

Article

# Change Point Estimation in Panel Data without Boundary Issue

Barbora Peřtová <sup>1,†</sup> and Michal Peřta <sup>2,\*,†</sup>

<sup>1</sup> Department of Medical Informatics and Biostatistics, Institute of Computer Science, The Czech Academy of Sciences, Pod Vodárenskou věží 271/2, 18207 Prague 8, Czech Republic; pestova@cs.cas.cz

<sup>2</sup> Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 18675 Prague 8, Czech Republic

\* Correspondence: michal.pesta@mff.cuni.cz; Tel.: +420-608-427-147

† These authors contributed equally to this work.

Academic Editor: Qihe Tang

Received: 28 August 2016; Accepted: 17 January 2017; Published: 22 January 2017

**Abstract:** Panel data of our interest consist of a moderate number of panels, while the panels contain a small number of observations. An estimator of common breaks in panel means without a boundary issue for this kind of scenario is proposed. In particular, the novel estimator is able to detect a common break point even when the change happens immediately after the first time point or just before the last observation period. Another advantage of the elaborated change point estimator is that it results in the last observation in situations with no structural breaks. The consistency of the change point estimator in panel data is established. The results are illustrated through a simulation study. As a by-product of the developed estimation technique, a theoretical utilization for correlation structure estimation, hypothesis testing and bootstrapping in panel data is demonstrated. A practical application to non-life insurance is presented, as well.

**Keywords:** change point; estimation; consistency; panel data; short panels; boundary issue; structural change; bootstrap; non-life insurance; change in claim amounts

**MSC:** 62F10; 62F40; 62H12; 62H15; 62E20; 62P05

**JEL:** C330; C130; C120; C150; G220

## 1. Introduction and Main Aims

The problem of an unknown common change in means of the panels is studied, where the panel data consist of  $N$  panels, and each panel contains  $T$  observations over time. Various values of the change are possible for each panel at some unknown common time  $\tau = 1, \dots, T$ . The panels are considered to be independent, although this restriction can be weakened. On the other hand, within the panels, the observations are generally not assumed to be independent. This is in accordance with typical assumptions that one can make about real data. A common dependence structure is supposed to be present over the panels. Our main goal is to construct an estimator of a possible change point, which is consistent even in the case of no structural break.

### 1.1. Current State of the Art

Tests for change point detection in panel data have been proposed by [1] for sufficiently large panel sizes  $T$ , i.e., the limiting results were derived under the assumption that  $T$  increases over all limits. Testing procedures for the change in panel means with  $T$  fixed, which can be relatively small or moderate, were considered in [2]. The change point estimation in panel data for fixed, as well as

for unbounded  $T$  was studied by [3]. However, the panel change point estimator in [3] is derived only for a situation, where one knows for sure that the change in means occurred within the given time period. This restriction can become insurmountable for some further utilization of the change point estimator, as will be demonstrated later in this paper. In [4], a consistent change point estimator was introduced, requiring no definite knowledge about the existence of the change point in the given panel data. In the case of no change being present, the estimator picks the last observation, which means that no structural break is identified. However, this estimator has several disadvantages. It assumes a certain kind of homoscedasticity in the panels. Further, it does not take into account the possibility that the change may occur right after the first time point. It also assumes conditions that may be viewed as too complicated with regard to verification and model checking. The remaining task is, therefore, to develop a change point estimator that is consistent regardless of the change's presence/absence. Moreover, such an estimator would gain from allowing heteroscedasticity in the panels, having a broader scope of applications. Besides that, the applicability of the estimator is enhanced by simple consistency conditions and no boundary issue. The boundary issue means that the change point can neither be detected nor estimated when being close to the beginning or to the end of the observation regime.

Further on, Kim [5,6] dealt with the change point estimator under cross-sectional dependence in the panels modeled by a common factor and expanded the estimation problem for more complicated types of structural changes. The first and second order asymptotics that can be used to derive consistent confidence intervals for the time of change in panel data were established by [7]. The panel length  $T$  was considered as unbounded and depending on the number of panels  $N$ . However, there is some literature on the short panel change point framework where also weighting functions, as we employ later on, are suggested, cf. [8].

### 1.2. Motivation in Non-Life Insurance

Structural changes in panel data, especially common breaks in means, are wide-spread phenomena. Our primary motivation comes from the non-life insurance business, where associations in many countries uniting several insurance companies collect claim amounts paid by every insurance company each year. Such a database of cumulative claim payments can be viewed as panel data, where insurance company  $i = 1, \dots, N$  provides the total claim amount  $Y_{i,t}$  paid in year  $t = 1, \dots, T$  into the common database. The members of the association can consequently profit from the joint database.

For the whole association, it is important to know whether a possible change in the claim amounts occurred during the observed time horizon. Usually, the time period is relatively short, e.g., 10–15 years. To be more specific, a widely-used and very standard actuarial method for predicting future claim amounts, called chain ladder, assumes a kind of stability of the historical claim amounts. The formal necessary and sufficient condition is derived in [9]. This paper shows a way to detect a possible historical instability.

### 1.3. Structure of the Paper

The remainder of the paper is organized as follows. Section 2 introduces an abrupt change point model together with stochastic assumptions. An estimator for the change point in panel means is proposed in Section 3. Consequently, the consistency of the considered change point estimator is derived, which covers the first main theoretical contribution. Section 4 contains a simulation study that illustrates the finite sample performance of the estimator. It numerically emphasizes the advantages and disadvantages of the proposed approach. The second main theoretical contribution lies in the panel correlation structure estimation and in the bootstrap add-on justification for hypothesis testing, all provided in Section 5. A practical application of the developed approach to an actuarial problem is presented in Section 6. Proofs are given in the Appendix.

## 2. Abrupt Change in Panel Data

Let us consider the panel change point model

$$Y_{i,t} = \mu_i + \delta_i \mathcal{I}\{t > \tau\} + \sigma_i \varepsilon_{i,t}, \quad 1 \leq i \leq N, 1 \leq t \leq T; \quad (1)$$

where  $\sigma_i > 0$  are unknown variance-scaling panel-specific parameters and  $T$  is fixed, not depending on  $N$ . The possible common change point time is denoted by  $\tau \in \{1, \dots, T\}$ . A situation where  $\tau = T$  corresponds to no change in means of the panels. The means  $\mu_i$  are panel individual. The amount of the break in the mean, which can also differ for every panel, is denoted by  $\delta_i$ . There is at most one change per panel in Model (1), and the type of change in the panel mean is abrupt.

Furthermore, it is assumed that the sequences of panel disturbances  $\{\varepsilon_{i,t}\}_t$  are independent. At the same time, the errors within each panel form a weakly-stationary sequence with a common correlation structure. This can be formalized in the following assumption.

**Assumption 1.** The vectors  $[\varepsilon_{i,1}, \dots, \varepsilon_{i,T}]^\top$  existing on a probability space  $(\Omega, \mathcal{F}, P)$  are iid for  $i = 1, \dots, N$  with  $E\varepsilon_{i,t} = 0$  and  $\text{Var} \varepsilon_{i,t} = 1$ , having the autocorrelation function

$$\rho_t = \text{Corr}(\varepsilon_{i,s}, \varepsilon_{i,s+t}) = \text{Cov}(\varepsilon_{i,s}, \varepsilon_{i,s+t}), \quad \forall s \in \{1, \dots, T-t\},$$

which is independent of the lag  $s$ , the cumulative autocorrelation function

$$r(t) = \text{Var} \sum_{s=1}^t \varepsilon_{i,s} = \sum_{|s|<t} (t - |s|) \rho_s,$$

and the shifted cumulative correlation function

$$R(t, v) = \text{Cov} \left( \sum_{s=1}^t \varepsilon_{i,s}, \sum_{u=t+1}^v \varepsilon_{i,u} \right) = \sum_{s=1}^t \sum_{u=t+1}^v \rho_{u-s}, \quad t < v$$

for all  $i = 1, \dots, N$  and  $t, v = 1, \dots, T$ . The covariance matrix  $\Lambda := \text{Var} \left[ \sum_{s=1}^1 \varepsilon_{1,s}, \dots, \sum_{s=1}^T \varepsilon_{1,s} \right]^\top$  is non-singular.

The sequence  $\{\varepsilon_{i,t}\}_{t=1}^T$  can be viewed as a part of a weakly-stationary process. Note that the within-panel dependent errors do not necessarily need to be linear processes. GARCH processes are a plausible alternative, for instance.

The assumption of independent panels can be relaxed. It would, however, make the setup much more complex, cf. [5]. Consequently, probabilistic tools for dependent data need to be used (e.g., suitable versions of the central limit theorem). Nevertheless, assuming that the claim amounts for different insurance companies are independent is reasonable with regard to real-life experience.

**Assumption 2.** There exist constants  $\underline{\sigma}, \bar{\sigma} > 0$  not depending on  $N$ , such that

$$\underline{\sigma} \leq \sigma_i \leq \bar{\sigma}, \quad 1 \leq i \leq N.$$

The assumption of the bounded panel variances from both below and above allows for heteroscedasticity between the panels. In the case when the equiboundedness cannot be satisfied, the panel model can be generalized by introducing weights  $w_{i,t}$ , which are supposed to be known. Subsequently, claim ratios  $Y_{i,t}/w_{i,t}$  can be modeled. Being particular in actuarial practice, it would mean normalizing the total claim amount by the premium received (considered as the weight), since bigger insurance companies are expected to have higher variability in total claim amounts paid.

### 3. Change Point Estimator

A consistent estimator of the change point in panel data is proposed in [3], but under circumstances that the change occurred for sure. In our situation, we do not know whether a change has occurred or not. Therefore, we modify the estimate proposed by [3] in the following way. If the panel means change somewhere inside  $\{1, \dots, T - 1\}$ , let the estimate select this break point. If there is no change in panel means, the estimator points out the very last time point  $T$  with the probability going to one. In other words, the value of the change point estimate can be  $T$ , meaning no change. This is in contrast to [3], where  $T$  is not reachable.

Our estimator of the time of change  $\tau$  in panel data is defined as

$$\hat{\tau}_N := \arg \min_{t=1, \dots, T} \sum_{i=1}^N \left\{ \frac{1}{w(t)} \sum_{s=1}^t (Y_{i,s} - \bar{Y}_{i,t})^2 + \frac{1}{w(T-t)} \sum_{s=t+1}^T (Y_{i,s} - \tilde{Y}_{i,t})^2 \right\}, \tag{2}$$

where  $\bar{Y}_{i,t}$  is the average of the first  $t$  observations in panel  $i$  and  $\tilde{Y}_{i,t}$  is the average of the last  $T - t$  observations in panel  $i$ , i.e.,

$$\bar{Y}_{i,t} = \frac{1}{t} \sum_{s=1}^t Y_{i,s} \quad \text{and} \quad \tilde{Y}_{i,t} = \frac{1}{T-t} \sum_{s=t+1}^T Y_{i,s}.$$

By convention, the value of an empty sum is zero. A sequence of positive weights  $\{w(t)\}_{t=0}^T$  is specified later on.

#### 3.1. Consistency

We postulate additional assumptions on the panel change point model (1) in order to derive the estimator’s consistency. The following conditions take into account that the length  $T$  of the observation regime is fixed; that the length  $T$  does not depend on the number of panels  $N$ ; and that the length  $T$  can even be relatively small.

**Assumption 3.** Let  $g(t) := \frac{t}{w(t)} \left(1 - \frac{r(t)}{t^2}\right)$  for  $t \in \{1, \dots, T\}$ ,  $g(0) \equiv 0$ , and

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left\{ \frac{\tau}{\tau+1} \sum_{i=1}^N \delta_i^2 - (g(\tau) + g(T-\tau)) \max_{t=1, \dots, T} w(t) \sum_{i=1}^N \sigma_i^2 \right\} = \infty,$$

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left\{ \frac{T-\tau}{T-\tau+1} \sum_{i=1}^N \delta_i^2 - (g(\tau) + g(T-\tau)) \max_{t=1, \dots, T} w(t) \sum_{i=1}^N \sigma_i^2 \right\} = \infty.$$

**Assumption 4.**  $\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1}^N \delta_i^2 = 0$ .

**Assumption 5.**  $E\varepsilon_{1,t}^4 < \infty$ ,  $t \in \{1, \dots, T\}$ .

**Theorem 1** (Change point estimator consistency). Under Assumptions 1–5,

$$\lim_{N \rightarrow \infty} P[\hat{\tau}_N = \tau] = 1.$$

The formally-postulated estimator’s consistency in Theorem 1 can be practically interpreted: as one observes more and more panels, the probability that the proposed estimator is different from the true unknown change point gets smaller and smaller.

Assumption 3 is not restrictive at all, although it may be seen as a complicated one. For example in the case of independent observations within the panel (i.e.,  $r(t) = t$ ) and the weight function  $w(t) = t^q$ ,  $q \geq 2$  for  $t \in \{1, \dots, T\}$ ,  $w(0) = 1$ , the sequence  $\{g(t)\}_{t=2}^T$  becomes  $\{t^{1-q} - t^{-q}\}_{t=2}^T$  and is

non-increasing. Then, Assumption 3 is automatically fulfilled, if  $q = 2$  and  $\frac{1}{\sqrt{N}} \sum_{i=1}^N (\delta_i^2 - T^2 \sigma_i^2) \rightarrow \infty$  as  $N \rightarrow \infty$ . This also gives us an idea how to choose the weights  $w(t)$ . Condition Assumptions 1 and 3, which we impose on the model errors, only pertain to the correlation structure. Hence, our results hold for nearly all stationary time series models of interest, including nonlinear time series, like the ARCH and GARCH processes. Moreover, Assumption 3 controls the trade-off between the size of breaks and the variability of errors. It may be considered as a detectability assumption, because it specifies the value of the signal-to-noise ratio for finding the consistent estimator.

Assumptions 3 and 4 are satisfied, for instance, if  $0 < \delta \leq \delta_i \leq \Delta$  for all  $i$ 's (a common lower and upper threshold for the means' shifts),  $\delta^2 = \mathcal{O}(N^\zeta)$ ,  $\zeta > 0$  and  $\Delta^2/N \rightarrow 0$  as  $N \rightarrow \infty$  (bearing in mind Assumption 2). Another suitable example of  $\delta_i$ 's for the conditions in Assumptions 3 and 4 can be  $\delta_i = Ki^\eta$  for some  $K > 0$  and  $0 < \eta < 1/2$ . Condition Assumptions 3 and 4 do not require each panel to have a break. Sometimes, a more restrictive assumption can be assumed instead of Assumptions 3 and 4, e.g.,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_i^2 = \infty. \quad (3)$$

On the one hand, this assumption might be considered as too strong, because a common fixed (not depending on  $N$ ) value of  $\delta = \delta_i$  for all  $i$ 's does not fulfill (3). On the other hand, (3) is satisfied when  $\delta_j^2/N \rightarrow \infty$  as  $N \rightarrow \infty$  for some  $j \in \mathbb{N}$  and  $\delta_i = 0$  for all  $i \neq j$ . This stands for a situation when all of the panels do not change in mean, except one panel having a sufficiently large change in mean with respect to the number of panels. Let us notice that one could replace Assumption 3 with a stronger assumption from (3), but it would mean the detectability relation disappearing between the size of breaks and the variability of errors. One would also lose an idea of how to choose the weights. Furthermore, Assumptions E1 and E2 from [4] are more restrictive than Assumption 3, which makes the presented approach even more general.

Various competing consistent estimators of a possible change point can be suggested, e.g., the maximizer of  $\sum_{i=1}^N [\sum_{s=1}^t (Y_{i,s} - t\bar{Y}_{i,T})]^2$ , as in [7]. To show the consistency of this estimator, one needs to postulate different assumptions on the cumulative autocorrelation function, and this may be rather complex.

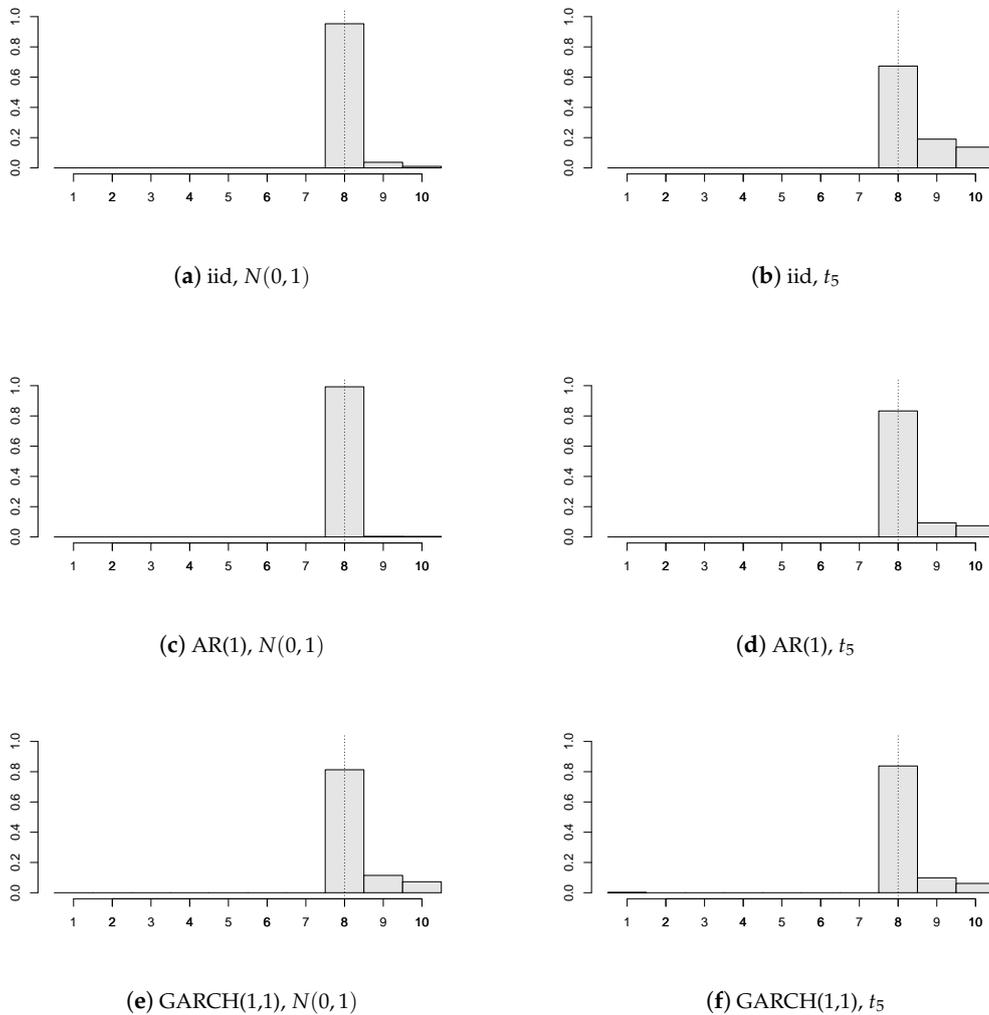
In our opinion, it is erroneously assumed in [3] that only the second moment of the errors is sufficient to prove the consistency result. In particular, Lemma A.1 from [3] has to require the finite fourth errors' moments, which coincides with Assumption 5.

#### 4. Simulation Study

A simulation experiment was performed to study the finite sample properties of the change point estimator for a common abrupt change in panel means. In particular, the interest lies in the empirical distributions of the proposed estimator visualized via histograms. Random samples of panel data (2000 each time) are generated from the panel change point model (1). The panel size is set to  $T = 10$  in order to demonstrate the performance of the estimator in the case of small panel length. The number of panels considered is  $N = 2, 5, 10, 20, 50$ .

The correlation structure within each panel is modeled via random vectors generated from iid, AR(1) and GARCH(1,1) sequences. The considered AR(1) process has coefficient  $\phi = 0.3$ . In the case of the GARCH(1,1) process, we use coefficients  $\alpha_0 = 1$ ,  $\alpha_1 = 0.1$  and  $\beta_1 = 0.2$ , which, according to ([10], Example 1), give a strictly stationary process. In all three sequences, the innovations are obtained as iid random variables from a standard normal  $N(0, 1)$  or Student  $t_5$  distribution multiplied by a suitable constant so that the errors possess unit variance (see Assumption 1). The variance-scaling parameters are kept constant for all panels, i.e.,  $\sigma_i = \sigma$  for all  $i$ . The sequence of weights is chosen as  $\{w(t) = t^2\}_{t=1}^{10}$  and  $w(0) = 1$ . Monte Carlo simulation scenarios are produced as all possible combinations of the above-mentioned settings, and a selection of the results is listed below.

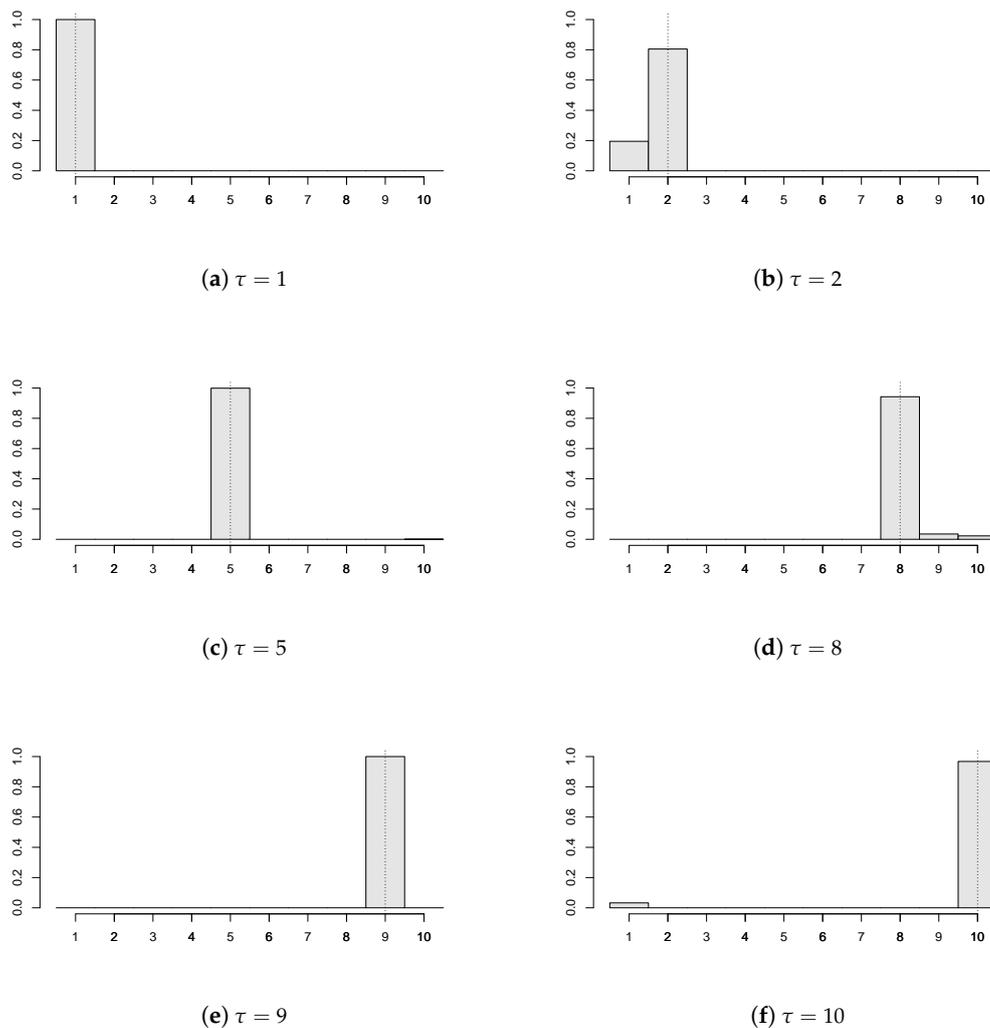
Firstly, we examine the impact of the errors' distribution and the correlation structure on the change point estimator. Figure 1 contains six different structures of model disturbances, where  $\tau = 8$  (depicted by the dotted vertical line),  $N = 20$ ,  $\sigma = 0.2$ , and all of the panels are subject to the break of value  $\delta_i \sim U[0, 2]$  (i.e., the breaks are independently and uniformly distributed on  $[0, 2]$ ).



**Figure 1.** Histograms of the estimated change points  $\hat{\tau}_N$  for various structures and distributions of the panel disturbances ( $\tau = 8, T = 10, N = 20, \sigma = 0.2$ ; all of the panels are subject to a break of size  $\delta_i \sim U[0, 2]$ ).

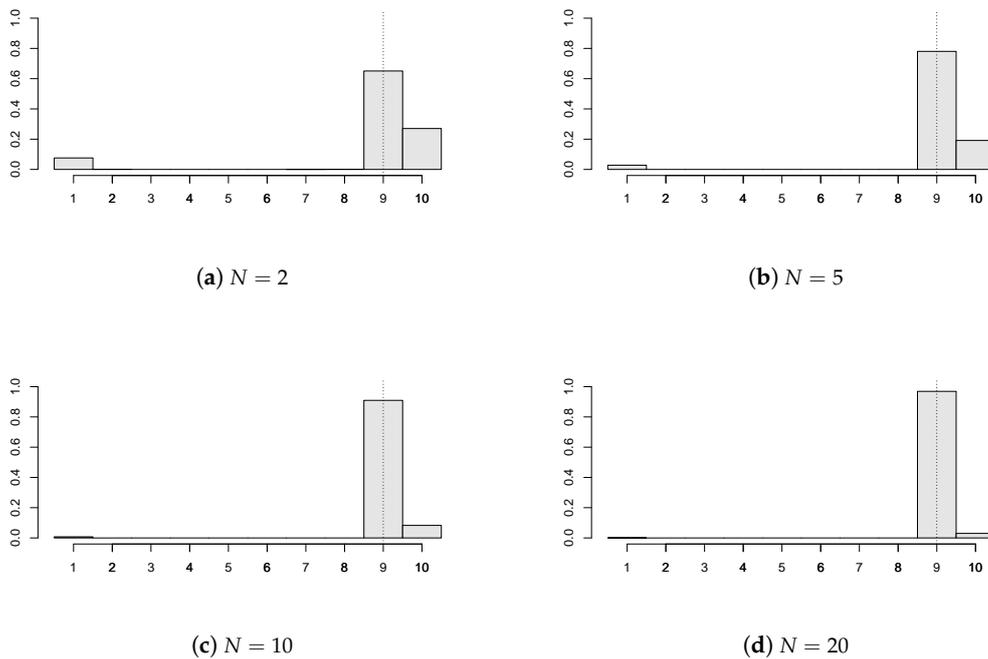
It can be concluded that the precision of our change point estimator is satisfactory even for a relatively small number of panels regardless of the errors' structure. Furthermore, innovations with heavier tails yield less precise estimators than innovations with lighter tails (i.e., compare Figure 1a,c,e, versus Figure 1b,d,f). One may notice that the AR(1) errors' model gives the best estimator's precision from three correlation structures. This should not be considered as a surprise, because our chosen AR(1) model has a positive autoregression coefficient ( $\phi = 0.3$ ), and therefore, the values of  $g(\tau)$  and  $g(T - \tau)$  from Assumption 3 are larger than the values corresponding to the iid errors' structure. Hence, one can say that the detectability Assumption 3 is satisfied more easily. Loosely speaking, the stronger the positive correlations within the panel, the more "deterministic the behavior" of the random noise and the better the estimator's precision.

Figure 2 demonstrates that the proposed estimator works reasonably for various locations of the unknown change point. Particularly, six values of the common change point (again, depicted by the dotted vertical line) are chosen ( $\tau = 1, 2, 5, 8, 9, 10$ ) with  $N = 20$ ,  $\sigma = 0.2$ ; 75% of the panels have a break  $\delta_i \sim U[0, 2]$ , and the panel disturbances come from AR(1) with  $N(0, 1)$  innovations. Recall that  $\tau = 10$  corresponds to the ‘no change’ situation, and the empirical distribution of the estimator concentrates mainly at the last time point, which is in coherence with the change point formulation from (1).



**Figure 2.** Histograms of the estimated change points  $\hat{\tau}_N$  for various values of the change point  $\tau$  ( $T = 10$ ,  $N = 20$ ,  $\sigma = 0.2$ , 75% of the panels are subject to a break of size  $\delta_i \sim U[0, 2]$  panel disturbances from AR(1) with  $N(0, 1)$  innovations).

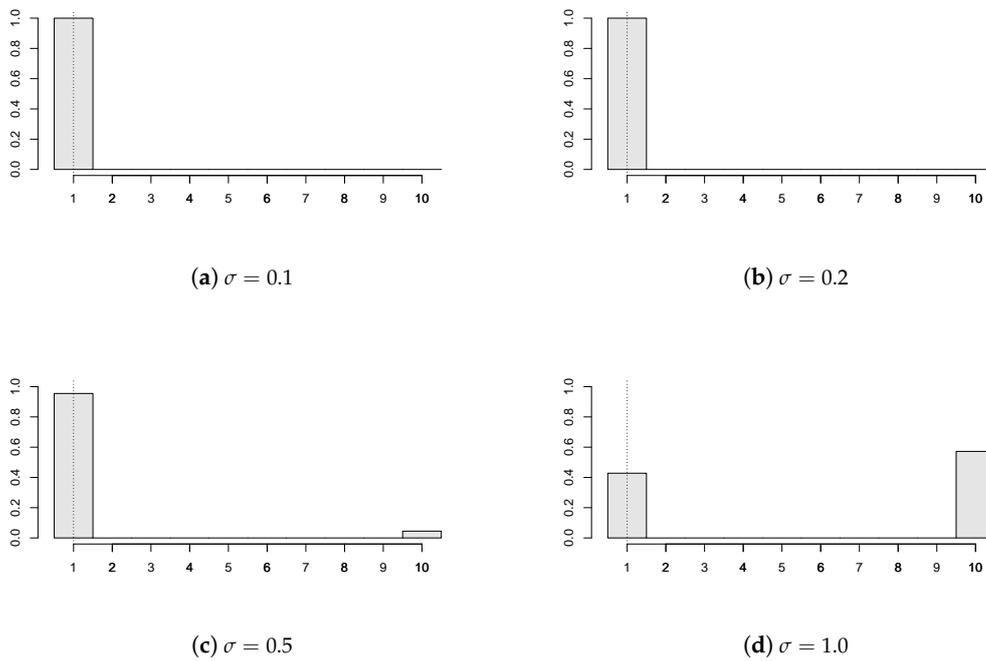
In Figure 3, the impact of the number of panels ( $N = 2, 5, 10, 20$ ) is investigated when  $\tau = 9$ ,  $\sigma = 0.2$ ; 50% of the panels have a break  $\delta_i \sim U[0, 2]$ , and the panel disturbances are AR(1) with  $t_5$  innovations. It is clear that the precision of  $\hat{\tau}_N$  improves markedly as  $N$  increases. A higher number of panels, i.e.,  $N = 50$ , were also taken into account, and then, 100% precision was achieved. Moreover, longer panels were also simulated (e.g.,  $T = 25$ ), but these results are not presented here. This is due to the simple reason that the precision of the estimator increases as the panel size gets bigger, which is straightforward and expected.



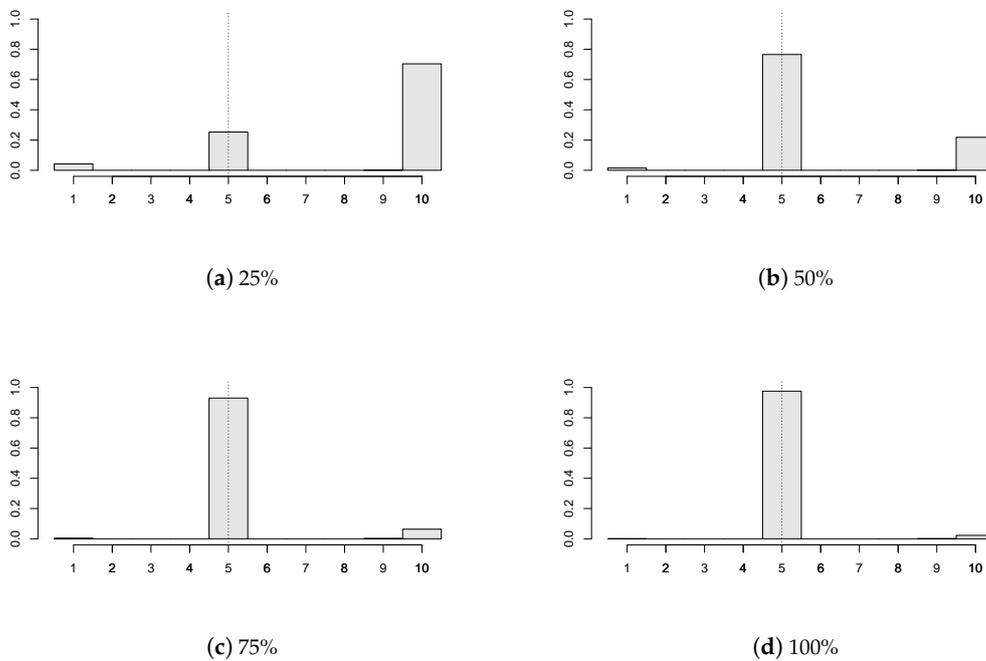
**Figure 3.** Histograms of the estimated change points  $\hat{\tau}_N$  for various values of  $N$  ( $\tau = 9$ ,  $T = 10$ ,  $\sigma = 0.2$ , 50% of the panels are subject to a break of size  $\delta_i \sim U[0, 2]$ , panel disturbances from AR(1) with  $t_5$  innovations).

Figure 4 shows the effect of panel variability on the estimator's performance. In particular, various values of the variance-scaling parameter are considered ( $\sigma = 0.1, 0.2, 0.5, 1.0$ ), where  $\tau = 1$ ,  $N = 10$ ; all of the panels have a break  $\delta_i \sim U[0, 2]$ , and the panel disturbances come from GARCH(1,1) with  $N(0, 1)$  innovations. It can be seen that the less volatile the observations, the more precise the change point estimate. The panel's variability under the considered dependency can be too high compared to the change size. Then, it would be rather difficult to detect a possible change, as for instance in Figure 4d, which corresponds to  $\sigma = 1.0$ .

In Figure 5, we examine how different portions of the panels with a change in mean influence the estimator's precision. Four cases were considered: 25%, 50%, 75%; and all of the panels have a break  $\delta_i \sim U[0, 2]$ . Here,  $\tau = 5$ ,  $N = 20$ ,  $\sigma = 0.2$ ; and the panel disturbances are GARCH(1,1) with  $t_5$  innovations. One can conclude that higher precision is obtained when a larger portion of panels is subject to change in the mean. If a small number of panels contain a break (for example, Figure 5a), then the change point estimator does not perform well.



**Figure 4.** Histograms of the estimated change points  $\hat{\tau}_N$  for various values of  $\sigma$  ( $\tau = 1, T = 10, N = 10$ , all of the panels are subject to a break of size  $\delta_i \sim U[0, 2]$ , panel disturbances from GARCH(1,1) with  $N(0, 1)$  innovations).



**Figure 5.** Histograms of the estimated change points  $\hat{\tau}_N$  when various portion of the panels are subject to a break of size  $\delta_i \sim U[0, 2]$  ( $\tau = 5, T = 10, N = 20, \sigma = 0.2$ , panel disturbances from GARCH(1,1) with  $t_5$  innovations).

### 5. Theoretical Usage in Hypothesis Testing

Estimation of structural breaks can become an important mid-step in many statistical procedures, e.g., estimation of the panel correlation structure or bootstrapping in hypothesis testing for the change point.

A possible theoretical application of the change point estimation can be motivated as follows. It is required to test the null hypothesis of no change in the means:

$$H_0 : \tau = T$$

against the alternative that at least one panel has a change in mean:

$$H_1 : \tau < T \quad \text{and} \quad \exists i \in \{1, \dots, N\} : \delta_i \neq 0.$$

Generally, a test statistic  $S_{N,T}$  for the change point detection may be constructed as a continuous function of the sums of cumulative residuals, cf. [2]. In particular,

$$S_{N,T} \equiv \mathcal{S} \left( \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{r=1}^s (Y_{i,r} - \bar{Y}_{i,t}) \right\}_{s=1,t=2}^{t-1,T}, \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{r=s+1}^T (Y_{i,r} - \tilde{Y}_{i,t}) \right\}_{s=t,t=1}^{T-1,T-1} \right),$$

where  $\mathcal{S}(\cdot, \cdot) : \mathbb{R}^{T \times (T-1)/2} \times \mathbb{R}^{T \times (T-1)/2} \rightarrow \mathbb{R}$  is continuous. To illustrate, a ratio type test statistic, discussed in [11],

$$S_{N,T} = \max_{t=2, \dots, T-2} \frac{\max_{s=1, \dots, t-1} \left| \sum_{i=1}^N \sum_{r=1}^s (Y_{i,r} - \bar{Y}_{i,t}) \right|}{\max_{s=t, \dots, T-1} \left| \sum_{i=1}^N \sum_{r=s+1}^T (Y_{i,r} - \tilde{Y}_{i,t}) \right|}$$

is generated by the continuous function

$$\mathcal{S} \left( \{a_{s,t}\}_{s=1,t=2}^{t-1,T}, \{b_{s,t}\}_{s=t,t=1}^{T-1,T-1} \right) = \max_{t=2, \dots, T-2} \frac{\max_{s=1, \dots, t-1} |a_{s,t}|}{\max_{s=t, \dots, T-1} |b_{s,t}|}.$$

The test statistic under the null would typically have a known limiting distribution (up to some unknown parameters), which is a function of a Gaussian vector or process. However, its correlation structure is unknown and needs to be estimated. Alternatively, one can avoid the estimation of the correlation structure (which may be considered as a nuisance parameter) by applying a suitable bootstrap procedure. The consistent change point estimator plays also an important role in the validity of the bootstrap algorithm.

#### 5.1. Estimation of Correlation Structure

Since the panels are considered to be independent and the number of panels may be sufficiently large, one can estimate the correlation structure of the errors  $[\varepsilon_{1,1}, \dots, \varepsilon_{1,T}]^\top$  empirically. We base the errors' estimates on residuals

$$\hat{e}_{i,t} := \begin{cases} Y_{i,t} - \bar{Y}_{i,\hat{\tau}_N}, & t \leq \hat{\tau}_N, \\ Y_{i,t} - \tilde{Y}_{i,\hat{\tau}_N}, & t > \hat{\tau}_N. \end{cases} \tag{4}$$

One may notice that the estimators that cannot result in the last time point are less suitable in the calculation of residuals.

Then, the empirical version of the autocorrelation function is

$$\hat{\rho}_t := \frac{1}{N(T-t)} \sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2} \sum_{s=1}^{T-t} \hat{e}_{i,s} \hat{e}_{i,s+t}$$

where  $\hat{\sigma}_i^2 := \frac{1}{T} \sum_{s=1}^T \hat{e}_{i,s}^2$  is the estimate of the variance parameter  $\sigma_i^2$ . Consequently, the cumulative autocorrelation function and shifted cumulative correlation function are estimated by

$$\hat{r}(t) = \sum_{|s|<t} (t - |s|)\hat{\rho}_s \quad \text{and} \quad \hat{R}(t, v) = \sum_{s=1}^t \sum_{u=t+1}^v \hat{\rho}_{u-s}, \quad t < v.$$

### 5.2. Bootstrapping

A wide range of literature has been published on bootstrapping in the change point problem, e.g., [2,12,13]. We build up the bootstrap test on the resampling with the replacement of row vectors  $\{\{\hat{e}_{i,1}, \dots, \hat{e}_{i,T}\}\}_{i=1, \dots, N}$  corresponding to the individual panels. This provides bootstrapped row vectors  $\{\{\hat{e}_{i,1}^*, \dots, \hat{e}_{i,T}^*\}\}_{i=1, \dots, N}$ . Then, the bootstrapped residuals  $\hat{e}_{i,t}^*$  are centered by their conditional expectation  $\frac{1}{N} \sum_{i=1}^N \hat{e}_{i,t}$ , yielding

$$\hat{Y}_{i,t}^* := \hat{e}_{i,t}^* - \frac{1}{N} \sum_{i=1}^N \hat{e}_{i,t}.$$

The bootstrap test statistic is just a modification of the original statistic  $\mathcal{S}_{N,T}$ , where the original observations  $Y_{i,t}$  are replaced by their bootstrap counterparts  $\hat{Y}_{i,t}^*$ :

$$\mathcal{S}_{N,T}^* \equiv \mathcal{S} \left( \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{r=1}^s (\hat{Y}_{i,r}^* - \tilde{Y}_{i,t}^*) \right\}_{s=1, t=2}^{t-1, T}, \left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^N \sum_{r=s+1}^T (\hat{Y}_{i,r}^* - \tilde{Y}_{i,t}^*) \right\}_{s=t, t=1}^{T-1, T-1} \right),$$

such that

$$\tilde{Y}_{i,t}^* = \frac{1}{t} \sum_{s=1}^t \hat{Y}_{i,s}^* \quad \text{and} \quad \tilde{Y}_{i,t}^* = \frac{1}{T-t} \sum_{s=t+1}^T \hat{Y}_{i,s}^*.$$

The idea behind bootstrapping is to mimic the original distribution of the test statistic by the distribution of the bootstrap test statistic, conditionally on the original data denoted by  $\mathbb{Y} \equiv \{Y_{i,t}\}_{i,t=1}^{N,T}$ . Recall that it is not known whether some common change in panel means occurred or not. In other words, one does not know whether the data come from the null or the alternative hypothesis.

**Theorem 2** (Bootstrap justification). *Suppose that Assumptions 1, 2 and 5 hold. Then,*

(i) under  $H_0$ ,

$$\mathcal{S}_{N,T} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \mathcal{L};$$

(ii) under additional Assumptions 3, 4 and under  $H_0$ , as well as under  $H_1$ ,

$$\mathcal{S}_{N,T}^* | \mathbb{Y} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \mathcal{L}^* \quad \text{in probability P};$$

(iii) under additional Assumptions 3, 4 and under  $H_0$ ,  $\mathcal{L}$  and  $\mathcal{L}^*$  coincide.

The validity of the bootstrap test is assured by Theorem 2. Indeed, the conditional asymptotic distribution of the bootstrap test statistic does not converge to infinity (in probability) under the alternative. In other words, the second part of Theorem 2 holds under  $H_0$ , as well as  $H_1$ . In contrast to the bootstrap version of the test statistics, the original test statistic typically explodes over all bounds under the alternative. That is why the bootstrap test statistic can be used correctly to reject the null in favor of the alternative, having sufficiently large  $N$ . Moreover, Theorem 2 states that the conditional distribution of the bootstrap test statistic and the unconditional distribution of the original test statistic coincide under the null. Additionally, that is the reason why the bootstrap test approximately keeps the same level as the original test based on the asymptotics (i.e., the test based on the asymptotic distribution of  $\mathcal{S}_{N,T}$ ).

A practical choice of the test statistic  $\mathcal{S}_{N,T}$  can be obtained from, e.g., [14]:

$$\mathcal{S}_{N,T} = \max_{t=2,\dots,T-2} \frac{\sum_{s=1}^{t-1} \left[ \sum_{i=1}^N \sum_{r=1}^s (Y_{i,r} - \bar{Y}_{i,t}) \right]^2}{\sum_{s=t}^{T-1} \left[ \sum_{i=1}^N \sum_{r=s+1}^T (Y_{i,r} - \tilde{Y}_{i,t}) \right]^2}$$

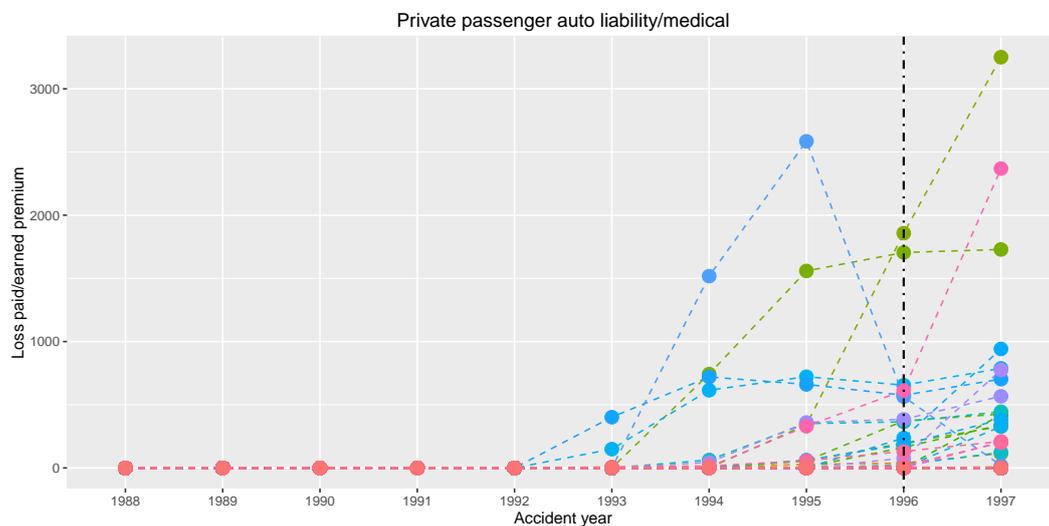
or

$$\mathcal{S}_{N,T} = \max_{t=2,\dots,T-2} \frac{\max_{s=1,\dots,t-1} \sum_{i=1}^N \sum_{r=1}^s (Y_{i,r} - \bar{Y}_{i,t}) - \min_{s=1,\dots,t-1} \sum_{i=1}^N \sum_{r=1}^s (Y_{i,r} - \bar{Y}_{i,t})}{\max_{s=t,\dots,T-1} \sum_{i=1}^N \sum_{r=s+1}^T (Y_{i,r} - \tilde{Y}_{i,t}) - \min_{s=t,\dots,T-1} \sum_{i=1}^N \sum_{r=s+1}^T (Y_{i,r} - \tilde{Y}_{i,t})}$$

Theorem 2 assures that the previously-described bootstrap algorithm can be used in hypothesis testing (change point detection) for the above-mentioned test statistics. Now, the simulated (empirical) distribution of the bootstrap test statistic can be used to calculate the bootstrap critical value, which will be compared to the value of the original test statistic in order to reject the null or not.

### 6. Practical Application in Non-Life Insurance

As mentioned in the Introduction, our primary motivation for the change point estimation in panel data comes from the non-life insurance business. The dataset is provided by the National Association of Insurance Commissioners (NAIC) database; see [15]. We concentrate on the ‘private passenger auto liability/medical’ insurance line of business. The data collect records from  $N = 146$  insurance companies. Each insurance company provides  $T = 10$  yearly total claim amounts starting from year 1988 up to year 1997. One can consider normalizing the claim amounts by the premium received by company  $i$  in year  $t$ . That is thinking of panel data  $Y_{i,t} / p_{i,t}$ , where  $p_{i,t}$  is the mentioned premium. This may yield a stabilization of series’ variability, which corresponds to Assumption 2 of bounded variances. Figure 6 graphically shows series of the normalized claim amounts.

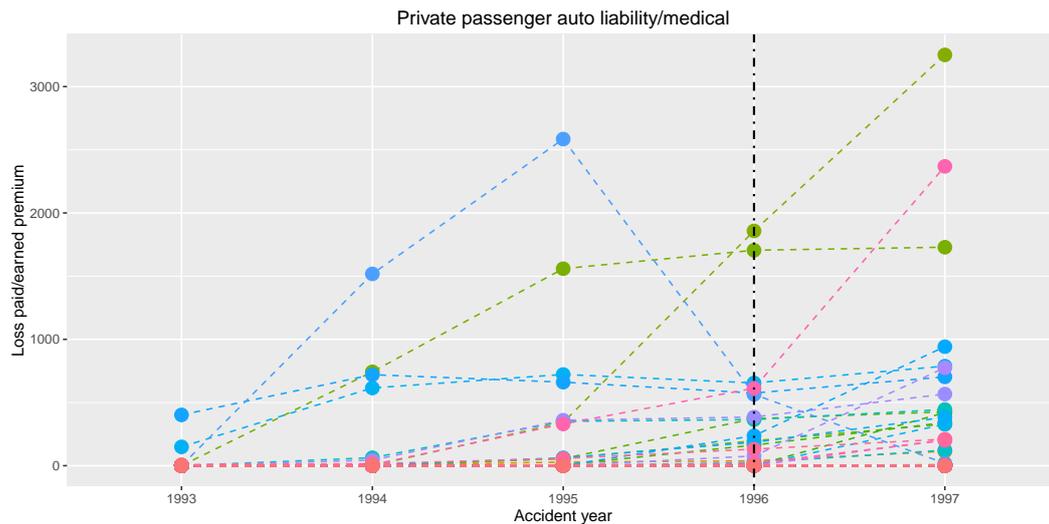


**Figure 6.** Development of the yearly total claim amounts normalized by the earned premium together with the estimated change point  $\hat{\tau}_{146} = 9$  (corresponding to year 1996).

The data are considered as panel data in the way that each insurance company corresponds to one panel, which is formed by the company’s yearly total claim amounts normalized by the earned premium. The length of the panel is quite short. This is very typical in insurance business, because considering longer panels may invoke incomparability between the early and the late claim amounts due to changing market or policies’ conditions over time.

We want to estimate a possible change in the normalized claim amounts occurring in a common year, assuming that the normalized claim amounts are approximately constant in the years before and after the possible change for every insurance company. Our change point estimator gives  $\hat{\tau}_{146} = 9$  (i.e., year 1996) using the sequence of weights  $\{t^2\}_{t=1}^{10}$  and  $w(0) = 1$ , which corresponds to the increased values for the last observed year 1997 shown in Figure 6.

An interesting finding comes out, when one concentrates only on the second half of the observation period; see Figure 7. The shortening of the time window up to years 1993–1997 yields the same change position, i.e., year 1996.



**Figure 7.** Development of the yearly total claim amounts normalized by the earned premium together for the second half of the original observation period with the estimated change point  $\hat{\tau}_{146} = 4$  (corresponding to year 1996).

Furthermore, the empirical cumulative autocorrelation function can be obtained. The correlation matrix is estimated as proposed in Section 5.1. The subsequence from the empirical cumulative autocorrelation function is obtained  $\{\hat{r}(t)\}_{t=1}^{10} = \{1.0, 2.5, 3.6, 4.6, 5.6, 6.6, 7.6, 8.6, 9.6, 10.6\}$ .

Dependent panels may be taken into account, and the presented work might be generalized for some kind of asymptotic independence over the panels or prescribed dependence among the panels. Nevertheless, our incentive is determined by a problem from non-life insurance, where the association of insurance companies consists of a relatively high number of insurance companies. Thus, the portfolio of yearly claims is so diversified, that the panels corresponding to insurance companies' yearly claims may be viewed as independent, and neither natural ordering, nor clustering has to be assumed.

### 7. Results and Conclusions

The change point problem in panel data with a fixed panel size is considered in this paper. A possible occurrence of common breaks in panel means is estimated. We introduce the change point estimator without the boundary issue, meaning that it can estimate the change close to the extremities of the studied time interval. The consistency of the estimator is proven regardless of the presence/absence of the change in panel means under relatively simple conditions.

The simulation study illustrates that the proposed change point estimator behaves sufficiently well for small panel sizes, a relatively moderate number of panels and various errors' dependence structures. The theoretical usage of the change point estimator is outlined for the estimation of the within-panel correlation structure and for the validity of the bootstrap procedures in hypothesis testing.

Finally, the proposed method is applied to an actuarial problem, for which the change point analysis in panel data provides an appealing approach.

**Acknowledgments:** Institutional support to Barbora Peřtová was provided by RVO:67985807. The research of Michal Peřta was supported by the Czech Science Foundation project GAČR No. 15-04774Y. The authors would like to thank three anonymous referees for careful reading of the paper and for providing suggestions that improved this paper.

**Author Contributions:** Both authors equally contributed to invention of the proposed change point estimator, to the derivation of its consistency, to designing and programming the simulation study and to showing its theoretical and practical applications.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Proofs

**Proof of Theorem 1.** Let us define

$$S_N^{(i,L)}(t) := \frac{1}{w(t)} \sum_{s=1}^t (Y_{i,s} - \bar{Y}_{i,t})^2, \quad S_N^{(i,R)}(t) := \frac{1}{w(T-t)} \sum_{s=t+1}^T (Y_{i,s} - \tilde{Y}_{i,t})^2$$

and, consequently,  $S_N(t) := \frac{1}{N} \sum_{i=1}^N \{S_N^{(i,L)}(t) + S_N^{(i,R)}(t)\}$ . Then,

$$S_N^{(i,L)}(t) = \begin{cases} \frac{\sigma_i^2}{w(t)} \sum_{s=1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2, & t \leq \tau, \\ \frac{1}{w(t)} \left[ \sum_{s=1}^{\tau} (\sigma_i \varepsilon_{i,s} - \sigma_i \bar{\varepsilon}_{i,t} - \frac{t-\tau}{t} \delta_i)^2 + \sum_{s=\tau+1}^t (\sigma_i \varepsilon_{i,s} - \sigma_i \bar{\varepsilon}_{i,t} + \frac{\tau-t}{t} \delta_i)^2 \right], & t > \tau; \end{cases}$$

where  $\bar{\varepsilon}_{i,t} = \frac{1}{t} \sum_{s=1}^t \varepsilon_{i,s}$ . Similarly,

$$S_N^{(i,R)}(t) = \begin{cases} \frac{\sigma_i^2}{w(T-t)} \sum_{s=t+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2, & t > \tau, \\ \frac{1}{w(T-t)} \left[ \sum_{s=t+1}^{\tau} (\sigma_i \varepsilon_{i,s} - \sigma_i \tilde{\varepsilon}_{i,t} - \frac{T-\tau}{T-t} \delta_i)^2 + \sum_{s=\tau+1}^T (\sigma_i \varepsilon_{i,s} - \sigma_i \tilde{\varepsilon}_{i,t} + \frac{\tau-t}{T-t} \delta_i)^2 \right], & t \leq \tau < T, \\ \frac{\sigma_i^2}{w(T-t)} \sum_{s=t+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2, & t \leq \tau = T; \end{cases}$$

where  $\tilde{\varepsilon}_{i,t} = \frac{1}{T-t} \sum_{s=t+1}^T \varepsilon_{i,s}$  for  $t < T$  and  $\tilde{\varepsilon}_{i,T} \equiv 0$ . By the definition of the cumulative autocorrelation function, we have for  $2 \leq t \leq \tau$ :

$$ES_N^{(i,L)}(t) = \frac{\sigma_i^2}{w(t)} \sum_{s=1}^t E(\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2 = \frac{\sigma_i^2}{w(t)} \sum_{s=1}^t \left[ 1 - \frac{2}{t} \sum_{r=1}^t E\varepsilon_{i,s} \varepsilon_{i,r} + \frac{1}{t^2} r(t) \right] = \frac{\sigma_i^2}{w(t)} \left( t - \frac{r(t)}{t} \right).$$

Clearly,  $S_N^{(i,L)}(1) = 0$ . In the remaining case when  $t > \tau$ , one can calculate

$$\begin{aligned} ES_N^{(i,L)}(t) &= \frac{\sigma_i^2}{w(t)} \left( t - \frac{r(t)}{t} \right) + \frac{\tau}{w(t)} \left( \frac{t-\tau}{t} \right)^2 \delta_i^2 + \frac{t-\tau}{w(t)} \left( \frac{\tau}{t} \right)^2 \delta_i^2 \\ &= \frac{\sigma_i^2 t}{w(t)} \left( 1 - \frac{r(t)}{t^2} \right) + \frac{\tau(t-\tau)}{tw(t)} \delta_i^2. \end{aligned}$$

By the definition of an empty sum,  $S_N^{(i,R)}(T) = 0$  and, moreover,  $S_N^{(i,R)}(T-1) = 0$ . For  $T-1 > t > \tau$ ,

$$\begin{aligned} ES_N^{(i,R)}(t) &= \frac{\sigma_i^2}{w(T-t)} \sum_{s=t+1}^T E(\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 \\ &= \frac{\sigma_i^2}{w(T-t)} \sum_{s=t+1}^T \left[ 1 - \frac{2}{T-t} \sum_{r=t+1}^T E\varepsilon_{i,s} \varepsilon_{i,r} + \frac{1}{(T-t)^2} r(T-t) \right] \end{aligned}$$

$$= \frac{\sigma_i^2}{w(T-t)} \left( T-t - \frac{r(T-t)}{T-t} \right).$$

The same result is obtained for  $T = \tau > t$ . In the remaining case  $T - 1 \geq \tau \geq t$ , such that  $T - 1 > t$ ,

$$\begin{aligned} \text{ES}_N^{(i,R)}(t) &= \frac{\sigma_i^2}{w(T-t)} \left( T-t - \frac{r(T-t)}{T-t} \right) + \frac{\tau-t}{w(T-t)} \left( \frac{T-\tau}{T-t} \right)^2 \delta_i^2 + \frac{T-\tau}{w(T-t)} \left( \frac{\tau-t}{T-t} \right)^2 \delta_i^2 \\ &= \frac{\sigma_i^2(T-t)}{w(T-t)} \left( 1 - \frac{r(T-t)}{(T-t)^2} \right) + \frac{(\tau-t)(T-\tau)}{(T-t)w(T-t)} \delta_i^2. \end{aligned}$$

Realize that  $S_N^{(i,L)}(t) + S_N^{(i,R)}(t) - \text{ES}_N^{(i,L)}(t) - \text{ES}_N^{(i,R)}(t)$  are independent with zero mean for fixed  $t$  and  $i = 1, \dots, N$ , but they are not identically distributed. Due to Assumption 5 and the stochastic Cauchy-Schwarz inequality, for  $t \leq \tau < T$ , it holds

$$\begin{aligned} \text{Var } S_N(t) &= \frac{1}{N^2} \sum_{i=1}^N \left\{ \frac{\sigma_i^4}{w^2(t)} \text{Var} \left[ \sum_{s=1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2 \right] \right. \\ &\quad + \frac{1}{w^2(T-t)} \text{Var} \left[ \sigma_i^2 \sum_{s=t+1}^{\tau} (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 - 2 \frac{T-\tau}{T-t} \sigma_i \delta_i \sum_{s=t+1}^{\tau} (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t}) + \left( \frac{T-\tau}{T-t} \right)^2 \delta_i^2 \right. \\ &\quad \left. \left. + \sigma_i^2 \sum_{s=\tau+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 + 2 \frac{\tau-t}{T-t} \sigma_i \delta_i \sum_{s=\tau+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t}) + \left( \frac{\tau-t}{T-t} \right)^2 \delta_i^2 \right] \right. \\ &\quad \left. + \frac{2\sigma_i^2}{w(t)w(T-t)} \text{Cov} \left[ \sum_{s=1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2, \sigma_i^2 \sum_{s=t+1}^{\tau} (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 - 2 \frac{T-\tau}{T-t} \sigma_i \delta_i \sum_{s=t+1}^{\tau} (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t}) \right. \right. \\ &\quad \left. \left. + \left( \frac{T-\tau}{T-t} \right)^2 \delta_i^2 + \sigma_i^2 \sum_{s=\tau+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 + 2 \frac{\tau-t}{T-t} \sigma_i \delta_i \sum_{s=\tau+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t}) + \left( \frac{\tau-t}{T-t} \right)^2 \delta_i^2 \right] \right\} \\ &\leq \frac{1}{N} C_1(t, \tau, \sigma, \bar{\sigma}) + \frac{1}{N^2} C_2(t, \tau, \sigma, \bar{\sigma}) \sum_{i=1}^N \delta_i^2 + \frac{1}{N^2} C_3(t, \tau, \sigma, \bar{\sigma}) \left| \sum_{i=1}^N \delta_i \right|, \end{aligned}$$

where  $C_1(t, \tau, \sigma, \bar{\sigma}) > 0$ ,  $C_2(t, \tau, \sigma, \bar{\sigma}) \geq 0$  and  $C_3(t, \tau, \sigma, \bar{\sigma}) \geq 0$  are some constants not depending on  $N$ . If  $t < \tau = T$ , then

$$\begin{aligned} \text{Var } S_N(t) &= \frac{1}{N^2} \sum_{i=1}^N \left\{ \frac{\sigma_i^4}{w^2(t)} \text{Var} \left[ \sum_{s=1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2 \right] + \frac{\sigma_i^4}{w^2(T-t)} \text{Var} \left[ \sum_{s=t+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 \right] \right. \\ &\quad \left. + \frac{2\sigma_i^4}{w(t)w(T-t)} \text{Cov} \left[ \sum_{s=1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2, \sum_{s=t+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 \right] \right\} \leq \frac{1}{N} C_4(t, \tau, \sigma, \bar{\sigma}), \end{aligned}$$

where  $C_4(t, \tau, \sigma, \bar{\sigma}) > 0$  does not depend on  $N$ . In the case of  $t = \tau = T$ , we also have  $\text{Var } S_N(T) \leq \frac{1}{N} C_5(T, \tau, \sigma, \bar{\sigma})$ , where  $C_5(T, \tau, \sigma, \bar{\sigma}) > 0$  does not depend on  $N$ . Finally, if  $t > \tau$ , then

$$\begin{aligned} \text{Var } S_N(t) &= \frac{1}{N^2} \sum_{i=1}^N \left\{ \frac{\sigma_i^4}{w^2(T-t)} \text{Var} \left[ \sum_{s=t+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2 \right] + \frac{1}{w^2(t)} \text{Var} \left[ \sigma_i^2 \sum_{s=1}^{\tau} (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2 \right. \right. \\ &\quad \left. \left. - 2 \frac{t-\tau}{t} \sigma_i \delta_i \sum_{s=1}^{\tau} (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t}) + \left( \frac{t-\tau}{t} \right)^2 \delta_i^2 + \sigma_i^2 \sum_{s=\tau+1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2 \right. \right. \\ &\quad \left. \left. + 2 \frac{\tau}{t} \sigma_i \delta_i \sum_{s=\tau+1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t}) + \left( \frac{\tau}{t} \right)^2 \delta_i^2 \right] \right\} \end{aligned}$$

$$\begin{aligned}
 & + \frac{2\sigma_i^2}{w(T-t)w(t)} \text{Cov} \left[ \sum_{s=t+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,t})^2, \sigma_i^2 \sum_{s=1}^{\tau} (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2 - 2\frac{t-\tau}{t} \sigma_i \delta_i \sum_{s=1}^{\tau} (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t}) \right. \\
 & \left. + \left( \frac{t-\tau}{t} \right)^2 \delta_i^2 + \sigma_i^2 \sum_{s=\tau+1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t})^2 + 2\frac{\tau}{t} \sigma_i \delta_i \sum_{s=\tau+1}^t (\varepsilon_{i,s} - \bar{\varepsilon}_{i,t}) + \left( \frac{\tau}{t} \right)^2 \delta_i^2 \right] \\
 & \leq \frac{1}{N} D_1(t, \tau, \underline{\sigma}, \bar{\sigma}) + \frac{1}{N^2} D_2(t, \tau, \underline{\sigma}, \bar{\sigma}) \sum_{i=1}^N \delta_i^2 + \frac{1}{N^2} D_3(t, \tau, \underline{\sigma}, \bar{\sigma}) \left| \sum_{i=1}^N \delta_i \right|,
 \end{aligned}$$

where  $D_1(t, \tau, \underline{\sigma}, \bar{\sigma}) > 0$ ,  $D_2(t, \tau, \underline{\sigma}, \bar{\sigma}) \geq 0$  and  $D_3(t, \tau, \underline{\sigma}, \bar{\sigma}) \geq 0$  do not depend on  $N$ . The Chebyshev inequality provides  $S_N(t) - ES_N(t) = \mathcal{O}_P \left( \sqrt{\text{Var } S_N(t)} \right)$  as  $N \rightarrow \infty$ . According to Assumption 4 and the Cauchy–Schwarz inequality, we have

$$\frac{1}{N^2} \left| \sum_{i=1}^N \delta_i \right| \leq \frac{1}{N} \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2} \rightarrow 0, \quad N \rightarrow \infty.$$

Since the index set  $\{1, \dots, T\}$  is finite and  $\tau$  is finite, as well, then

$$\max_{1 \leq t \leq T} \text{Var } S_N(t) \leq \frac{1}{N} K_1(\underline{\sigma}, \bar{\sigma}) + K_2(\underline{\sigma}, \bar{\sigma}) \frac{1}{N^2} \sum_{i=1}^N \delta_i^2 + K_3(\underline{\sigma}, \bar{\sigma}) \frac{1}{N^2} \left| \sum_{i=1}^N \delta_i \right| \leq \frac{1}{N} K_4(\underline{\sigma}, \bar{\sigma}),$$

where  $K_1(\underline{\sigma}, \bar{\sigma}) > 0$ ,  $K_2(\underline{\sigma}, \bar{\sigma}) \geq 0$ ,  $K_3(\underline{\sigma}, \bar{\sigma}) \geq 0$  and  $K_4(\underline{\sigma}, \bar{\sigma}) > 0$  are constants not depending on  $N$ . Thus, we also have uniform stochastic boundedness, i.e.,

$$\max_{1 \leq t \leq T} |S_N(t) - ES_N(t)| = \mathcal{O}_P \left( \frac{1}{\sqrt{N}} \right), \quad N \rightarrow \infty.$$

Adding and subtracting, one has

$$\begin{aligned}
 S_N(t) - S_N(\tau) & = S_N(t) - ES_N(t) - [S_N(\tau) - ES_N(\tau)] + ES_N(t) - ES_N(\tau) \\
 & \geq -2 \max_{1 \leq r \leq T} |S_N(r) - ES_N(r)| + ES_N(t) - ES_N(\tau) \\
 & = -2 \max_{1 \leq r \leq T} |S_N(r) - ES_N(r)| + \frac{1}{N} \left( \sum_{i=1}^N \sigma_i^2 \right) \left[ \frac{t}{w(t)} \left( 1 - \frac{r(t)}{t^2} \right) - \frac{\tau}{w(\tau)} \left( 1 - \frac{r(\tau)}{\tau^2} \right) \right. \\
 & \quad \left. + \mathcal{I}\{t < T\} \frac{T-t}{w(T-t)} \left( 1 - \frac{r(T-t)}{(T-t)^2} \right) - \mathcal{I}\{\tau < T\} \frac{T-\tau}{w(T-\tau)} \left( 1 - \frac{r(T-\tau)}{(T-\tau)^2} \right) \right] \\
 & \quad + \frac{1}{N} \left( \sum_{i=1}^N \delta_i^2 \right) \left[ \mathcal{I}\{t > \tau\} \frac{\tau(t-\tau)}{tw(t)} + \mathcal{I}\{t < \tau\} \frac{(\tau-t)(T-\tau)}{(T-t)w(T-t)} \right] \tag{A1}
 \end{aligned}$$

for each  $t \in \{1, \dots, T\}$ . Particularly, Inequality (A1) holds for  $\hat{\tau}_N$ . Note that  $\hat{\tau}_N = \arg \min_t S_N(t)$ . Hence,  $S_N(\hat{\tau}_N) - S_N(\tau) \leq 0$ . Therefore,

$$\begin{aligned}
 & 2\sqrt{N} \max_{1 \leq r \leq T} |S_N(r) - ES_N(r)| \tag{A2} \\
 & \geq \left[ \mathcal{I}\{\hat{\tau}_N > \tau\} \frac{\tau(\hat{\tau}_N - \tau)}{\hat{\tau}_N w(\hat{\tau}_N)} + \mathcal{I}\{\hat{\tau}_N < \tau\} \frac{(\tau - \hat{\tau}_N)(T - \tau)}{(T - \hat{\tau}_N)w(T - \hat{\tau}_N)} \right] \frac{1}{\sqrt{N}} \sum_{i=1}^N \delta_i^2 \\
 & \quad + \left[ \frac{\hat{\tau}_N}{w(\hat{\tau}_N)} \left( 1 - \frac{r(\hat{\tau}_N)}{\hat{\tau}_N^2} \right) - \frac{\tau}{w(\tau)} \left( 1 - \frac{r(\tau)}{\tau^2} \right) \right. \\
 & \quad \left. + \mathcal{I}\{\hat{\tau}_N < T\} \frac{T - \hat{\tau}_N}{w(T - \hat{\tau}_N)} \left( 1 - \frac{r(T - \hat{\tau}_N)}{(T - \hat{\tau}_N)^2} \right) - \mathcal{I}\{\tau < T\} \frac{T - \tau}{w(T - \tau)} \left( 1 - \frac{r(T - \tau)}{(T - \tau)^2} \right) \right] \frac{1}{\sqrt{N}} \sum_{i=1}^N \sigma_i^2
 \end{aligned}$$

$$\begin{aligned}
 &= \mathcal{I}\{\widehat{\tau}_N > \tau\} \frac{1}{\sqrt{N}} \left\{ \frac{\tau(\widehat{\tau}_N - \tau)}{\widehat{\tau}_N w(\widehat{\tau}_N)} \sum_{i=1}^N \delta_i^2 + [g(\widehat{\tau}_N) - g(\tau) + g(T - \widehat{\tau}_N) - g(T - \tau)] \sum_{i=1}^N \sigma_i^2 \right\} \\
 &\quad + \mathcal{I}\{\widehat{\tau}_N < \tau\} \frac{1}{\sqrt{N}} \left\{ \frac{(\tau - \widehat{\tau}_N)(T - \tau)}{(T - \widehat{\tau}_N)w(T - \widehat{\tau}_N)} \sum_{i=1}^N \delta_i^2 + [g(\widehat{\tau}_N) - g(\tau) + g(T - \widehat{\tau}_N) - g(T - \tau)] \sum_{i=1}^N \sigma_i^2 \right\} \\
 &\geq \mathcal{I}\{\widehat{\tau}_N > \tau\} \frac{1}{\sqrt{N}} \left\{ \frac{\tau}{w(\widehat{\tau}_N)} \left(1 - \frac{\tau}{\widehat{\tau}_N}\right) \sum_{i=1}^N \delta_i^2 - [g(\tau) + g(T - \tau)] \sum_{i=1}^N \sigma_i^2 \right\} \\
 &\quad + \mathcal{I}\{\widehat{\tau}_N < \tau\} \frac{1}{\sqrt{N}} \left\{ \frac{T - \tau}{w(T - \widehat{\tau}_N)} \left(1 - \frac{T - \tau}{T - \widehat{\tau}_N}\right) \sum_{i=1}^N \delta_i^2 - [g(\tau) + g(T - \tau)] \sum_{i=1}^N \sigma_i^2 \right\} \\
 &\geq \mathcal{I}\{\widehat{\tau}_N > \tau\} \frac{1}{\sqrt{N}} \left\{ \frac{\tau}{(\tau + 1) \max_{t=1, \dots, T} w(t)} \sum_{i=1}^N \delta_i^2 - [g(\tau) + g(T - \tau)] \sum_{i=1}^N \sigma_i^2 \right\} \tag{A3}
 \end{aligned}$$

$$\begin{aligned}
 &\quad + \mathcal{I}\{\widehat{\tau}_N < \tau\} \frac{1}{\sqrt{N}} \left\{ \frac{T - \tau}{(T - \tau + 1) \max_{t=1, \dots, T} w(t)} \sum_{i=1}^N \delta_i^2 - [g(\tau) + g(T - \tau)] \sum_{i=1}^N \sigma_i^2 \right\}, \tag{A4}
 \end{aligned}$$

where  $g(t) = \frac{t}{w(t)} \left(1 - \frac{r(t)}{t^2}\right) \geq 0$  for  $t \in \{1, \dots, T\}$  and  $g(0) \equiv 0$ . Since the expression in (A2) is  $\mathcal{O}_{\mathcal{P}}(1)$  as  $N \rightarrow \infty$ , we have  $\mathcal{I}\{\widehat{\tau}_N > \tau\} \xrightarrow{\mathbb{P}} 0$ , as well as  $\mathcal{I}\{\widehat{\tau}_N < \tau\} \xrightarrow{\mathbb{P}} 0$ , due to Assumption 3 applied in (A3) and (A4). Hence,  $\mathbb{P}[\widehat{\tau}_N = \tau] \rightarrow 1$  as  $N \rightarrow \infty$ .  $\square$

**Proof of Theorem 2.** (i) Let us define

$$U_N(t) := \sum_{i=1}^N \sum_{s=1}^t (Y_{i,s} - \mu_i).$$

Using the multivariate Lyapunov CLT for a sequence of  $T$ -dimensional independent random vectors  $\left\{ \sigma_i \left[ \sum_{s=1}^1 \varepsilon_{i,s}, \dots, \sum_{s=1}^T \varepsilon_{i,s} \right]^\top \right\}_{i \in \mathbb{N}}$ , we have under  $H_0$ :

$$\Sigma_N^{-1/2} [U_N(1), \dots, U_N(T)]^\top \xrightarrow[N \rightarrow \infty]{\mathcal{D}} [X_1, \dots, X_T]^\top,$$

such that

$$\Sigma_N = \sum_{i=1}^N \sigma_i^2 \text{Var} \left[ \sum_{s=1}^1 \varepsilon_{i,s}, \dots, \sum_{s=1}^T \varepsilon_{i,s} \right]^\top = \zeta_N^2 \Lambda,$$

where  $\Lambda = \text{Var} \left[ \sum_{s=1}^1 \varepsilon_{1,s}, \dots, \sum_{s=1}^T \varepsilon_{1,s} \right]^\top$  is the positive definite covariance matrix with respect to Assumption 1 and  $0 < N\sigma^2 \leq \zeta_N^2 := \sum_{i=1}^N \sigma_i^2 \leq N\bar{\sigma}^2$  according to Assumption 2. The limiting  $T$ -dimensional random vector  $[X_1, \dots, X_T]^\top$  has a multivariate normal distribution with zero mean and the identity covariance matrix. The Lyapunov condition is satisfied due to the Jensen inequality, the Cramér–Wold theorem and Assumption 5, i.e.,

$$\begin{aligned}
 &\left( \mathbf{a}^\top \Sigma_N \mathbf{a} \right)^{-\frac{2+\chi}{2}} \sum_{i=1}^N \mathbb{E} \left| \mathbf{a}^\top \sigma_i \left[ \sum_{s=1}^1 \varepsilon_{i,s}, \dots, \sum_{s=1}^T \varepsilon_{i,s} \right]^\top \right|^{2+\chi} \\
 &= \zeta_N^{-2-\chi} \left( \mathbf{a}^\top \Lambda \mathbf{a} \right)^{-\frac{2+\chi}{2}} \sum_{i=1}^N \sigma_i^{2+\chi} \mathbb{E} \left| \sum_{t=1}^T a_t \sum_{s=1}^t \varepsilon_{i,s} \right|^{2+\chi} \\
 &\leq T^{1+\chi} \zeta_N^{-2-\chi} \left( \mathbf{a}^\top \Lambda \mathbf{a} \right)^{-\frac{2+\chi}{2}} \sum_{i=1}^N \sigma_i^{2+\chi} \sum_{t=1}^T \mathbb{E} \left( a_t \sum_{s=1}^t \varepsilon_{i,s} \right)^{2+\chi}
 \end{aligned}$$

$$\begin{aligned} &\leq T^{1+\chi} \zeta_N^{-2-\chi} (\mathbf{a}^\top \mathbf{\Lambda} \mathbf{a})^{-\frac{2+\chi}{2}} \sum_{i=1}^N \sigma_i^{2+\chi} \sum_{t=1}^T |a_t|^{2+\chi} t^{1+\chi} \sum_{s=1}^t \mathbb{E} |\varepsilon_{t,s}|^{2+\chi} \\ &\leq \varrho \zeta_N^{-2-\chi} \sum_{i=1}^N \sigma_i^{2+\chi} \leq \varrho (N \underline{\sigma}^2)^{-\frac{2+\chi}{2}} N \bar{\sigma}^{2+\chi} = \varrho \underline{\sigma}^{-2-\chi} \bar{\sigma}^{2+\chi} N^{-\frac{\chi}{2}} \rightarrow 0, \quad N \rightarrow \infty \end{aligned} \tag{A5}$$

for arbitrary fixed  $\mathbf{0} \neq \mathbf{a} = [a_1, \dots, a_T]^\top \in \mathbb{R}^T$  and some  $0 < \chi \leq 2$ , where

$$\varrho = T^{1+\chi} (\mathbf{a}^\top \mathbf{\Lambda} \mathbf{a})^{-\frac{2+\chi}{2}} \sum_{t=1}^T |a_t|^{2+\chi} t^{1+\chi} \sum_{s=1}^t \mathbb{E} |\varepsilon_{1,s}|^{2+\chi}$$

is a positive constant not depending on  $N$ . The  $t$ -th diagonal element of the covariance matrix  $\mathbf{\Lambda}$  is

$$\text{Var} \sum_{s=1}^t \varepsilon_{1,s} = r(t)$$

and the upper off-diagonal element on position  $(t, v)$  is

$$\text{Cov} \left( \sum_{s=1}^t \varepsilon_{1,s}, \sum_{u=1}^v \varepsilon_{1,u} \right) = \text{Var} \sum_{s=1}^t \varepsilon_{1,s} + \text{Cov} \left( \sum_{s=1}^t \varepsilon_{1,s}, \sum_{u=t+1}^v \varepsilon_{1,u} \right) = r(t) + R(t, v), \quad t < v.$$

Moreover, let us define the reverse analogue of  $U_N(t)$ , i.e.,

$$V_N(t) := \sum_{i=1}^N \sum_{s=t+1}^T (Y_{i,s} - \mu_i) = U_N(T) - U_N(t).$$

Hence,

$$U_N(s) - \frac{s}{t} U_N(t) = \sum_{i=1}^N \left\{ \sum_{r=1}^s \left[ (Y_{i,r} - \mu_i) - \frac{1}{t} \sum_{v=1}^t (Y_{i,v} - \mu_i) \right] \right\} = \sum_{i=1}^N \sum_{r=1}^s (Y_{i,r} - \bar{Y}_{i,t})$$

and, consequently,

$$V_N(s) - \frac{T-s}{T-t} V_N(t) = \sum_{i=1}^N \left\{ \sum_{r=s+1}^T \left[ (Y_{i,r} - \mu_i) - \frac{1}{T-t} \sum_{v=t+1}^T (Y_{i,v} - \mu_i) \right] \right\} = \sum_{i=1}^N \sum_{r=s+1}^T (Y_{i,r} - \tilde{Y}_{i,t}),$$

for  $t < T$ . Then, under  $H_0$ :

$$\tilde{\Sigma}_N^{-1/2} [V_N(1), \dots, V_N(T-1)]^\top \xrightarrow[N \rightarrow \infty]{\mathcal{D}} [Z_1, \dots, Z_{T-1}]^\top,$$

where  $Z_t := X_T - X_t$  and

$$\tilde{\Sigma}_N = \sum_{i=1}^N \sigma_i^2 \text{Var} \left[ \sum_{s=2}^T \varepsilon_{i,s}, \dots, \sum_{s=T}^T \varepsilon_{i,s} \right]^\top = \zeta_N^2 \tilde{\mathbf{\Lambda}}$$

for  $\tilde{\mathbf{\Lambda}} = \text{Var} \left[ \sum_{s=2}^T \varepsilon_{1,s}, \dots, \sum_{s=T}^T \varepsilon_{1,s} \right]^\top$ . Using the continuous mapping theorem, we end up with

$$\mathcal{S}_{N,T} \xrightarrow[N \rightarrow \infty]{\mathcal{D}} \mathcal{L}$$

such that the law  $\mathcal{L}$  corresponds to the distribution of

$$\mathcal{S} \left( \left\{ X_s - \frac{s}{t} X_t \right\}_{s=1, t=1}^{t-1, T}, \left\{ Z_s - \frac{T-s}{T-t} Z_t \right\}_{s=t, t=1}^{T-1, T-1} \right).$$

(ii) Let us define  $\widehat{\epsilon}_{i,t} := \sum_{s=1}^t \widehat{e}_{i,s}$ ,  $\widehat{\epsilon}_{i,t}^* := \sum_{s=1}^t \widehat{e}_{i,s}^*$ ,

$$\widehat{U}_N(t) := \sum_{i=1}^N \sum_{s=1}^t \widehat{e}_{i,s} = \sum_{i=1}^N \widehat{\epsilon}_{i,t},$$

and

$$\widehat{U}_N^*(t) := \sum_{i=1}^N \sum_{s=1}^t \widehat{Y}_{i,s}^* = \sum_{i=1}^N \sum_{s=1}^t \left( \widehat{e}_{i,s}^* - \frac{1}{N} \sum_{i=1}^N \widehat{e}_{i,s} \right) = \sum_{i=1}^N \sum_{s=1}^t (\widehat{e}_{i,s}^* - \widehat{e}_{i,s}) = \sum_{i=1}^N (\widehat{\epsilon}_{i,t}^* - \widehat{\epsilon}_{i,t}).$$

Realize that  $\widehat{\epsilon}_{i,t}$  depends on  $\widehat{\tau}_N$ , and hence, it depends on  $N$ .

Let us calculate  $\lim_{N \rightarrow \infty} \mathbf{\Gamma}_{i,N}$ , where  $\mathbf{\Gamma}_{i,N} = \text{Var} [\widehat{\epsilon}_{i,1}, \dots, \widehat{\epsilon}_{i,T}]^\top$ . Using the law of total variance,

$$\begin{aligned} \text{Var} \widehat{\epsilon}_{i,t} &= \text{E}[\text{Var} \{\widehat{\epsilon}_{i,t} | \widehat{\tau}_N\}] + \text{Var} [\text{E}\{\widehat{\epsilon}_{i,t} | \widehat{\tau}_N\}] = \sum_{\pi=1}^T \text{P}[\widehat{\tau}_N = \pi] \text{Var} [\widehat{\epsilon}_{i,t} | \widehat{\tau}_N = \pi] \\ &\quad + \sum_{\pi=1}^T \text{P}[\widehat{\tau}_N = \pi] \{ \text{E}[\widehat{\epsilon}_{i,t} | \widehat{\tau}_N = \pi] \}^2 - \left\{ \sum_{\pi=1}^T \text{P}[\widehat{\tau}_N = \pi] \text{E}[\widehat{\epsilon}_{i,t} | \widehat{\tau}_N = \pi] \right\}^2. \end{aligned}$$

Since  $\lim_{N \rightarrow \infty} \text{P}[\widehat{\tau}_N = \tau] = 1$  and  $\text{E}[\widehat{\epsilon}_{i,t} | \widehat{\tau}_N = \tau] = 0$ , then

$$\lim_{N \rightarrow \infty} \text{Var} \widehat{\epsilon}_{i,t} = \lim_{N \rightarrow \infty} \text{Var} [\widehat{\epsilon}_{i,t} | \widehat{\tau}_N = \tau].$$

Similarly with the covariance, i.e., after applying the law of total covariance, we have

$$\lim_{N \rightarrow \infty} \text{Cov} (\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v}) = \lim_{N \rightarrow \infty} \text{Cov} (\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v} | \widehat{\tau}_N = \tau).$$

Note that

$$(\widehat{\epsilon}_{i,t} | \widehat{\tau}_N = \tau) = \begin{cases} \sigma_i(\epsilon_{i,t} - \bar{\epsilon}_{i,\tau}), & t \leq \tau; \\ \sigma_i(\epsilon_{i,t} - \tilde{\epsilon}_{i,\tau}), & t > \tau; \end{cases}$$

where  $\bar{\epsilon}_{i,t} = \frac{1}{t} \sum_{s=1}^t \epsilon_{i,s}$  and  $\tilde{\epsilon}_{i,t} = \frac{1}{T-t} \sum_{s=t+1}^T \epsilon_{i,s}$ . Taking into account the definitions of  $r(t)$  and  $R(t, v)$  together with some simple algebra, we obtain that  $\text{Var} [\widehat{\epsilon}_{i,t} | \widehat{\tau}_N = \tau] = \sigma_i^2 \gamma_{t,t}(\tau)$  and  $\text{Cov} (\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v} | \widehat{\tau}_N = \tau) = \sigma_i^2 \gamma_{t,v}(\tau)$  for  $t < v$ , such that

$$\gamma_{t,t}(\tau) = \begin{cases} r(t) + \frac{t^2}{\tau^2} r(\tau) - \frac{2t}{\tau} [r(t) + R(t, \tau)], & t < \tau; \\ 0, & t = \tau; \\ r(t - \tau) + \frac{(t-\tau)^2}{(T-\tau)^2} r(T - \tau) - \frac{2(t-\tau)}{T-\tau} [r(t - \tau) + R(t - \tau, T - \tau)], & t > \tau; \end{cases}$$

and

$$\gamma_{t,v}(\tau) = \begin{cases} 0, & t = \tau \text{ or } v = \tau, \\ r(t) + R(t, v) + \frac{tv}{\tau^2} r(\tau) - \frac{v}{\tau} [r(t) + R(t, \tau)] - \frac{t}{\tau} [r(v) + R(v, \tau)], & t < v < \tau; \\ S(t, v, \tau + 1 - t) + \frac{t(v-\tau)}{\tau(T-\tau)} R(\tau, T) - \frac{v-\tau}{T-\tau} S(t, T, \tau + 1 - t) - \frac{t}{\tau} R(\tau, v), & t < \tau < v; \\ r(t - \tau) + R(t - \tau, v - \tau) + \frac{(t-\tau)(v-\tau)}{(T-\tau)^2} r(T - \tau) - \frac{v-\tau}{T-\tau} [r(t - \tau) + R(t - \tau, T - \tau)] \\ \quad - \frac{t-\tau}{T-\tau} [r(v - \tau) + R(v - \tau, T - \tau)], & \tau < t < v; \end{cases}$$

where

$$S(t, v, d) = \text{Cov} \left( \sum_{s=1}^t \varepsilon_{i,s}, \sum_{u=t+d}^v \varepsilon_{i,u} \right) = \sum_{s=1}^t \sum_{u=t+d}^v \rho_{u-s}, \quad \forall i \in \mathbb{N}.$$

Thus,  $\lim_{N \rightarrow \infty} \Gamma_{i,N} = \sigma_i^2 \Gamma(\tau)$ , where the matrix  $\Gamma(\tau) = \{\gamma_{t,v}(\tau)\}_{t,v=1}^{T,T}$  is symmetric and does not depend on  $i$ . The matrix  $\Gamma(\tau)$  is singular. Nevertheless, omitting the  $\tau$ -th row and the  $\tau$ -th column from  $\Gamma(\tau)$ , one obtains matrix  $\tilde{\Gamma}(\tau)$ , i.e.,  $\tilde{\Gamma}(\tau) := \Gamma_{-\tau,-\tau}(\tau)$ , which has a full rank of  $T - 1$  due to Assumption 1 and

$$(\hat{\varepsilon}_{i,t} | \hat{\tau}_N = \tau) = \begin{cases} \sigma_i (\sum_{s=1}^t \varepsilon_{i,s} - t\bar{\varepsilon}_{i,\tau}), & t < \tau; \\ 0, & t = \tau; \\ \sigma_i (\sum_{s=\tau+1}^t \varepsilon_{i,s} - (t - \tau)\tilde{\varepsilon}_{i,\tau}), & t > \tau. \end{cases}$$

Let us define random vectors

$$\begin{aligned} \mathbf{U}_N &:= [U_N(1), \dots, U_N(\hat{\tau}_N - 1), U_N(\hat{\tau}_N + 1), \dots, U_N(T)]^\top, \\ \hat{\mathbf{U}}_N &:= [\hat{U}_N(1), \dots, \hat{U}_N(\hat{\tau}_N - 1), \hat{U}_N(\hat{\tau}_N + 1), \dots, \hat{U}_N(T)]^\top, \\ \hat{\mathbf{U}}_N^* &:= [\hat{U}_N^*(1), \dots, \hat{U}_N^*(\hat{\tau}_N - 1), \hat{U}_N^*(\hat{\tau}_N + 1), \dots, \hat{U}_N^*(T)]^\top, \end{aligned}$$

i.e., they do not contain elements with argument  $\hat{\tau}_N$ . The law of total probability provides

$$\begin{aligned} & \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N^* \leq \mathbf{x} | \mathbb{Y} \right] - \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} \right] \\ &= \sum_{\pi=1}^T \left\{ \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N^* \leq \mathbf{x} | \mathbb{Y}, \hat{\tau}_N = \pi \right] - \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} | \hat{\tau}_N = \pi \right] \right\} \mathbb{P}[\hat{\tau}_N = \pi] \end{aligned} \tag{A6}$$

for all  $\mathbf{x} \in \mathbb{R}^{T-1}$ . Since Assumption 5 holds, then according to the bootstrap multivariate CLT by [16] (Theorem 2.4) for (conditionally) independent and not identically distributed zero mean  $(T - 1)$ -dimensional random vectors  $\boldsymbol{\zeta}_N = [[\hat{\varepsilon}_{i,1}, \dots, \hat{\varepsilon}_{i,\hat{\tau}_N-1}, \hat{\varepsilon}_{i,\hat{\tau}_N+1}, \dots, \hat{\varepsilon}_{i,T}]^\top | \hat{\tau}_N = \tau]$ , we have

$$\mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N^* \leq \mathbf{x} | \mathbb{Y}, \hat{\tau}_N = \tau \right] - \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} | \hat{\tau}_N = \tau \right] \xrightarrow[N \rightarrow \infty]{\mathbb{P}} 0 \tag{A7}$$

for all  $\mathbf{x} \in \mathbb{R}^{T-1}$ . Theorem 1, Relations (A6) and (A7) imply

$$\mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N^* \leq \mathbf{x} | \mathbb{Y} \right] - \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} \right] \xrightarrow[N \rightarrow \infty]{\mathbb{P}} 0 \tag{A8}$$

for all  $\mathbf{x} \in \mathbb{R}^{T-1}$ .

Using the law of total probability again, we obtain

$$\mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} \right] = \sum_{\pi=1}^T \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} | \hat{\tau}_N = \pi \right] \mathbb{P}[\hat{\tau}_N = \pi]. \tag{A9}$$

The consistency result  $\lim_{N \rightarrow \infty} \mathbb{P}[\hat{\tau}_N = \tau] = 1$  from Theorem 1 and Equation (A9) give

$$\mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} \right] - \mathbb{P} \left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} | \hat{\tau}_N = \tau \right] \xrightarrow[N \rightarrow \infty]{\mathbb{P}} 0. \tag{A10}$$

Since the Lyapunov CLT provides that  $[\zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{\mathbf{U}}_N \leq \mathbf{x} | \hat{\tau}_N = \tau]$  has an approximate multivariate normal distribution with zero mean and identity covariance matrix, Relation (A10) gives that

the limiting distribution of  $\zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{U}_N$  is the same. Note that the Lyapunov condition for  $\left[ \zeta_N^{-1} \tilde{\Gamma}^{-1/2}(\tau) \hat{U}_N \leq \mathbf{x} \mid \hat{\tau}_N = \tau \right]$  can be checked in a similar manner as in (A5), i.e.,

$$\begin{aligned} & \zeta_N^{-2-\chi} \left( \mathbf{b}^\top \tilde{\Gamma}(\tau) \mathbf{b} \right)^{-\frac{2+\chi}{2}} \\ & \times \sum_{i=1}^N \mathbb{E} \left| \mathbf{b}^\top \sigma_i \left[ \sum_{s=1}^1 (\varepsilon_{i,s} - \bar{\varepsilon}_{i,\tau}), \dots, \sum_{s=1}^{\tau-1} (\varepsilon_{i,s} - \bar{\varepsilon}_{i,\tau}), \sum_{s=\tau+1}^{\tau+1} (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,\tau}), \dots, \sum_{s=\tau+1}^T (\varepsilon_{i,s} - \tilde{\varepsilon}_{i,\tau}) \right] \right|^{2+\chi} \\ & \leq \tilde{q} \zeta_N^{-2-\chi} \sum_{i=1}^N \sigma_i^{2+\chi} \leq \tilde{q} \left( N \bar{\sigma}^2 \right)^{-\frac{2+\chi}{2}} N \bar{\sigma}^{2+\chi} = \tilde{q} \bar{\sigma}^{-2-\chi} \bar{\sigma}^{2+\chi} N^{-\frac{\chi}{2}} \rightarrow 0, \quad N \rightarrow \infty \end{aligned}$$

for arbitrary fixed  $\mathbf{0} \neq \mathbf{b} = [b_1, \dots, b_{T-1}]^\top \in \mathbb{R}^{T-1}$ , some  $0 < \chi \leq 2$  and some positive constant  $\tilde{q}$  not depending on  $N$ .

Bear in mind that

$$\sum_{i=1}^N \sum_{r=1}^s \left( \hat{Y}_{i,r}^* - \tilde{Y}_{i,t}^* \right) = \sum_{i=1}^N \left\{ \left[ \sum_{r=1}^s \hat{Y}_{i,r}^* \right] - \frac{s}{t} \sum_{v=1}^t \hat{Y}_{i,v}^* \right\} = \hat{U}_N^*(s) - \frac{s}{t} \hat{U}_N^*(t)$$

and

$$\sum_{i=1}^N \sum_{r=s+1}^T \left( \hat{Y}_{i,r}^* - \tilde{Y}_{i,t}^* \right) = \hat{V}_N^*(s) - \frac{T-s}{T-t} \hat{V}_N^*(t).$$

Applying the continuous mapping theorem completes the second part of the proof.

(iii) Under  $H_0$ , Theorem 1 provides

$$\lim_{N \rightarrow \infty} \mathbb{P}[\hat{\tau}_N = T] = 1.$$

Then, in view of (4),

$$\lim_{N \rightarrow \infty} \mathbb{P} \left[ \hat{U}_N(s) - \frac{s}{t} \hat{U}_N(t) = U_N(s) - \frac{s}{t} U_N(t) \right] = 1, \quad 1 \leq s \leq t \leq T.$$

Finally, it is sufficient to realize the definition of  $\mathcal{S}_{N,T}$  and  $\mathcal{S}_{N,T}^*$  together with (A8).  $\square$

### References

1. Horváth, L.; Hušková, M. Change-point Detection in Panel Data. *J. Time Ser. Anal.* **2012**, *33*, 631–648.
2. Peštová, B.; Pešta, M. Testing Structural Changes in Panel Data with Small Fixed Panel Size and Bootstrap. *Metrika* **2015**, *78*, 665–689.
3. Bai, J. Common Breaks in Means and Variances for Panel Data. *J. Econom.* **2010**, *157*, 78–92.
4. Peštová, B.; Pešta, M. Erratum to: Testing Structural Changes in Panel Data with Small Fixed Panel Size and Bootstrap. *Metrika* **2016**, *79*, 237–238.
5. Kim, D. Estimating a common deterministic time trend break in large panels with cross sectional dependence. *J. Econom.* **2011**, *164*, 310–330.
6. Kim, D. Common breaks in time trends for large panel data with a factor structure. *Econom. J.* **2014**, *17*, 301–337.
7. Horváth, L.; Hušková, M.; Rice, G.; Wang, J. Estimation of the time of change in panel data. *ArXiv* **2015**, arXiv:math.ST/1503.04455.
8. Baltagi, B.H.; Feng, Q.; Kao, C. Estimation of heterogeneous panels with structural breaks. *J. Econom.* **2016**, *191*, 176–195.
9. Pešta, M.; Hudecová, Š. Asymptotic consistency and inconsistency of the chain ladder. *Insur. Math. Econ.* **2012**, *51*, 472–479.

10. Lindner, A.M. Stationarity, mixing, distributional properties and moments of GARCH(p,q)-processes. In *Handbook of Financial Time Series*; Andersen, T.G., Davis, R.A., Kreiss, J.P., Mikosch, T., Eds.; Springer: Berlin, Germany, 2009; pp. 481–496.
11. Peřtová, B.; Peřta, M. Change Point in Panel Data with Small Fixed Panel Size: Ratio and Non-Ratio Test Statistics. *ArXiv* **2016**, arXiv:math.ME/1608.05670.
12. Huřková, M.; Kirch, C. Bootstrapping sequential change-point tests for linear regression. *Metrika* **2012**, *75*, 673–708.
13. Huřková, M.; Kirch, C.; Prářková, Z.; Steinebach, J. On the detection of changes in autoregressive time series, II. Resampling procedures. *J. Stat. Plan. Inference* **2008**, *138*, 1697–1721.
14. Peřtová, B.; Peřta, M. Change Point Detection in Panel Data with Small Fixed Panel Size. In *Proceedings ITISE 2016*; Valenzuela, O., Rojas, F., Ruiz, G., Pomares, H., Rojas, I., Eds.; University of Granada: Granada, Spain, 2016; pp. 194–205.
15. Meyers, G.G.; Shi, P. Loss Reserving Data Pulled From NAIC Schedule P, 2011. Available online: [http://www.casact.org/research/index.cfm?fa=loss\\_reserves\\_data](http://www.casact.org/research/index.cfm?fa=loss_reserves_data) (accessed on 10 June 2014).
16. Peřta, M. Total least squares and bootstrapping with application in calibration. *Statistics* **2013**, *47*, 966–991.



© 2017 by the authors; 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 (<http://creativecommons.org/licenses/by/4.0/>).