbol{\beta}}\bar{\Sigma}_{\boldsymbol{\varepsilon}}),\\\boldsymbol{\varphi}(t)&:=(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\bar{\Sigma}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_\zeta(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}})+(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\bar{\Sigma}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\beta}}^\top)(\boldsymbol{\Delta}_t-\boldsymbol{\Delta}_\zeta)(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\tilde{\boldsymbol{\beta}}\bar{\Sigma}_{\boldsymbol{\varepsilon}}),\\\tilde{\boldsymbol{\varphi}}(t)&:=(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\bar{\Sigma}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_{-\zeta}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\boldsymbol{\beta}\bar{\Sigma}_{\boldsymbol{\varepsilon}})+(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\bar{\Sigma}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\beta}}^\top)(\boldsymbol{\Delta}_{-t}-\boldsymbol{\Delta}_{-\zeta})(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}+\tilde{\boldsymbol{\beta}}\bar{\Sigma}_{\boldsymbol{\varepsilon}}),\\\eta(t)&:=\lambda_{min}((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_t(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\boldsymbol{\beta}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top)+t\sigma^2\boldsymbol{I}_p)-t\sigma^2,\\\tilde{\eta}(t)&:=\lambda_{min}((\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_{-t}(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}}+\boldsymbol{\beta}\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\boldsymbol{\varepsilon}}^\top)+(1-t)\sigma^2\boldsymbol{I}_p)-(1-t)\sigma^2.\end{aligned}$$

The estimator can be rewritten as

$$\hat{\tau}_n = \underset{1\le k\le n-1}{\operatorname{argmax}} \frac{\frac{1}{n}\left|\lambda_k - \frac{k}{n}\lambda_n\right| + \frac{1}{n}\left|\widetilde{\lambda}_k - \frac{n-k}{n}\widetilde{\lambda}_0\right|}{\max_{1\le i<k}\frac{1}{\sqrt{n}}\left|\lambda_i - \frac{i}{k}\lambda_k\right| + \max_{k<i\le n}\frac{1}{\sqrt{n}}\left|\widetilde{\lambda}_i - \frac{n-i}{n-k}\widetilde{\lambda}_k\right|}. \tag{A11}$$

We will treat the numerator $N_n(k)$ and the denominator $D_n(k)$ of the above stated ratio separately. Let us use notations from the previous proofs and let us recall Assumption $\mathcal{C}$, $\mathcal{D}$, and relations (A6). If $[nt] \le \tau$, then

$$\frac{1}{n}\tilde{\boldsymbol{Y}}_{[nt]}^\top \tilde{\boldsymbol{Y}}_{[nt]} = (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^\top + \bar{\Sigma}_\varepsilon \boldsymbol{\beta}^\top)\boldsymbol{\Delta}_t(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} + \boldsymbol{\beta}\bar{\Sigma}_\varepsilon) + t\sigma^2 + o(1)$$

almost surely as $n \to \infty$. Otherwise, if $[nt] > \tau$, then

$$\begin{aligned}\frac{1}{n}\tilde{\boldsymbol{Y}}_{[nt]}^\top \tilde{\boldsymbol{Y}}_{[nt]} &= (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^\top + \bar{\Sigma}_\varepsilon \boldsymbol{\beta}^\top)\boldsymbol{\Delta}_\zeta(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} + \boldsymbol{\beta}\bar{\Sigma}_\varepsilon) + t\sigma^2 \\ &\quad + (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^\top + \bar{\Sigma}_\varepsilon(\boldsymbol{\beta}+\boldsymbol{\delta})^\top)(\boldsymbol{\Delta}_t - \boldsymbol{\Delta}_\zeta)(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} + (\boldsymbol{\beta}+\boldsymbol{\delta})\bar{\Sigma}_\varepsilon) + o(1)\end{aligned}$$

almost surely as $n \to \infty$. In both cases, we have

$$\begin{aligned}\frac{1}{n}[\tilde{\boldsymbol{X}}_{[nt]}, \tilde{\boldsymbol{Y}}_{[nt]}]^\top [\tilde{\boldsymbol{X}}_{[nt]}, \tilde{\boldsymbol{Y}}_{[nt]}] &\xrightarrow[n\to\infty]{\text{a.s.}} \begin{bmatrix} \boldsymbol{\vartheta}^\top \boldsymbol{\Delta}_t \boldsymbol{\vartheta} + t\sigma^2 \boldsymbol{I}_p & \boldsymbol{G} \\ \boldsymbol{G}^\top & (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon}^\top + \bar{\Sigma}_\varepsilon \boldsymbol{\beta}^\top)\boldsymbol{\Delta}_t(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} + \boldsymbol{\beta}\bar{\Sigma}_\varepsilon) + t\sigma^2 \end{bmatrix} \\ &= t\sigma^2 \boldsymbol{I}_{p+1} + \boldsymbol{\Sigma}^{-1/2}\begin{bmatrix} \boldsymbol{I}_p \\ \boldsymbol{\beta}^\top \end{bmatrix} \boldsymbol{\Delta}_t [\boldsymbol{I}_p, \boldsymbol{\beta}] \boldsymbol{\Sigma}^{-1/2},\end{aligned}$$

where $\boldsymbol{G} := (\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta}} + \bar{\Sigma}_{\boldsymbol{\Theta},\varepsilon}\boldsymbol{\beta}^\top)\boldsymbol{\Delta}_t(\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\Theta},\varepsilon} + \boldsymbol{\beta}\bar{\Sigma}_\varepsilon)$. Therefore, for the Frobenius matrix norm $\|\cdot\|_F$,

$$\begin{aligned}&\lim_{n\to\infty}\left|\lambda_{min}\left(\frac{1}{n}[\tilde{\boldsymbol{X}}_{[nt]}, \tilde{\boldsymbol{Y}}_{[nt]}]^\top [\tilde{\boldsymbol{X}}_{[nt]}, \tilde{\boldsymbol{Y}}_{[nt]}]\right) - \frac{[nt]}{n}\lambda_{min}\left(\frac{1}{n}[\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}]^\top [\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}]\right)\right| \\ &\qquad =: \lambda_{dif}(t) \le \left\|\boldsymbol{\Sigma}^{-1/2}\begin{bmatrix} \boldsymbol{I}_p \\ \boldsymbol{\beta}^\top \end{bmatrix} \boldsymbol{\Delta}_{-t} [\boldsymbol{I}_p, \boldsymbol{\beta}] \boldsymbol{\Sigma}^{-1/2}\right\|_F\end{aligned}$$

uniformly in $t$ almost surely, because $|\lambda_{min}(\boldsymbol{A}) - \lambda_{min}(\boldsymbol{B})| \le \|\boldsymbol{A} - \boldsymbol{B}\|_F$ due to proof of Lemma 2.3 in [21].

For $k = \tau$, Proposition 1 together with the continuous mapping theorem yield that the denominator from (A11)

$$D_n(\tau) \xrightarrow[n\to\infty]{\mathsf{D}} \frac{\phi^2 v}{1 + \|\boldsymbol{\alpha}\|_2^2}\left\{\sup_{0\le t\le\zeta}\left|\mathcal{W}(t) - \frac{t}{\zeta}\mathcal{W}(\zeta)\right| + \sup_{\zeta<t\le 1}\left|\widetilde{\mathcal{W}}(t) - \frac{1-t}{1-\zeta}\widetilde{\mathcal{W}}(\zeta)\right|\right\} =: W,$$

where the limit $W$ is strictly positive almost surely. We conclude that $N_n(\tau)/D_n(\tau)$ converge in distribution to the random variable $\{\lambda_{dif}(\zeta) + \widetilde{\lambda}_{dif}(\zeta)\}/W$ such that $\widetilde{\lambda}_{dif}(t) := \lim_{n\to\infty}|\widetilde{\lambda}_{[nt]} - \frac{n-[nt]}{n}\widetilde{\lambda}_0|$. For $k = [nt]$ with $t > \zeta$, we obtain

$$\begin{aligned}&\max_{1\le i<[nt]}\frac{1}{\sqrt{n}}\left|\lambda_i - \frac{i}{[nt]}\lambda_{[nt]}\right| + \max_{[nt]<i\le n}\frac{1}{\sqrt{n}}\left|\widetilde{\lambda}_i - \frac{n-i}{n-[nt]}\widetilde{\lambda}_{[nt]}\right| \\ &\ge \frac{1}{\sqrt{n}}\left|\lambda_{[n\zeta]} - \frac{[n\zeta]}{[nt]}\lambda_{[nt]}\right| \ge \frac{1}{\sqrt{n}}\left|\left|\lambda_{[n\zeta]} - [n\zeta]\sigma^2\right| - \frac{[n\zeta]}{[nt]}\left|\lambda_{[nt]} - [nt]\sigma^2\right|\right| \\ &\approx \left|O_{\mathsf{P}}(1) - \sqrt{n}\frac{\zeta}{t}\left|\frac{\lambda_{[nt]}}{n} - t\sigma^2\right|\right|\end{aligned}$$

$$\approx \left| O_{\mathsf{P}}(1) - 2\sqrt{n}\frac{\zeta}{t}\frac{\left|\eta(t)\kappa(t) - \boldsymbol{\varphi}(t)^\top \boldsymbol{\varphi}(t)\right|}{\kappa(t) + \eta(t) + \sqrt{(\kappa(t) - \eta(t))^2 + 4\boldsymbol{\varphi}(t)^\top \boldsymbol{\varphi}(t)}} \right| \xrightarrow[n\to\infty]{\mathsf{P}} \infty$$

according to the proof of Theorem 2 and assumption (10). The convergence holds uniformly for all $t$ outside any right $\epsilon$-neighborhood of $\zeta$. It follows that for an arbitrary $\epsilon > 0$,

$$\max_{k:(k-\tau)\geq n\epsilon} \frac{N_n(k)}{D_n(k)} = O_{\mathsf{P}}\left( \frac{1}{\sqrt{n}\inf_{t\in(\zeta,1)} \left|\eta(t)\kappa(t) - \boldsymbol{\varphi}(t)^\top \boldsymbol{\varphi}(t)\right|} \right). \tag{A12}$$

Similar arguments can be applied in the case $t < \zeta$. Assumption (11) provides, for an arbitrary $\epsilon > 0$,

$$\max_{k:(\tau-k)\geq n\epsilon} \frac{N_n(k)}{D_n(k)} = O_{\mathsf{P}}\left(\tilde{\psi}_n^{-1}\right). \tag{A13}$$

Relation (12) is equivalent to the statement that for every $\tilde{\epsilon} > 0$, there exists $\tilde{M} > 0$, such that $\mathsf{P}[(\psi_n \wedge \tilde{\psi}_n)|\hat{\tau}/n - \zeta| > \tilde{M}] < \tilde{\epsilon}$. Thus, it is sufficient to show for any $\tilde{\epsilon} > 0$, there exists $\tilde{M} > 0$, such that

$$\mathsf{P}\left[ \max_{k:(\psi_n\wedge\tilde{\psi}_n)|k-\tau|\geq \tilde{M}n} \frac{N_n(k)}{D_n(k)} > \frac{N_n(\tau)}{D_n(\tau)} \right] < \tilde{\epsilon}. \tag{A14}$$

By choosing $\epsilon = \tilde{M}/(\psi_n \wedge \tilde{\psi}_n)$, relations (A12) and (A13) yield

$$\max_{k:(\psi_n\wedge\tilde{\psi}_n)|k-\tau|\geq \tilde{M}n} \frac{N_n(k)}{D_n(k)} = O_{\mathsf{P}}\left( \frac{1}{\psi_n \wedge \tilde{\psi}_n} \right). \tag{A15}$$

Since $N_n(\tau)/D_n(\tau)$ is bounded in probability, (A15) implies (A14). □

**Proof of Lemma 1.** The relative Weyl theorem [63] applied on the multiplicative perturbation matrix $\boldsymbol{\Sigma}^{1/2}\underline{\boldsymbol{\Sigma}}^{-1/2}$ provides

$$|\underline{\lambda}_i - \lambda_i| \leq \lambda_i \|\underline{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\underline{\boldsymbol{\Sigma}}^{-1/2} - \boldsymbol{I}_{p+1}\|_2, \quad i = 1, \ldots, n.$$

Then, realizing $\lambda_i = O_{\mathsf{P}}(\sqrt{n})$ as $n \to \infty$ completes the proof. □

**Proof of Theorem 3.** Since $[\underset{\sim}{\boldsymbol{\Theta}}, \underset{\sim}{\varepsilon}]$ are the linearly transformed original errors $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$, the sequence $\{[\underset{\sim}{\boldsymbol{\Theta}}_{n,\bullet}, \underset{\sim}{\varepsilon}_n]\}_{n=1}^\infty$ is $\alpha$-mixing with the coefficients $\alpha([\underset{\sim}{\boldsymbol{\Theta}}_{\circ,\bullet}, \underset{\sim}{\varepsilon}_\circ], n) = \mathcal{O}(n^{-1-\omega})$ as $n \to \infty$ due to [32]. Moreover, $\sup_{n\in\mathbb{N}} \mathsf{E}\,|\underset{\sim}{\Theta}_{n,j}|^{4+\omega} < \infty$, $j \in \{1, \ldots, p\}$ and $\sup_{n\in\mathbb{N}} \mathsf{E}\,|\underset{\sim}{\varepsilon}_n|^{4+\omega} < \infty$. Then, Assumption $\mathcal{E}$ is satisfied for $\{[\underset{\sim}{\boldsymbol{\Theta}}_{n,\bullet}, \underset{\sim}{\varepsilon}_n]\}_{n=1}^\infty$ with $\boldsymbol{\Sigma} = \boldsymbol{I}_{p+1}$. Theorem 2 and Corollary 1 can be applied to $[\underset{\sim}{\boldsymbol{X}}, \underset{\sim}{\boldsymbol{Y}}]$, where the resulting assertions hold for $\underset{\sim}{\lambda}_i$'s and $\underset{\sim}{\widetilde{\lambda}}_i$'s instead of $\lambda_i$'s and $\widetilde{\lambda}_i$'s, respectively.

Let us realize that $\underset{\sim}{\lambda}_i = O_{\mathsf{P}}(\sqrt{n})$ as $n \to \infty$. Finally, it is sufficient to employ the relative Weyl theorem [63] and we get

$$|\underline{\lambda}_i - \underset{\sim}{\lambda}_i| \leq \underset{\sim}{\lambda}_i \|\underline{\boldsymbol{\Sigma}}^{-1/2}\underset{\leftarrow}{\boldsymbol{\Sigma}}\underline{\boldsymbol{\Sigma}}^{-1/2} - \boldsymbol{I}_{p+1}\|_2, \quad i = 1, \ldots, \tau;$$
$$|\underline{\lambda}_i - \underset{\sim}{\lambda}_i| \leq \underset{\sim}{\lambda}_i \|\underline{\boldsymbol{\Sigma}}^{-1/2}\underset{\rightarrow}{\boldsymbol{\Sigma}}\underline{\boldsymbol{\Sigma}}^{-1/2} - \boldsymbol{I}_{p+1}\|_2, \quad i = \tau+1, \ldots, n.$$

Therefore, one can replace $\underset{\sim}{\lambda}_i$'s and $\underset{\sim}{\widetilde{\lambda}}_i$'s by $\underline{\lambda}_i$'s and $\underline{\widetilde{\lambda}}_i$'s, respectively, in the test statistics and changepoint estimator. □

## References


1. Gleser, L.J. Estimation in a multivariate "errors in variables" regression model: Large sample results. *Ann. Stat.* **1981**, *9*, 24–44. [CrossRef]
2. Fuller, W.A. *Measurement Error Models*; Wiley: New York, NY, USA, 1987.
3. Van Huffel, S.; Vandewalle, J. *The Total Least Squares Problem: Computational Aspects and Analysis*; SIAM: Philadelphia, PA, USA, 1991.
4. Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2006.
5. Buonaccorsi, J.P. *Measurement Error: Models, Methods, and Applications*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2010.
6. Yi, G.Y. *Statistical Analysis with Measurement Error or Misclassification*; Spriger: New York, NY, USA, 2017.
7. Horváth, L. Detecting changes in linear regressions. *J. Am. Stat. Assoc.* **1995**, *26*, 189–208. [CrossRef]
8. Aue, A.; Horváth, L.; Hušková, M.; Kokoszka, P. Testing for changes in polynomial regression. *Bernoulli* **2008**, *14*, 637–660. [CrossRef]
9. Chang, Y.P.; Huang, W.T. Inferences for the linear errors-in-variables with changepoint models. *J. Am. Stat. Assoc.* **1997**, *92*, 171–178. [CrossRef]
10. Staudenmayer, J.; Spiegelman, D. Segmented regression in the presence of covariate measurement error in main study/validation study designs. *Biometrics* **2002**, *58*, 871–877.
11. Carroll, R.J.; Roeder, K.; Wasserman, L. Flexible parametric measurement error models. *Biometrics* **1999**, *55*, 44–54. [CrossRef]
12. Gössl, C.; Küchenhoff, H. Bayesian analysis of logistic regression with an unknown change point and covariate measurement error. *Statist. Med.* **2001**, *20*, 3109–3121.
13. Kukush, A.; Markovsky, I.; Van Huffel, S. Estimation in a linear multivariate measurement error model with a change point in the data. *Comput. Stat. Data Anal.* **2007**, *52*, 1167–1182.
14. Dong, C.; Tan, C.; Jin, B.; Miao, B. Inference on the change point estimator of variance in measurement error models. *Lith. Math. J.* **2016**, *56*, 474–491. [CrossRef]
15. Booth, J.G.; Hall, P. Bootstrap confidence regions for functional relationships in errors-in-variables models. *Ann. Stat.* **1993**, *21*, 1780–1791.
16. Stefanski, L.A. Measurement error models. *J. Am. Stat. Assoc.* **2000**, *95*, 1353–1358. [CrossRef]
17. Golub, G.H.; Van Loan, C.F. An analysis of the total least squares problem. *SIAM J. Numer. Anal.* **1980**, *17*, 883–893. [CrossRef]
18. Pešta, M. Unitarily invariant errors-in-variables estimation. *Stat. Pap.* **2016**, *57*, 1041–1057. [CrossRef]
19. Gallo, P.P. Consistency of regression estimates when some variables are subject to error. *Commun. Stat. A-Theor.* **1982**, *11*, 973–983. [CrossRef]
20. Pešta, M. Strongly consistent estimation in dependent errors-in-variables. *Acta Univ. Carol. Math. Phys.* **2011**, *52*, 69–79.
21. Gallo, P.P. Properties of Estimators in Errors-in-Variables Models. Ph.D. Thesis, University of North Carolina, Chapel Hill, NC, USA, 1982.
22. Pešta, M. Total least squares and bootstrapping with application in calibration. *Statistics* **2013**, *47*, 966–991. [CrossRef]
23. Pešta, M. Block bootstrap for dependent errors-in-variables. *Commun. Stat. A-Theor.* **2017**, *46*, 1871–1897. [CrossRef]
24. Bland, J.M.; Altman, D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1986**, *1*, 307–310. [CrossRef]
25. Lord, F.M. Testing if two measuring procedures measure the same dimension. *Psychol. Bull.* **1973**, *79*, 71–72. [CrossRef]
26. Wright, D.J. Forecasting data published at irregular time intervals using an extension of Holt's method. *Manag. Sci.* **1986**, *32*, 499–510. [CrossRef]
27. Gleser, L.J.; Watson, G.S. Estimation of a linear transformation. *Biometrika* **1973**, *60*, 525–534. [CrossRef]
28. Ryu, S.J.; Lee, H.K. Estimation of linear transformation by analyzing the periodicity of interpolation. *Pattern Recogn. Lett.* **2014**, *36*, 89–99. [CrossRef]
29. Ochoa, J.; León, A.; Ramírez, I.; Lopera, C.; Bernal, E.; Arbeláez, M. Prevalence of tuberculosis infection in healthcare workers of the public hospital network in Medellín, Colombia: A Bayesian approach. *Epidemiol. Infect.* **2017**, *145*, 1095–1106. [CrossRef]
30. Hosseini, M.; Jiang, Y.; Yekkehkhany, A.; Berlin, R.R.; Sha, L. A mobile geo-communication dataset for physiology-aware DASH in rural ambulance transport. In *Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17)*; Association for Computing Machinery: New York, NY, USA, 2017; pp. 158–163. [CrossRef]
31. Li, M.; Liu, Z.; Li, X.; Liu, Y. Dynamic risk assessment in healthcare based on Bayesian approach. *Reliab. Eng. Syst. Saf.* **2019**, *189*, 327–334. [CrossRef]
32. Bradley, R.C. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2005**, *2*, 107–144. [CrossRef]
33. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis*; John Wiley & Sons: New York, NY, USA, 1958.
34. Billingsley, P. *Convergence of Probability Measures*, 1st ed.; John Wiley & Sons: New York, NY, USA, 1968.
35. Ibragimov, I.A.; Linnik, Y.V. *Independent and Stationary Sequences of Random Variables*; Wolters-Noordhoff: Groningen, The Netherlands, 1971.
36. Rosenblatt, M. *Markov Processes: Structure and Asymptotic Behavior*; Springer-Verlag: Berlin, Germany, 1971.

37. Cheng, C.L.; Riu, J. On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics* **2006**, *48*, 511–519. [CrossRef]
38. Guo, Y.; Little, R.J. Regression analysis with covariates that have heteroscedastic measurement error. *Statist. Med.* **2011**, *30*, 2278–2294. [CrossRef] [PubMed]
39. Li, M.; Ma, Y.; Li, R. Semiparametric regression for measurement error model with heteroscedastic error. *J. Multivar. Anal.* **2019**, *171*, 320–338. [CrossRef]
40. Nussbaum, M. Asymptotic optimality of estimators of a linear functional relation if the ratio of the error variances is known. *Statistics* **1977**, *8*, 173–198.
41. Kao, C.; Trapani, L.; Urga, G. Testing for instability in covariance structures. *Bernoulli* **2018**, *24*, 740–771. [CrossRef]
42. Van Huffel, S.; Vandewalle, J. Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in $A$ are subject to error. *SIAM J. Matrix Anal. Appl.* **1989**, *10*, 294–315. [CrossRef]
43. Shao, X.; Zhang, X. Testing for change points in time series. *J. Am. Stat. Assoc.* **2010**, *105*, 1228–1240. [CrossRef]
44. Csörgő, M.; Horváth, L. *Limit Theorems in Change-Point Analysis*; Wiley: Chichester, UK, 1997.
45. Fang, H.; Ding, S.; Li, X.; Yang, W. Asymptotic approximations of ratio moments based on dependent sequences. *Mathematics* **2020**, *8*, 361. [CrossRef]
46. Ding, S.; Li, X.; Dong, X.; Yang, W. The consistency of the CUSUM-type estimator of the change-point and its application. *Mathematics* **2020**, *8*, 2113. [CrossRef]
47. Kirch, C. Resampling Methods for the Change Analysis of Dependent Data. Ph.D. Thesis, University of Cologne, Cologne, Germany, 2006.
48. Peštová, B.; Pešta, M. Variance estimation free tests for structural changes in regression. In *Nonparametric Statistics*; Bertail, P., Blanke, D., Cornillon, P.A., Matzner-Løber, E., Eds.; Springer: Cham, Switzerland, 2018; Volume 250, pp. 357–373.
49. Pešta, M.; Hudecová, Š. Asymptotic consistency and inconsistency of the chain ladder. *Insur. Math. Econ.* **2012**, *51*, 472–479. [CrossRef]
50. Hudecová, Š.; Pešta, M. Modeling Dependencies in Claims Reserving with GEE. *Insur. Math. Econ.* **2013**, *53*, 786–794. [CrossRef]
51. Pešta, M.; Okhrin, O. Conditional least squares and copulae in claims reserving for a single line of business. *Insur. Math. Econ.* **2014**, *56*, 28–37. [CrossRef]
52. Peštová, B.; Pešta, M. Change point estimation in panel data without boundary issue. *Risks* **2017**, *5*, 7. [CrossRef]
53. Maciak, M.; Peštová, B.; Pešta, M. Structural breaks in dependent, heteroscedastic, and extremal panel data. *Kybernetika* **2018**, *54*, 1106–1121. [CrossRef]
54. Pešta, M.; Peštová, B.; Maciak, M. Changepoint estimation for dependent and non-stationary panels. *Appl. Math-Czech* **2020**, *65*, 299–310. [CrossRef]
55. Maciak, M.; Pešta, M.; Peštová, B. Changepoint in dependent and non-stationary panels. *Stat. Pap.* **2020**, *61*, 1385–1407. [CrossRef]
56. Pešta, M. Asymptotics for weakly dependent errors-in-variables. *Kybernetika* **2013**, *49*, 692–704.
57. Chen, X.; Wu, Y. Strong law for mixing sequence. *Acta Math. Appl. Sin.* **1989**, *5*, 367–371. [CrossRef]
58. Herrndorf, N. Stationary strongly mixing sequences not satisfying the central limit theorem. *Ann. Probab.* **1983**, *11*, 809–813. [CrossRef]
59. Lin, Z.; Lu, C. *Limit Theory for Mixing Dependent Random Variables*; Springer-Verlag: New York, NY, USA, 1997.
60. Pešta, M.; Wendler, M. Nuisance-parameter-free changepoint detection in non-stationary series. *TEST* **2020**, *29*, 379–408. [CrossRef]
61. Dembo, A. Bounds on the extreme eigenvalues of positive-definite Toeplitz matrices. *IEEE Trans. Inform. Theory* **1988**, *34*, 352–355. [CrossRef]
62. Ma, E.M.; Zarowski, C.J. On lower bounds for the smallest eigenvalue of a Hermitian positive-definite matrix. *IEEE Trans. Inform. Theory* **1995**, *41*, 539–540. [CrossRef]
63. Nakatsukasa, Y. Absolute and relative Weyl theorems for generalized eigenvalue problems. *Linear Algebra Appl.* **2010**, *432*, 242–248. [CrossRef]