CrossMark

Testing structural changes in panel data with small fixed panel size and bootstrap

Barbora Peštová · Michal Pešta

Abstract Panel data of our interest consist of a moderate or relatively large number of panels, while the panels contain a small number of observations. This paper establishes testing procedures to detect a possible common change in means of the panels. To this end, we consider a ratio type test statistic and derive its asymptotic distribution under the no change null hypothesis. Moreover, we prove the consistency of the test under the alternative. The main advantage of such an approach is that the variance of the observations neither has to be known nor estimated. On the other hand, the correlation structure is required to be calculated. To overcome this issue, a bootstrap technique is proposed in the way of a completely data driven approach without any tuning parameters. The validity of the bootstrap algorithm is shown. As a by-product of the developed tests, we introduce a common break point estimate and prove its consistency. The results are illustrated through a simulation study. An application of the procedure to actuarial data is presented.

B. Peštová · M. Pešta (✉)
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, 18675 Prague 8, Czech Republic
e-mail: Michal.Pesta@mff.cuni.cz

1 Introduction

The problem of an unknown common change in means of the panels is studied here, 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 $\tau = 1, \ldots, N$. The panels are considered to be independent, but this restriction can be weakened. In spite of that, observations within the panel are usually not independent. It is supposed that a common unknown dependence structure is present over the panels.

Tests for change point detection in the panel data have been proposed only in case when the panel size T is sufficiently large, i.e., T increases over all limits from an asymptotic point of view, cf. Chan et al. (2013) or Horváth and Hušková (2012). However, the change point estimation has already been studied for finite T not depending on the number of panels N, see Bai (2010). The remaining task is to develop testing procedures to decide whether a common change point is present or not in the panels, while taking into account that the length T of each observation regime is fixed and can be relatively small.

1.1 Motivation

Structural changes in panel data—especially *common breaks in means*—are wide spread phenomena. Our primary motivation comes from 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 $i = 1, \ldots, N$ provides the total claim amount $Y_{i,t}$ paid in year $t = 1, \ldots, T$ 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 Pešta and Hudecová (2012). This paper shows a way how to test for a possible historical instability.

2 Panel change point model

Let us consider the panel change point model

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

where $\sigma > 0$ is an unknown variance-scaling parameter and T is fixed, not depending on N. The possible *common change point time* is denoted by $\tau \in \{1, \ldots, T\}$. A situation where $\tau = T$ corresponds to *no change* in means of the panels. The means

μ_i are panel-individual. The amount of the break in mean, which can also differ for every panel, is denoted by δ_i. Furthermore, it is assumed that the sequences of panel disturbances $\{\varepsilon_{i,t}\}_t$ are independent and within each panel the errors form a weakly stationary sequence with a common correlation structure. This can be formalized in the following assumption.

Assumption A1 The vectors $[\varepsilon_{i,1}, \ldots, \varepsilon_{i,T}]^\top$ existing on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ are iid for $i = 1, \ldots, N$ with $\mathsf{E}\varepsilon_{i,t} = 0$ and $\mathsf{Var}\,\varepsilon_{i,t} = 1$, having the autocorrelation function

$$\rho_t = \mathsf{Corr}\left(\varepsilon_{i,s}, \varepsilon_{i,s+t}\right) = \mathsf{Cov}\left(\varepsilon_{i,s}, \varepsilon_{i,s+t}\right), \quad \forall s \in \{1, \ldots, T - t\},$$

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

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

and the shifted cumulative correlation function

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

for all $i = 1, \ldots, N$ and $t, v = 1, \ldots, T$.

The sequence $\{\varepsilon_{i,t}\}_{t=1}^{T}$ can be viewed as a part of a *weakly stationary* process. Note that the dependent errors within each panel do not necessarily need to be linear processes. For example, GARCH processes as error sequences are allowed as well. The assumption of independent panels can indeed be relaxed, but it would make the setup much more complex. 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. Moreover, the assumption of a common homoscedastic variance parameter σ can be generalized by introducing weights $w_{i,t}$, which are supposed to be known. Being particular in actuarial practice, it would mean to normalize the total claim amount by the premium received, since bigger insurance companies are expected to have higher variability in total claim amounts paid.

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

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

3 Test statistic and asymptotic results

We propose a *ratio type statistic* to test H_0 against H_1, because this type of statistic does not require estimation of the nuisance parameter for the variance. Generally, this is due to the fact that the variance parameter simply cancels out from the nominator and denominator of the statistic. In spite of that, the common variance could be estimated from all the panels, of which we possess a sufficient number. Nevertheless, we aim to construct a valid and completely data driven testing procedure without interfering estimation and plug-in estimates instead of nuisance parameters. A bootstrap add-on is going to serve this purpose as it is seen later on.

For surveys on ratio type test statistics, we refer to Chen and Tian (2014), Csörgő and Horváth (1997), Horváth et al. (2009), Liu et al. (2008), and Madurkayová (2011). Our particular panel change point test statistic is

$$\mathcal{R}_N(T) = \max_{t=2,\dots,T-2} \frac{\max_{s=1,\dots,t} \left| \sum_{i=1}^N \left[\sum_{r=1}^s \left(Y_{i,r} - \overline{Y}_{i,t} \right) \right] \right|}{\max_{s=t,\dots,T-1} \left| \sum_{i=1}^N \left[\sum_{r=s+1}^T \left(Y_{i,r} - \widetilde{Y}_{i,t} \right) \right] \right|},$$

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

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

An alternative way for testing the change in panel means could be a usage of CUSUM type statistics. For example, a maximum or minimum of a sum (not a ratio) of properly standardized or modified sums from our test statistic $\mathcal{R}_N(T)$. The theory, which follows, can be appropriately rewritten for such cases.

Firstly, we derive the behavior of the test statistics under the null hypothesis.

Theorem 1 (Under null) *Under hypothesis H_0 and Assumption A1*

$$\mathcal{R}_N(T) \xrightarrow[N\to\infty]{\mathscr{D}} \max_{t=2,\dots,T-2} \frac{\max_{s=1,\dots,t} \left| X_s - \frac{s}{t} X_t \right|}{\max_{s=t,\dots,T-1} \left| Z_s - \frac{T-s}{T-t} Z_t \right|},$$

where $Z_t := X_T - X_t$ and $[X_1, \dots, X_T]^\top$ is a multivariate normal random vector with zero mean and covariance matrix $\boldsymbol{\Lambda} = \{\lambda_{t,v}\}_{t,v=1}^{T,T}$ such that

$$\lambda_{t,t} = r(t) \quad \text{and} \quad \lambda_{t,v} = r(t) + R(t,v), \ t < v.$$

The limiting distribution does not depend on the variance nuisance parameter σ, but it depends on the unknown correlation structure of the panel change point model, which has to be estimated for testing purposes. The way of its estimation is shown in Sect. 4.1. Furthermore, Theorem 1 is just a theoretical mid-step for the bootstrap

test, where the correlation structure need not to be known. That is why the presence of unknown quantities in the asymptotic distribution is not troublesome.

Note that in case of independent observations within the panel, the correlation structure and, hence, the covariance matrix Λ is simplified such that $r(t) = t$ and $R(t, v) = 0$.

Next, we show how the test statistic behaves under the alternative.

Assumption A2 $\lim_{N\to\infty} \frac{1}{\sqrt{N}} \left| \sum_{i=1}^{N} \delta_i \right| = \infty$.

Theorem 2 (Under alternative) *If $\tau \leq T - 3$, then under Assumptions A1, A2 and alternative H_1*

$$\mathcal{R}_N(T) \xrightarrow[N\to\infty]{\mathsf{P}} \infty. \tag{2}$$

Assumption A2 is satisfied, for instance, if $0 < \delta \leq \delta_i \ \forall i$ (a common lower change point threshold) and $\delta\sqrt{N} \to \infty$, $N \to \infty$. Another suitable example of δ_is for the condition in Assumption A2, can be $0 < \delta_i = KN^{-1/2+\eta}$ for some $K > 0$ and $\eta > 0$. Or $\delta_i = Ci^{\alpha-1}\sqrt{N}$ may be used as well, where $\alpha \geq 0$ and $C > 0$. The assumption $\tau \leq T - 3$ means that there are at least three observations in the panel after the change point. It is also possible to redefine the test statistic by interchanging the nominator and the denominator of $\mathcal{R}_N(T)$. Afterwards, Theorem 2 for the modified test statistic would require three observations before the change point, i.e., $\tau \geq 3$.

Theorem 2 says that in presence of a structural change in the panel means, the test statistic explodes above all bounds. Hence, the procedure is consistent and the asymptotic distribution from Theorem 1 can be used to construct the test.

4 Change point estimation

Despite the fact that the aim of the paper is to establish testing procedures for detection of a panel mean change, it is necessary to construct a *consistent estimate* for a possible change point. There are two reasons for that: Firstly, the estimation of the covariance matrix Λ from Theorem 1 requires panels as vectors with elements having common mean (i.e., without a jump). Secondly, the bootstrap procedure, introduced later on, requires centered residuals to be resampled.

A consistent estimate of the change point in the panel data is proposed in Bai (2010), but under circumstances that the change occurred for sure. In our situation, we do not know whether a change occurs or not. Therefore, we modify the estimate proposed by Bai (2010) in the following way. If the panel means change somewhere inside $\{2, \ldots, T-1\}$, let the estimate consistently select this change. If there is no change in panel means, the estimate points out the very last time point T with probability going to one. In other words, the value of the change point estimate can be T meaning no change. This is in contrast with Bai (2010), where T is not reachable.

Let us define the estimate of τ:

$$\widehat{\tau}_N := \arg\max_{t=2,\ldots,T} \frac{1}{t} \sum_{i=1}^{N} \sum_{s=1}^{t} (Y_{i,s} - \overline{Y}_{i,t})^2. \tag{3}$$

Now, we show the desired property of consistency for the proposed change point estimate under the following assumptions.

Assumption C1 $L < \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i^2 < \infty$, where $L = -\infty$ if $\tau = T$ and $L = \max_{t=\tau+1,\ldots,T} \frac{\sigma^2 t^2}{\tau(t-\tau)} \left(\frac{r(\tau)}{\tau^2} - \frac{r(t)}{t^2} \right)$ otherwise.

Assumption C2 $\mathsf{E}\varepsilon_{1,t}^4 < \infty$, $t \in \{1, \ldots, T\}$.

Theorem 3 (Change point estimate consistency) *Suppose that $\tau \neq 1$ and the sequence $\{r(t)/t^2\}_{t=2}^{T}$ is decreasing. Then under Assumptions A1, C1, and C2*

$$\lim_{N \to \infty} \mathsf{P}[\widehat{\tau}_N = \tau] = 1.$$

Assumption C1 assures that the values of changes have to be large enough compared to the variability of the random noise in the panels and to the strength of dependencies within the panels as well. On one hand, Assumption C1 implies the usual assumption $\lim_{N \to \infty} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \delta_i^2 = \infty$ in change point analysis, cf. Bai (2010) or Horváth and Hušková (2012). On the other hand, Assumption C1 assures that $\lim_{N \to \infty} \frac{1}{N^2} \sum_{i=1}^{N} \delta_i^2 = 0$, which is not present when the panel size T is considered as unbounded, i.e., $T \to \infty$. Here, this second part is needed to control the asymptotic boundedness of the variability of $\frac{1}{t} \sum_{i=1}^{N} \sum_{s=1}^{t} (Y_{i,s} - \overline{Y}_{i,t})^2$, because a finite T cannot do that.

Similarly as in the previous section, Assumption C1 is satisfied for $0 < \delta \leq \delta_i < \Delta, \forall i$ (a common lower and upper bound for the change amount) and suitable σ and $r(t)$. Assumptions A2 and C1 are generally incomparable. The monotonicity assumption from Theorem 3 in not very restrictive at all. For example in case of independent observations within the panel, this assumption is automatically fulfilled, since $\{1/t\}_{t=2}^{T}$ is decreasing. Moreover, the weaker the dependency within the panel, the faster the decrease of $r(t)/t^2$.

One can check the proof of Theorem 3 and see that Assumption C1 can be replaced by more restrictive assumptions $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \delta_i^2 = \infty$ and $\lim_{N \to \infty} \frac{1}{N^2} \sum_{i=1}^{N} \delta_i^2 = 0$. This first assumption might be considered as too strong, because a common value of $\delta = \delta_i$ for all i does not fulfill it.

Various competing consistent estimates of a possible change point can be suggested, e.g., the maximizer of $\sum_{i=1}^{N} \left[\sum_{s=1}^{t} (Y_{i,s} - \overline{Y}_{i,T}) \right]^2$. To show the consistency, one needs to postulate different assumptions on the cumulative autocorrelation function and shifted cumulative correlation function compared to Theorem 3 and this may be rather complex.

4.1 Estimation of the 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 $[\varepsilon_{1,1}, \ldots, \varepsilon_{1,T}]^{\top}$ empirically. We base the errors' estimates on *residuals*

$$\widehat{e}_{i,t} := \begin{cases} Y_{i,t} - \overline{Y}_{i,\widehat{\tau}_N}, & t \le \widehat{\tau}_N, \\ Y_{i,t} - \widetilde{Y}_{i,\widehat{\tau}_N}, & t > \widehat{\tau}_N. \end{cases} \tag{4}$$

Then, the empirical version of the autocorrelation function is

$$\widehat{\rho}_t := \frac{1}{\widehat{\sigma}^2 NT} \sum_{i=1}^{N} \sum_{s=1}^{T-t} \widehat{e}_{i,s} \widehat{e}_{i,s+t}.$$

Consequently, the kernel estimation of the cumulative autocorrelation function and shifted cumulative correlation function is adopted in lines with Andrews (1991):

$$\widehat{r}(t) = \sum_{|s|<t} (t - |s|) \kappa \left(\frac{s}{h}\right) \widehat{\rho}_s,$$

$$\widehat{R}(t, v) = \sum_{s=1}^{t} \sum_{u=t+1}^{v} \kappa \left(\frac{u-s}{h}\right) \widehat{\rho}_{u-s}, \quad t < v;$$

where $h > 0$ stands for the window size and κ belongs to a class of kernels given by

$$\left\{ \kappa(\cdot) : \mathbb{R} \to [-1, 1] \,\middle|\, \kappa(0) = 1, \ \kappa(x) = \kappa(-x), \ \forall x, \ \int_{-\infty}^{+\infty} \kappa^2(x) \mathrm{d}x < \infty, \right.$$

$$\left. \kappa(\cdot) \text{ is continuos at 0 and at all but a finite number of other points} \right\}.$$

Since the variance parameter σ is not present in the limiting distribution of Theorem 1, it neither has to be estimated nor known. Nevertheless, one can use $\widehat{\sigma}^2 := \frac{1}{NT} \sum_{i=1}^{N} \sum_{s=1}^{T} \widehat{e}_{i,s}^2$.

5 Bootstrap and hypothesis testing

A wide range of literature has been published on bootstrapping in the change point problem, e.g., Hušková and Kirch (2012) or Hušková et al. (2008). We build up the bootstrap test on the resampling with replacement of row vectors $\{[\widehat{e}_{i,1}, \ldots, \widehat{e}_{i,T}]\}_{i=1,\ldots,N}$ corresponding to the panels. This provides bootstrapped row vectors $\{[\widehat{e}_{i,1}^*, \ldots, \widehat{e}_{i,T}^*]\}_{i=1,\ldots,N}$. Then, the bootstrapped residuals $\widehat{e}_{i,t}^*$ are *centered* by their conditional expectation $\frac{1}{N} \sum_{i=1}^{N} \widehat{e}_{i,t}$ yielding

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

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

$$\mathcal{R}_N^*(T) = \max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t} \left| \sum_{i=1}^{N} \left[\sum_{r=1}^{s} \left(\widehat{Y}_{i,r}^* - \overline{\widehat{Y}}_{i,t}^* \right) \right] \right|}{\max_{s=t,\ldots,T-1} \left| \sum_{i=1}^{N} \left[\sum_{r=s+1}^{T} \left(\widehat{Y}_{i,r}^* - \widetilde{\widehat{Y}}_{i,t}^* \right) \right] \right|},$$

such that

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

An *algorithm* for the bootstrap is illustratively shown in Procedure 1 and its validity will be proved in Theorem 4.

Procedure 1 Bootstrapping test statistic $\mathcal{R}_N(T)$.

Input: Panel data consisting of N panels with length T, i.e., N row vectors of observations $[Y_{i,1}, \ldots, Y_{i,T}]$.
Output: Bootstrap distribution of $\mathcal{R}_N(T)$, i.e., the empirical distribution where probability mass $1/B$
    concentrates at each of $_{(1)}\mathcal{R}_N^*(T), \ldots, {}_{(B)}\mathcal{R}_N^*(T)$.
1: estimate the change point by calculating $\widehat{\tau}_N$
2: compute residuals $\widehat{e}_{i,t}$
3: **for** $b = 1$ to B **do** // repeat in order to obtain the empirical distribution
4:    $\{[\widehat{e}_{i,1}^*, \ldots, \widehat{e}_{i,T}^*]\}_{i=1}^{N}$ resampled with replacement from original rows $\{[\widehat{e}_{i,1}, \ldots, \widehat{e}_{i,T}]\}_{i=1}^{N}$
5:    calculate bootstrap panel data $\widehat{Y}_{i,t}^*$
6:    compute bootstrap test statistics $_{(b)}\mathcal{R}_N^*(T)$
7: **end for**

5.1 Validity of the resampling procedure

The idea behind bootstrapping is to *mimic the original distribution* of the test statistic in some sense with the distribution of the bootstrap test statistic, conditionally on the original data denoted by $\mathbb{Y} \equiv \{Y_{i,t}\}_{i,t=1}^{N,T}$.

First of all, two simple and just technical assumptions are needed.

Assumption B1 $\{\varepsilon_{i,t}\}_t$ possesses the lagged cumulative correlation function

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

Assumption B2 $\lim_{N\to\infty} \mathsf{P}[\widehat{\tau}_N = \tau] = 1$.

Assumption B1 is not really an assumption, actually it is only a notation. Notice that $S(t, v, 1) \equiv R(t, v)$. Assumption B2 is satisfied for our estimate proposed in (3), if the assumptions of Theorem 2 hold. Assumption B2 is postulated in a rather broader sense, because we want to allow any other consistent estimate of τ to be used instead.

Realize that it is not known, whether the common panel means' change occurred or not. In other words, one does not know *whether the data come from the null or the alternative* hypothesis. Therefore, the following theorem holds under H_0 as well as H_1.

Theorem 4 (Bootstrap justification) *Under Assumptions A1, B1, B2, and C2*

$$\mathcal{R}_N^*(T)|\mathbb{Y} \xrightarrow[N\to\infty]{\mathscr{D}} \max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t}\left|\mathcal{X}_s - \frac{s}{t}\mathcal{X}_t\right|}{\max_{s=t,\ldots,T-1}\left|\mathcal{Z}_s - \frac{T-s}{T-t}\mathcal{Z}_t\right|} \quad \text{in probability } \mathsf{P},$$

where $\mathcal{Z}_t := \mathcal{X}_T - \mathcal{X}_t$ and $[\mathcal{X}_1, \ldots, \mathcal{X}_T]^\top$ is a multivariate normal random vector with zero mean and covariance matrix $\boldsymbol{\Gamma} = \{\gamma_{t,v}(\tau)\}_{t,v=1}^{T,T}$ such that

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

and

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

The validity of the bootstrap test is assured by Theorem 4. Indeed, the conditional asymptotic distribution of the bootstrap test statistic is a functional of a multivariate normal distribution under the null as well as under the alternative. It does not converge to infinity (in probability) under the alternative. That is why it can be used for correctly rejecting the null in favor of the alternative, having sufficiently large N. Moreover, the following theorem states that the conditional distribution of the bootstrap test statistic and the unconditional distribution of the original test statistic *coincide*. And that is the reason why the bootstrap test should approximately keep the same level as the original test based on the asymptotics from Theorem 1.

Theorem 5 (Bootstrap test consistency) *Under Assumptions A1, B2, C2 and hypothesis H_0, the asymptotic distribution of $\mathcal{R}_N(T)$ from Theorem 1 and the asymptotic distribution of $\mathcal{R}_N^*(T)|\mathbb{Y}$ from Theorem 4 coincide.*

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.

Finally, note that one cannot think about any local alternative in this setup, because τ has a discrete and finite support.

6 Simulations

A simulation experiment was performed to study the *finite sample* properties of the asymptotic and bootstrap test statistics for a common change in panel means. In particular, the interest lies in the empirical *sizes* of the proposed tests under the null hypothesis and in the empirical *rejection* rate (power) under the alternative. Random samples of panel data (5000 each time) are generated from the panel change point model (1). The panel size is set to $T = 10$ and $T = 25$ in order to demonstrate the performance of the testing approaches in case of small and intermediate panel length. The number of panels considered is $N = 50$ and $N = 200$.

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 case of GARCH(1,1) process, we use coefficients $\alpha_0 = 1$, $\alpha_1 = 0.1$, and $\beta_1 = 0.2$, which according to (Lindner 2009, [Example 1]) gives 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. Simulation scenarios are produced as all possible combinations of the above mentioned settings.

When using the asymptotic distribution from Theorem 1, the covariance matrix is estimated as proposed in Sect. 4.1 using the Parzen kernel

$$\kappa_P(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & 0 \le |x| \le 1/2; \\ 2(1 - |x|)^3, & 1/2 \le |x| \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

Several values of the smoothing window width h are tried from the interval $[2, 5]$ and all of them work fine providing comparable results. To simulate the asymptotic distribution of the test statistics, 2000 multivariate random vectors are generated using the pre-estimated covariance matrix.

The bootstrap approach does not need to estimate the covariance structure. The number of bootstrap replications used is 2000. To access the theoretical results under H_0 numerically, Table 1 provides the empirical specificity (one minus size) of the tests

Table 1 Empirical specificity (1−size) of the test under H_0 using the asymptotic and the **bootstrap** critical values, considering a significance level of 5 %

T	N	Innovations	IID		AR(1)		GARCH(1,1)	
10	50	$N(0, 1)$	.948	**.949**	.943	**.955**	.949	**.955**
		t_5	.949	**.954**	.941	**.956**	.946	**.953**
	200	$N(0, 1)$	.952	**.951**	.937	**.954**	.942	**.952**
		t_5	.948	**.953**	.935	**.960**	.944	**.953**
25	50	$N(0, 1)$	.948	**.951**	.929	**.952**	.954	**.959**
		t_5	.946	**.951**	.932	**.954**	.944	**.958**
	200	$N(0, 1)$	.950	**.950**	.927	**.951**	.947	**.949**
		t_5	.948	**.953**	.931	**.952**	.952	**.952**

Table 2 Empirical sensitivity (power) of the test under H_1 using the asymptotic and the **bootstrap** critical values, considering a significance level of 5 %

H_1	T	N	Innovations	IID		AR(1)		GARCH(1,1)	
33 %	10	50	N(0, 1)	.25	**.29**	.26	**.28**	.21	**.22**
			t_5	.18	**.18**	.19	**.21**	.20	**.24**
		200	N(0, 1)	.46	**.50**	.48	**.51**	.40	**.45**
			t_5	.37	**.38**	.39	**.41**	.40	**.47**
	25	50	N(0, 1)	.38	**.43**	.31	**.37**	.30	**.36**
			t_5	.34	**.32**	.27	**.30**	.32	**.37**
		200	N(0, 1)	.72	**.79**	.69	**.66**	.60	**.67**
			t_5	.53	**.54**	.51	**.52**	.58	**.59**
66 %	10	50	N(0, 1)	.45	**.52**	.47	**.50**	.39	**.41**
			t_5	.36	**.35**	.38	**.39**	.39	**.42**
		200	N(0, 1)	.76	**.84**	.81	**.81**	.70	**.77**
			t_5	.65	**.66**	.68	**.64**	.69	**.76**
	25	50	N(0, 1)	.71	**.79**	.63	**.66**	.60	**.7**
			t_5	.60	**.63**	.44	**.45**	.60	**.72**
		200	N(0, 1)	.96	**.98**	.95	**.95**	.90	**.94**
			t_5	.84	**.82**	.86	**.85**	.90	**.92**
100 %	10	50	N(0, 1)	.64	**.69**	.67	**.71**	.56	**.64**
			t_5	.52	**.48**	.49	**.51**	.55	**.62**
		200	N(0, 1)	.93	**.96**	.94	**.95**	.86	**.92**
			t_5	.84	**.83**	.86	**.83**	.87	**.92**
	25	50	N(0, 1)	.86	**.91**	.83	**.86**	.80	**.85**
			t_5	.76	**.79**	.66	**.67**	.79	**.85**
		200	N(0, 1)	1.00	**1.00**	.99	**.99**	.98	**.97**
			t_5	.98	**.97**	.98	**.97**	.99	**.99**

for both the asymptotic and bootstrap version of the panel change point test, where the significance level is $\alpha = 5$ %.

It may be seen that both approaches (using asymptotic and bootstrap distribution) are close to the theoretical value of specificity .95. As expected, the best results are achieved in case of independence within the panel, because there is no information overlap between two consecutive observations. The precision of not rejecting the null is increasing as the number of panels is getting higher and the panel is getting longer as well.

The performance of both testing procedures under H_1 in terms of the empirical rejection rates is shown in Table 2, where the change point is set to $\tau = \lfloor T/2 \rfloor$ and the change sizes δ_i are independently uniform on [1, 3] in 33 %, 66 % or in all panels.

One can conclude that the power of both tests increases as the panel size and the number of panels increase, which is straightforward and expected. It should be noticed that numerical instability issues may appear for larger T, when generating from a T-variate normal distribution. Moreover, higher power is obtained when a larger portion

Table 3 Empirical sensitivity of the test for small values of τ under H_1 using the asymptotic and the **bootstrap** critical values, considering a significance level of 5 %

T	N	τ	H_1, iid, $N(0, 1)$	
10	50	3	.59	**.63**
	200	3	.89	**.91**
25	50	5	.66	**.68**
	200	5	.94	**.96**

of panels is subject to have a change in mean. The test power drops when switching from independent observations within the panel to dependent ones. Innovations with heavier tails (i.e., t_5) yield smaller power than innovations with lighter tails. Generally, the bootstrap outperforms the classical asymptotics in all scenarios.

Let us mention that for finite sections of processes with a stronger dependence structure than taken into account in the simulation scenarios, Assumption C1 does not have to be fulfilled. For example, Assumption C1 is violated for AR(1) with coefficient $\phi = 0.9$, $\delta_i = 2$, $\sigma = 1$, standard normal or Student t_5 innovations, and $\tau = 5$ for $T = 10$ or $\tau = 12$ for $T = 25$. Here, the dependency under the considered variability is too strong compared to the change size. It is rather difficult to detect possible changes in such a setup.

Finally, an early change point is discussed very briefly. We stay with standard normal innovations, iid observations within the panel, the size of changes δ_i being independently uniform on [1, 3] in all panels, and the change point is $\tau = 3$ in case of $T = 10$ and $\tau = 5$ for $T = 25$. The empirical sensitivities of both tests for small values of τ are shown in Table 3.

When the change point is not in the middle of the panel, the power of the test generally falls down. The source of such decrease is that the left or right part of the panel possesses less observations with constant mean, which leads to a decrease of precision in the correlation estimation in case of the asymptotic test and in the change point estimation in case of the bootstrap test. Nevertheless, the bootstrap test again outperforms the asymptotic version and, moreover, provides solid results even for early or late change points (the late change points are not numerically demonstrated here).

7 Real data analysis

As mentioned in the introduction, our primary motivation for testing the panel mean change comes from the *insurance business*. The data set is provided by the National Association of Insurance Commissioners (NAIC) database, see Meyers and Shi (2011). We concentrate on the 'Commercial auto/truck liability/medical' insurance line of business. The data collect records from $N = 157$ insurance companies (one extreme insurance company was omitted from the analysis). Each insurance company provides $T = 10$ yearly total claim amounts starting from year 1988 up to year 1997. Figure 1 graphically shows series of claim amounts for 20 selected insurance companies (a plot with all 157 panels would be cluttered).

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. The
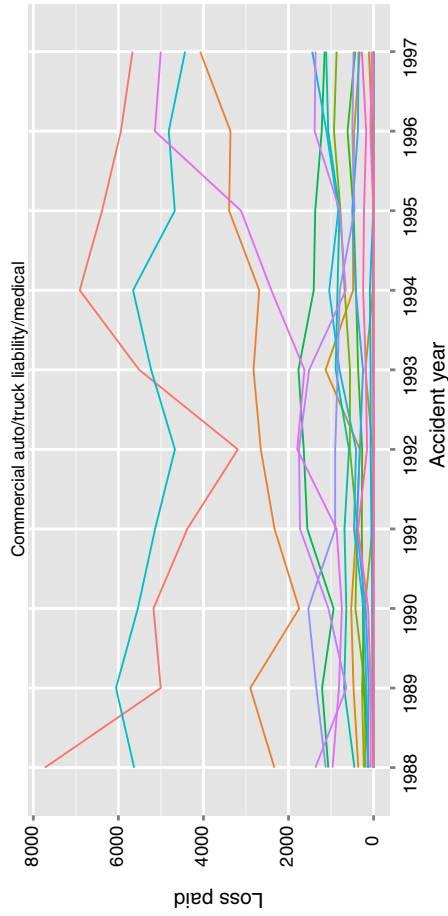
Fig. 1 Development of yearly total claim amounts for 20 selected insurance companies

length of the panel is quite short. This is very typical in insurance business, because considering longer panels may invoke incomparability between the early claim amounts and the late ones due to changing market or policies' conditions over time.

We want to test whether or not a change in the claim amounts occurred in a common year, assuming that the claim amounts are approximately constant in the years before and after the possible change for every insurance company. Our ratio type test statistic gives $\mathcal{R}_{157}(10) = 39.9$. The asymptotic critical value is 52.4 and the bootstrap critical value equals 203.1. These values mean that we do not reject the hypothesis of no change in panel means in both cases. The striking difference between the two critical values may come from the inefficient correlation structure estimation (since $T = 10$ is quite short) or from violation of the model assumptions.

That is why we also try to take the logarithms of claim amounts and to consider log amounts as the panel data observations. Nevertheless, we do not reject the hypothesis of no change in the panel means (i.e., means of log amounts) again. Additionally to that, 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 the assumption of a common variance. In spite of that, we again do not reject the null (neither by the asymptotic test, nor by the bootstrap one). For the sake of completeness, we may reveal that our estimate of the panel change point provides value $\widehat{\tau}_N = 10$ meaning no change in panels.

8 Conclusions

In this paper, we consider the change point problem in panel data with fixed panel size. Occurrence of common breaks in panel means is tested. We introduce a ratio type test statistic and derive its asymptotic properties. Under the null hypothesis of no change, the test statistic weakly converges to a functional of the multivariate normal random vector with zero mean and covariance structure depending on the intra-panel covariances. As shown in the paper, these covariances can be estimated and, consequently, used for testing whether a change in means occurred or not. This is indeed feasible, because the test statistic under the alternative converges to infinity in probability.

The secondary aim of the paper lies in proposing a consistent change point estimate, which is straightforwardly used for bootstrapping the test statistic. We establish the asymptotic behavior of the bootstrap version of the test statistic, regardless of the fact whether the data come from the null or the alternative hypothesis. Moreover, the asymptotic distribution of the bootstrap test statistic coincides with the original test statistic's limiting distribution. This provides justification for the bootstrap method. One of the main goals is to obtain a completely data driven testing approach whether the means remain the same during the observation period or not. The ratio type test statistic allows us to omit a variance estimation and the bootstrap technique overcomes estimation of the correlation structure. Hence, neither nuisance nor smoothing parameters are present in the whole testing process, which makes it very simple for practical use. Furthermore, the whole stochastic theory behind requires relatively simple assumptions, which are not too restrictive.

A simulation study illustrates that even for small panel size, both presented approaches—based on traditional asymptotics and on bootstrapping—work fine. One may judge that both methods keep the significance level under the null, while various simulation scenarios are considered. Besides that, the power of the test is slightly higher in case of the bootstrap. Finally, the proposed methods are applied to insurance data, for which the change point analysis in panel data provides an appealing approach.

8.1 Discussion

First of all, it has to be noted that the non-ratio CUSUM type test statistic can be used instead of the ratio type, but this requires to estimate the variance of the observations. The statements of theorems and proofs would become even less complicated. Omitting the usage of the bootstrap test statistic can especially be unreliable in short panels from a computational point of view. This is due to the fact that the bootstrap overcomes the issue of estimating the correlation structure.

Furthermore, our setup can be modified by considering large panel size, i.e., $T \to \infty$. Consequently, the whole theory leads to convergences to functionals of Gaussian processes with a covariance structure derived in a very similar fashion as for fixed T. However, our motivation is to develop techniques for fixed and relatively small panel size.

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.

Appendix 1: Supporting theorems

Suppose that $\{\boldsymbol{\xi}_n\}_{n=1}^{\infty}$ is a sequence of random variables/vectors existing on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. A *bootstrap version* of $\boldsymbol{\xi} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]^\top$ is its (randomly) resampled sequence with replacement—denoted by $\boldsymbol{\xi}^* \equiv [\boldsymbol{\xi}_1^*, \ldots, \boldsymbol{\xi}_n^*]^\top$—with the same length, where for each $i \in \{1, \ldots, n\}$ it holds that $\mathsf{P}_{\boldsymbol{\xi}}^*[\boldsymbol{\xi}_i^* = \boldsymbol{\xi}_j] \equiv \mathsf{P}[\boldsymbol{\xi}_i^* = \boldsymbol{\xi}_j|\boldsymbol{\xi}] = 1/n, \; j = 1, \ldots, n$. In the sequel, $\mathsf{P}_{\boldsymbol{\xi}}^*$ denotes the conditional probability given $\boldsymbol{\xi}$. So, $\boldsymbol{\xi}_i^*$ has a discrete uniform distribution on $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}$ for every $i = 1, \ldots, n$. The conditional expectation and variance given $\boldsymbol{\xi}$ are denoted by $\mathsf{E}_{\mathsf{P}_{\boldsymbol{\xi}}^*}$ and $\mathsf{Var}_{\mathsf{P}_{\boldsymbol{\xi}}^*}$.

If a statistic has an approximate normal distribution, one may be interested in the asymptotic comparison of the bootstrap distribution with the original one. A tool for assessing such an approximate closeness can be a *bootstrap central limit theorem* for triangular arrays.

Theorem 6 (Bootstrap CLT for triangular arrays) *Let $\{\xi_{n,k_n}\}_{n=1}^{\infty}$ be a triangular array of zero mean random variables on the same probability space such that the elements of the vector $[\xi_{n,1}, \ldots, \xi_{n,k_n}]^{\top}$ are iid for every $n \in \mathbb{N}$ satisfying*

$$\sup_{n \in \mathbb{N}} \mathsf{E}_{\mathsf{P}} \xi_{n,1}^4 < \infty \tag{5}$$

and $k_n \to \infty$ as $n \to \infty$. Suppose that $\boldsymbol{\xi}^* \equiv [\xi_{n,1}^*, \ldots, \xi_{n,k_n}^*]^{\top}$ is the bootstrapped version of $\boldsymbol{\xi} \equiv [\xi_{n,1}, \ldots, \xi_{n,k_n}]^{\top}$ and denote

$$\bar{\xi}_n := k_n^{-1} \sum_{i=1}^{k_n} \xi_{n,i}, \quad \bar{\xi}_n^* := k_n^{-1} \sum_{i=1}^{k_n} \xi_{n,i}^*, \quad and \quad \varsigma_n^2 := \mathsf{Var}_{\mathsf{P}} \xi_{n,1}.$$

If

$$\liminf_{n \to \infty} \varsigma_n^2 = \varsigma^2 > 0, \tag{6}$$

then

$$\sup_{x \in \mathbb{R}} \left| \mathsf{P}_{\boldsymbol{\xi}}^* \left[\frac{\sqrt{k_n}}{\sqrt{\varsigma_n^2}} \left(\bar{\xi}_n^* - \bar{\xi}_n \right) \leq x \right] - \mathsf{P} \left[\frac{\sqrt{k_n}}{\sqrt{\varsigma_n^2}} \bar{\xi}_n \leq x \right] \right| \xrightarrow[n \to \infty]{\mathsf{P}} 0.$$

Theorem 7 (Bootstrap multivariate CLT for triangular arrays) *Let $\{\boldsymbol{\xi}_{n,k_n}\}_{n=1}^{\infty}$ be a triangular array of zero mean q-dimensional random vectors on the same probability space such that the elements of the vector sequence $\{\boldsymbol{\xi}_{n,1}, \ldots, \boldsymbol{\xi}_{n,k_n}\}$ are iid for every $n \in \mathbb{N}$ satisfying*

$$\sup_{n \in \mathbb{N}} \mathsf{E}_{\mathsf{P}} |\xi_{n,1}^{(j)}|^4 < \infty, \quad j \in \{1, \ldots, q\}, \tag{7}$$

where $\boldsymbol{\xi}_{n,1} \equiv [\xi_{n,1}^{(1)}, \ldots, \xi_{n,1}^{(q)}]^{\top} \in \mathbb{R}^q$, $n \in \mathbb{N}$ and $k_n \to \infty$ as $n \to \infty$. Assume that $\boldsymbol{\Xi}^* \equiv [\boldsymbol{\xi}_{n,1}^*, \ldots, \boldsymbol{\xi}_{n,k_n}^*]^{\top}$ is the bootstrapped version of $\boldsymbol{\Xi} \equiv [\boldsymbol{\xi}_{n,1}, \ldots, \boldsymbol{\xi}_{n,k_n}]^{\top}$. Denote

$$\bar{\boldsymbol{\xi}}_n := k_n^{-1} \sum_{i=1}^{k_n} \boldsymbol{\xi}_{n,i}, \quad \bar{\boldsymbol{\xi}}_n^* := k_n^{-1} \sum_{i=1}^{k_n} \boldsymbol{\xi}_{n,i}^*, \quad and \quad \boldsymbol{\Gamma}_n := \mathsf{Var}_{\mathsf{P}} \boldsymbol{\xi}_{n,1}.$$

If

$$\liminf_{n \to \infty} \boldsymbol{\Gamma}_n = \boldsymbol{\Gamma} > \mathbf{0}, \tag{8}$$

then

$$\mathsf{P}_{\boldsymbol{\Xi}}^* \left[\sqrt{k_n} \boldsymbol{\Gamma}_n^{-1/2} \left(\bar{\boldsymbol{\xi}}_n^* - \bar{\boldsymbol{\xi}}_n \right) \leq \mathbf{x} \right] - \mathsf{P} \left[\sqrt{k_n} \boldsymbol{\Gamma}_n^{-1/2} \bar{\boldsymbol{\xi}}_n \leq \mathbf{x} \right] \xrightarrow[n \to \infty]{\mathsf{P}} 0, \quad \forall \mathbf{x} \in \mathbb{R}^q.$$

Appendix 2: Proofs

Proof (of Theorem 1) Let us define

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

Using the multivariate Lindeberg-Lévy CLT for a sequence of T-dimensional iid random vectors $\{[\sum_{s=1}^{1} \varepsilon_{i,s}, \ldots, \sum_{s=1}^{T} \varepsilon_{i,s}]^{\top}\}_{i \in \mathbb{N}}$, we have under H_0

$$[U_N(1), \ldots, U_N(T)]^{\top} \xrightarrow[N\to\infty]{\mathscr{D}} [X_1, \ldots, X_T]^{\top},$$

since $\mathsf{Var}\,[\sum_{s=1}^{1} \varepsilon_{1,s}, \ldots, \sum_{s=1}^{T} \varepsilon_{1,s}]^{\top} = \Lambda$. Indeed, the t-th diagonal element of the covariance matrix Λ is

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

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

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

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

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

Hence,

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

and, consequently,

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

$$= \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} \sum_{r=s+1}^{T} (Y_{i,r} - \widetilde{Y}_{i,t}).$$

Using the Cramér–Wold device, we end up with

$$\max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t} \left| U_N(s) - \frac{s}{t} U_N(t) \right|}{\max_{s=t,\ldots,T-1} \left| V_N(s) - \frac{T-s}{T-t} V_N(t) \right|}$$

$$\xrightarrow[N\to\infty]{\mathcal{D}} \max_{t=2,\ldots,T-2} \frac{\max_{s=1,\ldots,t} \left| X_s - \frac{s}{t} X_t \right|}{\max_{s=t,\ldots,T-1} \left| (X_T - X_s) - \frac{T-s}{T-t} (X_T - X_t) \right|}.$$

$\square$

Proof (of Theorem 2) Let $t = \tau + 1$. Then, under alternative H_1

$$\frac{1}{\sigma\sqrt{N}} \max_{s=1,\ldots,\tau+1} \left| \sum_{i=1}^{N} \left[\sum_{r=1}^{s} (Y_{i,r} - \overline{Y}_{i,\tau+1}) \right] \right|$$

$$\geq \frac{1}{\sigma\sqrt{N}} \left| \sum_{i=1}^{N} \sum_{r=1}^{\tau} (Y_{i,r} - \overline{Y}_{i,\tau+1}) \right|$$

$$= \frac{1}{\sigma\sqrt{N}} \left| \sum_{i=1}^{N} \sum_{r=1}^{\tau} \left(\mu_i + \sigma\varepsilon_{i,r} - \frac{1}{\tau+1} \sum_{v=1}^{\tau+1} (\mu_i + \sigma\varepsilon_{i,v}) - \frac{1}{\tau+1} \delta_i \right) \right|$$

$$= \frac{1}{\sqrt{N}} \left| \sum_{i=1}^{N} \sum_{r=1}^{\tau} (\varepsilon_{i,r} - \overline{\varepsilon}_{i,\tau+1}) - \frac{\tau}{\sigma(\tau+1)} \sum_{i=1}^{N} \delta_i \right|$$

$$= \mathcal{O}_{\mathsf{P}}(1) + \frac{\tau}{\sigma(\tau+1)\sqrt{N}} \left| \sum_{i=1}^{N} \delta_i \right| \xrightarrow{\mathsf{P}} \infty, \quad N \to \infty,$$

where $\overline{\varepsilon}_{i,\tau+1} = \frac{1}{\tau} \sum_{v=1}^{\tau+1} \varepsilon_{i,v}$.

Since there is no change after $\tau + 1$ and $\tau \leq T - 3$, then by Theorem 1 we have

$$\frac{1}{\sigma\sqrt{N}} \max_{s=\tau+1,\ldots,T-1} \left| \sum_{i=1}^{N} \sum_{r=s+1}^{T} (Y_{i,r} - \widetilde{Y}_{i,\tau+1}) \right| \xrightarrow[N\to\infty]{\mathcal{D}} \max_{s=\tau+1,\ldots,T-1} \left| Z_s - \frac{T-s}{T-\tau} Z_{\tau+1} \right|.$$

$\square$

Proof (of Theorem 3) Let us define $S_N^{(i)}(t) := \frac{1}{t}\sum_{s=1}^{t}(Y_{i,s} - \overline{Y}_{i,t})^2$ and, consequently, $S_N(t) := \frac{1}{N}\sum_{i=1}^{N} S_N^{(i)}(t)$. Then,

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

where $\overline{\varepsilon}_{i,t} = \frac{1}{t}\sum_{s=1}^{t}\varepsilon_{i,s}$. By the definition of the cumulative autocorrelation function, we have for $2 \leq t \leq \tau$

$$
\mathsf{E}S_N^{(i)}(t) = \frac{\sigma^2}{t}\sum_{s=1}^{t}\mathsf{E}(\varepsilon_{i,s} - \overline{\varepsilon}_{i,t})^2 = \frac{\sigma^2}{t}\sum_{s=1}^{t}\left[1 - \frac{2}{t}\sum_{r=1}^{t}\mathsf{E}\varepsilon_{i,s}\varepsilon_{i,r} + \frac{1}{t^2}r(t)\right]
$$
$$
= \sigma^2\left(1 - \frac{r(t)}{t^2}\right).
$$

In the other case when $t > \tau$, one can calculate

$$
\mathsf{E}S_N^{(i)}(t) = \sigma^2\left(1 - \frac{r(t)}{t^2}\right) + \frac{\tau}{t}\left(\frac{t-\tau}{t}\right)^2\delta_i^2 + \frac{t-\tau}{t}\left(\frac{\tau}{t}\right)^2\delta_i^2
$$
$$
= \sigma^2\left(1 - \frac{r(t)}{t^2}\right) + \frac{\tau(t-\tau)}{t^2}\delta_i^2.
$$

Realize that $S_N^{(i)}(t) - \mathsf{E}S_N^{(i)}(t)$ are independent with zero mean for fixed t and $i = 1, \ldots, N$. Due to Assumption C2, for $2 \leq t \leq \tau$ it holds

$$
\mathsf{Var}\, S_N(t) = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\sigma^4}{t^2}\mathsf{Var}\left[\sum_{s=1}^{t}(\varepsilon_{i,s} - \overline{\varepsilon}_{i,t})^2\right] = \frac{1}{N}C_1(t, \sigma),
$$

where $C_1(t, \sigma) > 0$ is some constant not depending on N. If $t > \tau$, then

$$
\mathsf{Var}\, S_N(t) = \frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{t^2}\mathsf{Var}\left[\sigma^2\sum_{s=1}^{\tau}(\varepsilon_{i,s} - \overline{\varepsilon}_{i,t})^2 \right.
$$
$$
- 2\frac{t-\tau}{t}\sigma\delta_i\sum_{s=1}^{\tau}(\varepsilon_{i,s} - \overline{\varepsilon}_{i,t}) + \left(\frac{t-\tau}{t}\right)^2\delta_i^2
$$
$$
\left. + \sigma^2\sum_{s=\tau+1}^{t}(\varepsilon_{i,s} - \overline{\varepsilon}_{i,t})^2 + 2\frac{\tau}{t}\sigma\delta_i\sum_{s=\tau+1}^{t}(\varepsilon_{i,s} - \overline{\varepsilon}_{i,t}) + \left(\frac{\tau}{t}\right)^2\delta_i^2\right]
$$
$$
\leq \frac{1}{N}C_2(t, \tau, \sigma) + \frac{1}{N^2}C_3(t, \tau, \sigma)\sum_{i=1}^{N}\delta_i^2 + \frac{1}{N^2}C_4(t, \tau, \sigma)\left|\sum_{i=1}^{N}\delta_i\right|,
$$

where $C_j(t, \tau, \sigma) > 0$ does not depend on N for $j = 2, 3, 4$.

The Chebyshev inequality provides $S_N(t) - \mathsf{E}S_N(t) = \mathcal{O}_\mathsf{P}\left(\sqrt{\mathsf{Var}\, S_N(t)}\right)$ as $N \to \infty$. According to Assumption C1 and the Cauchy-Schwarz inequality, we have

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

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

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

where $K_j(\sigma) > 0$ are constants not depending on N for $j = 1, 2, 3, 4$. Thus, we also have uniform stochastic boundedness, i.e.,

$$\max_{1 \leq t \leq T}|S_N(t) - \mathsf{E}S_N(t)| = \mathcal{O}_\mathsf{P}\left(\frac{1}{\sqrt{N}}\right), \quad N \to \infty.$$

Adding and subtracting, one has

$$\begin{aligned}
S_N(\tau) - S_N(t) &= S_N(\tau) - \mathsf{E}S_N(\tau) - [S_N(t) - \mathsf{E}S_N(t)] + \mathsf{E}S_N(\tau) - \mathsf{E}S_N(t) \\
&\geq -2\max_{1 \leq r \leq T}|S_N(r) - \mathsf{E}S_N(r)| + \mathsf{E}S_N(\tau) - \mathsf{E}S_N(t) \\
&= -2\max_{1 \leq r \leq T}|S_N(r) - \mathsf{E}S_N(r)| + \sigma^2\left(\frac{r(t)}{t^2} - \frac{r(\tau)}{\tau^2}\right) \\
&\quad + \mathcal{I}\{t > \tau\}\frac{\tau(t-\tau)}{t^2}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2.
\end{aligned}$$

The above inequality holds for each $t \in \{2, \ldots, T\}$ and, particularly, it holds for $\widehat{\tau}_N$. Note that $\widehat{\tau}_N = \arg\max_t S_N(t)$. Hence, $S_N(\tau) - S_N(\widehat{\tau}_N) \leq 0$. Therefore,

$$\begin{aligned}
&2\sqrt{N}\max_{1 \leq r \leq T}|S_N(r) - \mathsf{E}S_N(r)| \\
&\geq \sqrt{N}\left[\sigma^2\left(\frac{r(\widehat{\tau}_N)}{\widehat{\tau}_N^2} - \frac{r(\tau)}{\tau^2}\right) + \mathcal{I}\{\widehat{\tau}_N > \tau\}\frac{\tau(\widehat{\tau}_N - \tau)}{\widehat{\tau}_N^2}\frac{1}{N}\sum_{i=1}^{N}\delta_i^2\right]. \quad (9)
\end{aligned}$$

If $\widehat{\tau}_N > \tau$, then the left hand side of (9) is $\mathcal{O}_\mathsf{P}(1)$ as $N \to \infty$, but the right hand side is unbounded because of Assumption C1. So, if $\widehat{\tau}_N \leq \tau$, then

$$0 \xleftarrow[N \to \infty]{\mathsf{P}} 2\max_{1 \leq r \leq T}|S_N(r) - \mathsf{E}S_N(r)| \geq \sigma^2\left(\frac{r(\widehat{\tau}_N)}{\widehat{\tau}_N^2} - \frac{r(\tau)}{\tau^2}\right),$$

which yields due to the monotonicity of $r(t)/t^2$ that $\mathsf{P}[\widehat{\tau}_N = \tau] \to 1$ as $N \to \infty$. $\qquad \square$

Proof (of Theorem 4) Let us define $\widehat{\epsilon}_{i,t} := \sigma^{-1} \sum_{s=1}^{t} \widehat{e}_{i,s}$, $\widehat{\epsilon}_{i,t}^* := \sigma^{-1} \sum_{s=1}^{t} \widehat{e}_{i,s}^*$,

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

and

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

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

Realize that $\widehat{\epsilon}_{i,t}$ depends on $\widehat{\tau}_N$ and, hence, it depends on N. Thus, $\widehat{\epsilon}_{i,t} \equiv \widehat{\epsilon}_{i,t}(N)$. Since Assumption C2 holds, then according to the bootstrap multivariate CLT for triangular arrays (Theorem 7) of T-dimensional vectors $\boldsymbol{\xi}_{N,i} = [\widehat{\epsilon}_{i,1}(N), \ldots, \widehat{\epsilon}_{i,T}(N)]^\top$ with $k_N = N$, we have

$$\mathsf{P}\left[\boldsymbol{\Gamma}_N^{-1/2} [\widehat{U}_N^*(1), \ldots, \widehat{U}_N^*(T)]^\top \leq \mathbf{x} \big| \mathbb{Y} \right] - \mathsf{P}\left[\boldsymbol{\Gamma}_N^{-1/2} [\widehat{U}_N(1), \ldots, \widehat{U}_N(T)]^\top \leq \mathbf{x} \right]$$

$$\xrightarrow[N\to\infty]{\mathsf{P}} 0, \quad \forall \mathbf{x} \in \mathbb{R}^T,$$

where $\boldsymbol{\Gamma}_N = \mathsf{Var}\,[\widehat{\epsilon}_{i,1}, \ldots, \widehat{\epsilon}_{i,T}]^\top$.

Now, it is sufficient to realize that $[\widehat{U}_N(1), \ldots, \widehat{U}_N(T)]^\top$ has an approximate multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Gamma} = \lim_{N\to\infty} \boldsymbol{\Gamma}_N$. Using the law of total variance,

$$\mathsf{Var}\,\widehat{\epsilon}_{i,t} = \mathsf{E}[\mathsf{Var}\,\{\widehat{\epsilon}_{i,t}|\widehat{\tau}_N\}] + \mathsf{Var}\,[\mathsf{E}\{\widehat{\epsilon}_{i,t}|\widehat{\tau}_N\}]$$

$$= \sum_{\pi=1}^{T} \mathsf{P}[\widehat{\tau}_N = \pi]\mathsf{Var}\,[\widehat{\epsilon}_{i,t}|\widehat{\tau}_N = \pi] + \sum_{\pi=1}^{T} \mathsf{P}[\widehat{\tau}_N = \pi]\{\mathsf{E}[\widehat{\epsilon}_{i,t}|\widehat{\tau}_N = \pi]\}^2$$

$$- \left\{ \sum_{\pi=1}^{T} \mathsf{P}[\widehat{\tau}_N = \pi]\mathsf{E}[\widehat{\epsilon}_{i,t}|\widehat{\tau}_N = \pi] \right\}^2.$$

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

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

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

$$\lim_{N\to\infty} \mathsf{Cov}\,\left(\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v} \right) = \lim_{N\to\infty} \mathsf{Cov}\,\left(\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v}|\widehat{\tau}_N = \tau \right).$$

Note that

$$\left(\widehat{e}_{i,t}|\widehat{\tau}_N = \tau\right) = \begin{cases} \sigma(\varepsilon_{i,t} - \overline{\varepsilon}_{i,\tau}), & t \leq \tau; \\ \sigma(\varepsilon_{i,t} - \widetilde{\varepsilon}_{i,\tau}), & t > \tau; \end{cases}$$

where

$$\overline{\varepsilon}_{i,t} = \frac{1}{t} \sum_{s=1}^{t} \varepsilon_{i,s} \quad \text{and} \quad \widetilde{\varepsilon}_{i,t} = \frac{1}{T-t} \sum_{s=t+1}^{T} \varepsilon_{i,s}.$$

Taking into account the definitions of $r(t)$, $R(t, v)$, and $S(t, v, d)$ together with some simple algebra, we obtain that $\mathsf{Var}\,[\widehat{\epsilon}_{i,s}|\widehat{\tau}_N = \tau] = \gamma_{t,t}(\tau)$ and $\mathsf{Cov}\,\left(\widehat{\epsilon}_{i,t}, \widehat{\epsilon}_{i,v}|\widehat{\tau}_N = \tau\right) = \gamma_{t,v}(\tau)$ for $t < v$, where the elements $\gamma_{t,t}(\tau)$ and $\gamma_{t,v}(\tau)$ are as in the statement of Theorem 4.

Then the sum in the nominator of $\mathcal{R}_N^*(T)$ can be alternatively rewritten as

$$\frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} \sum_{r=1}^{s} \left(\widehat{Y}_{i,r}^* - \overline{\widehat{Y}}_{i,t}^*\right) = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^{N} \left\{\left[\sum_{r=1}^{s} \widehat{Y}_{i,r}^*\right] - \frac{s}{t} \sum_{v=1}^{t} \widehat{Y}_{i,v}^*\right\}$$

$$= \widehat{U}_N^*(s) - \frac{s}{t} \widehat{U}_N^*(t).$$

Concerning the denominator of $\mathcal{R}_N^*(T)$, one needs to perform a similar calculation as in the proof of Theorem 1 with $V_N(t)$, i.e., to define $\widehat{V}_N(t)$ and $\widehat{V}_N^*(t)$ analogously to $\widehat{U}_N(t)$ and $\widehat{U}_N^*(t)$ as $V_N(t)$ is to $U_N(t)$. Applying the Cramér–Wold theorem completes the proof. □

Proof (of Theorem 5) Recall the notation from the proof of Theorem 4. Under H_0, B2, and C2 it holds

$$\lim_{N \to \infty} \mathsf{P}[\widehat{\tau}_N = T] = 1.$$

Then in view of (4),

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

□

Proof (of Theorem 6) The Lyapunov condition (Billingsley 1986, [p. 371]) for a triangular array of random variables $\{\xi_{n,k_n}\}_{n=1}^{\infty}$ is satisfied due to (5) and (6), i.e., for $\omega = 2$:

$$\frac{1}{\sqrt{k_n \varsigma_n^2}^{2+\omega}} \sum_{i=1}^{k_n} \mathsf{E}|\xi_{n,i}|^{2+\omega} \leq \frac{k_n^{-\omega/2}}{\varsigma_n^{2+\omega}} \sup_{\iota \in \mathbb{N}} \mathsf{E}|\xi_{\iota,1}|^{2+\omega} \to 0, \quad n \to \infty.$$

Therefor, the CLT for $\{\xi_{n,k_n}\}_{n=1}^{\infty}$ holds and

$$\sup_{x\in\mathbb{R}}\left|\mathsf{P}\left[\frac{\sqrt{k_n}}{\sqrt{\varsigma_n^2}}\bar{\xi}_n \leq x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}dt\right| \xrightarrow[n\to\infty]{} 0.$$

Now, to prove the theorem, it suffices to show the following three statements:

(i) $\sup_{x\in\mathbb{R}}\left|\mathsf{P}^*_\xi\left[\frac{\sqrt{k_n}}{\sqrt{\mathsf{Var}_{\mathsf{P}^*_\xi}\xi^*_{n,1}}}\left(\bar{\xi}^*_n - \mathsf{E}_{\mathsf{P}^*_\xi}\bar{\xi}^*_n\right) \leq x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}dt\right| \xrightarrow[n\to\infty]{\mathsf{P}} 0$;

(ii) $\mathsf{Var}_{\mathsf{P}^*_\xi}\xi^*_{n,1} - \varsigma_n^2 \xrightarrow[n\to\infty]{\mathsf{P}} 0$;

(iii) $\mathsf{E}_{\mathsf{P}^*_\xi}\bar{\xi}^*_n = \bar{\xi}_n$, $[\mathsf{P}] - a.s.$

Proving *(iii)* is trivial, because $\mathsf{E}_{\mathsf{P}^*_\xi}\bar{\xi}^*_n = \mathsf{E}_{\mathsf{P}^*_\xi}\xi^*_{n,1} = k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i} = \bar{\xi}_n$, $[\mathsf{P}]$-*a.s.*

Let us calculate the conditional variance of the bootstrapped variable $\xi^*_{n,1}$:

$\mathsf{Var}_{\mathsf{P}^*_\xi}\xi^*_{n,1} = \mathsf{E}_{\mathsf{P}^*_\xi}\xi^{*2}_{n,1} - (\mathsf{E}_{\mathsf{P}^*_\xi}\xi^*_{n,1})^2 = k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}^2 - \left(k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}\right)^2$, $[\mathsf{P}]$-*a.s.*
The weak law of large numbers together with (5) provides

$$\bar{\xi}_n - n^{-1}\sum_{i=i}^{k_n}\mathsf{E}_{\mathsf{P}}\xi_{n,i} = \bar{\xi}_n \xrightarrow[n\to\infty]{\mathsf{P}} 0$$

and

$$0 \xleftarrow[n\to\infty]{\mathsf{P}} k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}^2 - \left(k_n^{-1}\sum_{i=1}^{k_n}\xi_{n,i}\right)^2 - \mathsf{E}_{\mathsf{P}}\xi_{n,1}^2 = \mathsf{Var}_{\mathsf{P}^*_\xi}\xi^*_{n,1} - \varsigma_n^2.$$

The last result of the WLLN is true, because (5) implies

$$k_n^{-2}\sum_{i=1}^{k_n}\mathsf{Var}_{\mathsf{P}}\xi_{n,i}^2 \leq k_n^{-2}\sum_{i=1}^{k_n}\mathsf{E}_{\mathsf{P}}\xi_{n,i}^4 \leq k_n^{-1}\sup_{\iota\in\mathbb{N}}\mathsf{E}_{\mathsf{P}}\xi_{\iota,1}^4 \xrightarrow[n\to\infty]{} 0.$$

Thus *(ii)* is proved.

The Berry–Esseen–Katz theorem (see Katz 1963) with $g(x) = |x|^\epsilon$, $\epsilon > 0$ for the bootstrapped sequence of *iid* (with respect to P^*) random variables $\{\xi^*_{n,i}\}_{i=1}^{k_n}$ results in

$$\sup_{x\in\mathbb{R}}\left|\mathsf{P}^*_\xi\left[\frac{\sqrt{k_n}}{\sqrt{\mathsf{Var}_{\mathsf{P}^*_\xi}\xi^*_{n,1}}}\left(\bar{\xi}^*_n - \mathsf{E}_{\mathsf{P}^*_\xi}\bar{\xi}^*_n\right) \leq x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}dt\right|$$

$$\leq Ck_n^{-\epsilon/2}\mathsf{E}_{\mathsf{P}^*_\xi}\left|\frac{\xi^*_{n,1} - \mathsf{E}_{\mathsf{P}^*_\xi}\xi^*_{n,1}}{\mathsf{Var}_{\mathsf{P}^*_\xi}\xi^*_{n,1}}\right|^{2+\epsilon} \quad [\mathsf{P}] - a.s., \tag{10}$$

for all $n \in \mathbb{N}$ where $C > 0$ is an absolute constant.

The Jensen inequality and Minkowski inequality provide an upper bound for the nominator from the right-hand side of (10):

$$
\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} |\xi^*_{n,1} - \mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} \xi^*_{n,1}|^{2+\epsilon} = k_n^{-1} \sum_{i=1}^{k_n} \left| \xi_{n,i} - k_n^{-1} \sum_{j=1}^{k_n} \xi_{n,j} \right|^{2+\epsilon}
$$

$$
\leq k_n^{-1} \left\{ \left(\sum_{i=1}^{k_n} |\xi_{n,i}|^{2+\epsilon} \right)^{1/(2+\epsilon)} + k_n^{-(1+\epsilon)/(2+\epsilon)} \left| \sum_{j=1}^{k_n} \xi_{n,j} \right| \right\}^{2+\epsilon}
$$

$$
\leq 2^{1+\epsilon} k_n^{-1} \sum_{i=1}^{k_n} |\xi_{n,i}|^{2+\epsilon} + 2^{1+\epsilon} \left| k_n^{-1} \sum_{i=1}^{k_n} \xi_{n,i} \right|^{2+\epsilon} \quad [\mathsf{P}] - a.s.
$$

The right-hand side of the previously derived upper bound is uniformly bounded in probability P, because of Markov's inequality and (5). Indeed, for fixed $\eta > 0$

$$
\mathsf{P}\left[k_n^{-1} \sum_{i=1}^{k_n} |\xi_{n,i}|^{2+\epsilon} \geq \eta \right] \leq \eta^{-1} k_n^{-1} \sum_{i=1}^{k_n} \mathsf{E}_{\mathsf{P}} |\xi_{n,i}|^{2+\epsilon}
$$

$$
\leq \eta^{-1} \sup_{\iota \in \mathbb{N}} \mathsf{E}_{\mathsf{P}} |\xi_{\iota,1}|^{2+\epsilon} < \infty, \quad \forall n \in \mathbb{N}
$$

and

$$
\mathsf{P}\left[\left| k_n^{-1} \sum_{i=1}^{k_n} \xi_{n,i} \right| \geq \eta \right] \leq \eta^{-1} k_n^{-1} \mathsf{E}_{\mathsf{P}} \left| \sum_{i=1}^{k_n} \xi_{n,i} \right|
$$

$$
\leq \eta^{-1} \sup_{\iota \in \mathbb{N}} \mathsf{E}_{\mathsf{P}} |\xi_{\iota,1}| < \infty, \quad \forall n \in \mathbb{N}.
$$

Since $\mathsf{E}_{\mathsf{P}^*_{\boldsymbol{\xi}}} |\xi^*_{n,1} - \mathsf{E}_{\mathsf{P}^*} \xi^*_{n,1}|^{2+\epsilon}$ is bounded in probability P uniformly over n and the denominator of the right-hand side of (10) is uniformly bounded away from zero due to (6), then the left-hand side of (10) converges in probability P to zero as n tends to infinity. So, *(i)* is proved as well. □

Proof (of Theorem 7) According to the Cramér–Wold theorem, it is sufficient to ensure that all assumptions of one-dimensional bootstrap CLT (6) for triangular arrays are valid for any linear combination of the elements of the random vector $\boldsymbol{\xi}_{n,1}$, $n \in \mathbb{N}$.

For arbitrary fixed $\mathbf{t} \in \mathbb{R}^q$ using the Jensen inequality, we get

$$
\sup_{n \in \mathbb{N}} \mathsf{E}_{\mathsf{P}} |\mathbf{t}^\top \boldsymbol{\xi}_{n,1}|^4 \leq q^3 \sup_{n \in \mathbb{N}} \sum_{j=1}^{q} t_j^4 \mathsf{E}_{\mathsf{P}} |\xi_{n,1}^{(j)}|^4 \leq q^4 \max_{j=1,\ldots,q} t_j^4 \sup_{n \in \mathbb{N}} \mathsf{E}_{\mathsf{P}} |\xi_{n,1}^{(j)}|^4 < \infty.
$$

Hence, assumption (7) implies assumption (5) for the random variables $\{\mathbf{t}^\top \boldsymbol{\xi}_{n,k_n}\}_{n \in \mathbb{N}}$.

Similarly, assumption (8) implies assumption (6) for such an arbitrary linear combination, i.e., positive definiteness of the matrix $\boldsymbol{\Gamma}$ yields

$$\liminf_{n\to\infty} \mathsf{Var}_{\mathsf{P}} \mathbf{t}^{\top} \boldsymbol{\xi}_{n,1} = \liminf_{n\to\infty} \mathbf{t}^{\top} \left(\mathsf{Var}_{\mathsf{P}} \boldsymbol{\xi}_{n,1} \right) \mathbf{t} \geq \mathbf{t}^{\top} \left(\liminf_{n\to\infty} \boldsymbol{\Gamma}_n \right) \mathbf{t} = \mathbf{t}^{\top} \boldsymbol{\Gamma} \mathbf{t} > 0.$$

$\square$

References

Andrews DWK (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59(3):817–858

Bai J (2010) Common breaks in means and variances for panel data. J Econom 157(1):78–92

Billingsley P (1986) Probability and measure, 2nd edn. Wiley, New York

Chan J, Horváth L, Hušková M (2013) Change-point detection in panel data. J Stat Plan Inference 143(5):955–970

Chen Z, Tian Z (2014) Ratio tests for variance change in nonparametric regression. Stat J Theor Appl Stat 48(1):1–16

Csörgő M, Horváth L (1997) Limit theorems in change-point analysis. Wiley, Chichester

Horváth L, Horváth Z, Hušková M (2009) Ratio tests for change point detection. In: Balakrishnan N, Peña EA, Silvapulle MJ (eds) Beyond parametrics in interdisciplinary research: Festschrift in honor of professor Pranab K. Sen, vol 1. IMS Collections, Beachwood, pp 293–304

Horváth L, Hušková M (2012) Change-point detection in panel data. J Time Ser Anal 33(4):631–648

Hušková M, Kirch C (2012) Bootstrapping sequential change-point tests for linear regression. Metrika 75(5):673–708

Hušková M, Kirch C, Prášková Z, Steinebach J (2008) On the detection of changes in autoregressive time series, II. Resampling procedures. J Stat Plan Inference 138(6):1697–1721

Katz ML (1963) Note on the Berry–Esseen theorem. Ann Math Stat 34(3):1107–1108

Lindner AM (2009) Stationarity, mixing, distributional properties and moments of GARCH(p, q)-processes. In: Andersen TG, Davis RA, Kreiss JP, Mikosch T (eds) Handbook of financial time series. Springer, Berlin, pp 481–496

Liu Y, Zou C, Zhang R (2008) Empirical likelihood ratio test for a change-point in linear regression model. Commun Stat Theory Methods 37(16):2551–2563

Madurkayová B (2011) Ratio type statistics for detection of changes in mean. Acta Univ Carol Math Phys 52(1):47–58

Meyers GG, Shi P (2011) Loss reserving data pulled from NAIC Schedule P. http://www.casact.org/research/index.cfm?fa=loss_reserves_data. [Online; Updated September 01, 2011; Accessed June 10, 2014]

Pešta M, Hudecová Š (2012) Asymptotic consistency and inconsistency of the chain ladder. Insur Math Econ 51(2):472–479