The New England Journal of Statistics in Data Science Volume 4, 105–130 (2026)
DOI: https://doi.org/10.51387/25-NEJSDS91

Locally Adaptive Modeling of Unconditional Heteroskedasticity

Matthias R. FENGLER*, Bruno JÄGER, and Ostap OKHRIN

Abstract

We study local change point detection in variance using generalized likelihood ratio tests. Building on [24], we utilize the multiplier bootstrap to approximate the unknown, non-asymptotic distribution of the test statistic and introduce a multiplicative bias correction that improves upon the existing additive version. This proposed correction offers a clearer interpretation of the bootstrap estimators while significantly reducing computational costs. Simulation results demonstrate that our method performs comparably to the original approach. We apply it to the growth rates of U.S. inflation, industrial production, and Bitcoin returns.

KEYWORDS AND PHRASES: Generalized likelihood ratio test, Multiplier bootstrap, Local change point detection, Economic and financial variance.

1. INTRODUCTION

Ever since the seminal contributions of [5], [1], and [26], modeling the variance of financial time series data has emerged as a key research area in econometrics. Given its critical importance to the financial industry, where precise estimates and forecasts of variance are essential for applications in risk management, portfolio management, and derivative pricing, two families of models have gained widespread acceptance and, by any standards, now dominate practical applications: (*i*) the ARCH(q) and GARCH(p, q) models [5, 1]; and (*ii*) the stochastic variance models [26].

An aspect that both model families share is their assumption of conditional heteroskedasticity and global stationarity in the underlying data sequence. However, an alternative approach to interpreting the stylized facts of financial data, and hence to modeling variance, is to assume nonstationarity triggered by structural breaks in the variance of unknown form, i.e., unconditional heteroskedasticity.[1] A prominent example is the IGARCH effect, which refers to the empirical observation that the autocorrelation function of squared log returns exhibits long-range dependence and that the sum of the estimated dynamic parameters of a GARCH$(1, 1)$ model often approaches unity as the sample size increases [16].

Acknowledging nonstationarity has led to the development of variance modeling approaches that aim to identify unknown breakpoints that indicate shifts between stationary data-generating processes. These methods enable the approximation of such processes using piecewise stationary models, which aim to detect intervals of homogeneity where the chosen model is supported at a given significance level.[2]

While various locally stationary linear and nonlinear models of variance can be proposed, our approach prioritizes simplicity by targeting the approximation of a constant variance [15, 27, 23, 4, 14]. This simplification has two benefits. It reduces the challenge of identifying the prevailing interval of homogeneity for accurate estimation of a complex model; additionally, it enhances intuition of practitioners, who often work with a rolling window estimator for variance. Put differently, we perform a local change point (LCP) detection of unconditional variance, effectively mimicking a rolling window estimator that selects its window width in a data-driven manner.

To test for intervals of homogeneity, we use the maximum across a set of generalized likelihood ratio test (LRT) statistics computed for a set of candidate breakpoints. We consider two tests: a test for homogeneity in variance (with arbitrary means) and a test for complete homogeneity, i.e., for constancy in mean *and* variance. With independently and identically normally distributed data, each of these tests would, according to Wilks' phenomenon [28], follow a $\chi^2(1)$ or $\chi^2(2)$ distribution with increasing sample size. However, because our test statistic transforms generalized likelihood ratios and the assumption of Gaussianity can be problematic, we use a multiplier bootstrap to approximate the distribution of the test statistics. This approach builds on theory developed by [21]. Its utility for calibrating critical values, particularly in small samples, has also been demonstrated in [24], [13], and [12].

Within the multiplier bootstrap framework for testing the variance, we propose substituting the traditional additive bias correction described by [24] with a multiplicative correction. This approach is particularly suitable for variance testing because it offers a compelling interpretation of the bootstrap estimates. Unlike under the additive bias correction, the bootstrap estimators can even be derived in closed

*Corresponding author.
[1]For a recent discussion of this approach, see [14].
[2]See [11], [15], [27], [23], [7], [29], [3], [20], [2], [4], [18] and [14].

form. Consequently, the bootstrap procedure simplifies to drawing the multiplier weights and calculating the bootstrap estimates analytically, with the testing decision based on their distribution. This approach significantly reduces the computational costs associated with the bootstrap.

Our extensive simulations suggest that the tests that incorporate the multiplicative bias correction possess similar discrepancies in size and power as the tests using the additive one. However, under misspecification, tests using the multiplicative bias correction exhibit slightly higher size discrepancies than those using the additive correction, but they more effectively detect unconditional changes in the LCP simulations. Compared to standard fixed window estimators, the tests also perform well, even when applied to a globally stationary, conditionally heteroskedastic GARCH framework.

In summary, compared to methods targeting the same goals, our methodological approach stands out for its remarkable simplicity and directness despite requiring bootstrap simulation. For instance, [23] and [18] analyze data in the frequency domain, resulting in tests with non-standard asymptotic distributions involving functionals of Brownian motions. Meanwhile, [15] requires extensive prior simulations to calibrate parameters that control the false alarm rate. Closely related, [4] apply piecewise linear functions; however, their approach is purely constructive, requires dynamic programming, and lacks an asymptotic analysis.

This paper is structured as follows. In Section 2, we outline our modeling approach, present the tests, and discuss the LCP detection algorithm incorporating the proposed homogeneity tests. Section 3 provides simulations, while we apply local change point detection to the growth rates of U.S. inflation and industrial production, as well as Bitcoin returns in Section 4. Section 5 concludes.

2. LOCALLY ADAPTIVE MODELING

2.1 Homogeneity Testing

Let $\{X_t\}$ be a sequence of random variables on a given probability space $(\Omega, \mathcal{F}, \mathrm{P})$. We assume X_t to be independently distributed according to a probability measure P_0, which belongs to a parametric family $\mathrm{P}(\theta)$, i.e., $\mathrm{P}_0 \in \{\mathrm{P}(\theta), \theta \in \Theta \subset \mathbb{R}^p\}$, with p being the number of parameters. On an interval $\mathcal{I} = [a, b]$ with length $n_{\mathcal{I}}$, we aim to test the null hypothesis H_0 of no change in θ. H_0 may depend on all or just a subset of the elements in θ. The alternative hypothesis is that there is a break point $\tau \in (a, b)$, leading to distinct parameterizations for X_t in the two intervals; specifically, one for $X_{t\in L_\tau} = \{X_t\}_{t\in L_\tau}$ where $L_\tau = [a, \tau]$, and another one for $X_{t\in R_\tau} = \{X_t\}_{t\in R_\tau}$ where $R_\tau = (\tau, b]$. Summarizing, we test

$$\mathrm{H}_0 : X_{t\in\mathcal{I}} \sim \mathrm{P}_0(\theta_{\mathcal{I}}), \qquad \theta_{\mathcal{I}} \in \Theta,$$

against the alternative

$$\mathrm{H}_1 : X_{t\in L_\tau} \sim \mathrm{P}_0(\theta_{L_\tau}) \text{ and } X_{t\in R_\tau} \sim \mathrm{P}_0(\theta_{R_\tau}) \,,$$

where $\theta_{L_\tau} \neq \theta_{R_\tau}$ and $L_\tau \cup R_\tau = \mathcal{I}$.

Following [20], we consider a set of candidate breakpoints $\mathcal{T}(\mathcal{I}) \subset \mathcal{I}$ to test H_0. For each candidate $\tau \in \mathcal{T}(\mathcal{I})$, we conduct a test belonging to the family of generalized likelihood ratio tests (LRT)—see [9]. More specifically, for every fixed τ, we employ the statistic

$$\begin{aligned} T_{\mathcal{I},\tau} = &\sup_{\theta\in\Theta} \mathcal{L}_{L_\tau}(\theta|X_{t\in L_\tau}) + \sup_{\theta\in\Theta} \mathcal{L}_{R_\tau}(\theta|X_{t\in R_\tau}) \\ &- \sup_{\theta\in\Theta} \mathcal{L}_{\mathcal{I}}(\theta|X_{t\in\mathcal{I}}), \end{aligned} \tag{2.1}$$

where $\mathcal{L}(\theta|X_{t\in j}) = \sum_{t\in j} \ell_t(\theta|X_t)$ is a log-likelihood function on the respective interval, for $j = \{L_\tau, R_\tau, \mathcal{I}\}$, and $\ell_t(\theta|X_t)$ is the log-likelihood contribution of a single observation X_t at time t. It is supposed to be clear from the context when the conditioning is on the random variable X_t (e.g., for the computation of a probability) or its realization x_t (e.g., for the computation of the value of a test statistic).

To test for the presence of a breakpoint jointly among the candidates, we employ the maximum statistic, as in [20], given by

$$T_{\mathcal{I}} = \max_{\tau\in\mathcal{T}(\mathcal{I})} T_{\mathcal{I},\tau}. \tag{2.2}$$

For a specified significance level α, we need to determine the critical value $\mathfrak{z}_{\mathcal{I},\alpha}$ such that

$$\mathrm{P}(T_{\mathcal{I}} \leq \mathfrak{z}_{\mathcal{I},\alpha}) = 1 - \alpha \,,$$

implying that we reject H_0 whenever $T_{\mathcal{I}} > \mathfrak{z}_{\mathcal{I},\alpha}$. For small samples or non-Gaussian distributions, this may prove to be difficult. Thus, we will compute $\mathfrak{z}_{\mathcal{I},\alpha}$ using the multiplier bootstrap.

2.2 Multiplier Bootstrap

Following [21] as well as [24], we use the multiplier bootstrap to approximate the unknown and non-asymptotic distribution of the test statistic. The multiplier bootstrap is carried out by re-weighting the log-likelihood function. The bootstrapped log-likelihood function is given by

$$\mathcal{L}_j^\flat(\theta|X_{t\in j}) = \sum_{t\in j} u_t \ell_t(\theta|x_t) \,, \tag{2.3}$$

where $\{u_t\}$ is an *i.i.d.* sample from a sub-Gaussian distribution, independent from X_t, with $\mathrm{E}[u_t] = 1$ and $\mathrm{Var}[u_t] = 1$; here, $j = \{L_\tau, R_\tau, \mathcal{I}\}$ indexes the three samples involved.

From (2.3), we receive the bootstrap estimator

$$\hat{\theta}_j^\flat = \arg\max_{\theta\in\Theta} \mathcal{L}_j^\flat(\theta|X_{t\in j}) \,. \tag{2.4}$$

As an important property, note that

$$\hat{\theta}_j = \arg\max_{\theta\in\Theta} \mathrm{E}^\flat[\mathcal{L}_j^\flat(\theta|X_{t\in j})] = \arg\max_{\theta\in\Theta} \mathcal{L}_j(\theta|X_{t\in j}) \,, \quad (2.5)$$

where $\mathrm{E}^\flat[\mathcal{L}_j^\flat(\theta|X_{t\in j})]$ is the expectation conditional on the sample over the measure for $\{u_t\}$—the so-called bootstrap world.

As established in [21, p. 2660, Theorem 2.2], $\mathrm{P}\{\mathcal{L}_j(\hat{\theta}_j) - \mathcal{L}_j(\theta_j) > \mathfrak{z}^2/2\}$ is close in probability to the corresponding random quantity in the bootstrap world across a broad range of $\mathfrak{z}$-values. Thus, the test statistic in the bootstrap world corresponding to (2.1), which is given by

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &\sup_{\theta\in\Theta} \mathcal{L}^\flat_{L_\tau}(\theta|X_{t\in L_\tau}) + \sup_{\theta\in\Theta} \mathcal{L}^\flat_{R_\tau}(\theta|X_{t\in R_\tau}) \\ &- \sup_{\theta\in\Theta}\Big\{ \mathcal{L}^\flat_{L_\tau}(\theta|X_{t\in L_\tau}) \\ &+ \mathcal{L}^\flat_{R_\tau}(\theta+\hat{\theta}_{R_\tau}-\hat{\theta}_{L_\tau}|X_{t\in R_\tau})\Big\}, \end{aligned} \quad (2.6)$$

where $\hat{\theta}_{L_\tau}$ and $\hat{\theta}_{R_\tau}$ are the maximum likelihood (ML) estimators on each subinterval, has a distribution close to $T_{\mathcal{I},\tau}$. Therefore, we use repeated bootstrap sampling to evaluate the critical value,

$$\mathfrak{z}_{\mathcal{I},\alpha} = \inf\{\mathfrak{z}\geq 0 : \mathrm{P}(T^\flat_{\mathcal{I},\tau} > \mathfrak{z}^2/2) \leq \alpha\} \,, \quad (2.7)$$

in a data-dependent way.

In the last term of (2.6), an additive bias correction in the form $\hat{\theta}_{R_\tau} - \hat{\theta}_{L_\tau}$ is used to ensure that the simulation is carried out under H_0 [24]. It adjusts the parameters estimated on the interval R_τ to align with those estimated on L_τ. The following section introduces a multiplicative bias correction, a natural choice when testing for variance.

2.3 Testing for Homogeneity in Variance

From this section on, we specialize the discussion to the particular case, where we assume $X_t \sim \mathcal{N}(\mu,\sigma^2)$. Besides testing for complete homogeneity, involving the mean and variance simultaneously, we are also interested in proposing a test solely for homogeneity in variance. Notably, even in the case of model misspecification, the bootstrap validity continues to apply and becomes conservative under significant deviations [21]. Therefore, the use of the normal distribution in this context carries the characteristics of a quasi-maximum likelihood approach.

In the first step, we discuss the test for homogeneity in variance. We consider a left sample $X_{t\in L_\tau}$, a right sample $X_{t\in R_\tau}$, and their combination $X_{t\in\mathcal{I}}$. We wish to test whether $X_{t\in\mathcal{I}} \sim \mathcal{N}(\mu,\sigma^2)$ against the alternative that $X_{t\in L_\tau} \sim \mathcal{N}(\mu,\sigma^2)$ and $X_{t\in R_\tau} \sim \mathcal{N}(\breve{\mu},\breve{\sigma}^2)$ with $\sigma^2 \neq \breve{\sigma}^2$ and arbitrary means.

Under these assumptions, for $j = \{L_\tau, R_\tau, \mathcal{I}\}$ with sample size n_j, the log-likelihood function is given by

$$\begin{aligned} \mathcal{L}_j\left(\mu,\sigma^2 \mid X_{t\in j}\right) = &-\frac{n_j}{2}\log(2\pi) - \frac{n_j}{2}\log\sigma^2 \\ &- \frac{1}{2\sigma^2}\sum_{t=1}^{n_j}(x_t-\mu)^2 \,, \end{aligned} \quad (2.8)$$

and the LRT for homogeneity in variance becomes

$$\begin{aligned} T_{\mathcal{I},\tau} = &\sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}_{R_\tau}\left(\mu,\sigma^2 \mid X_{t\in R_\tau}\right) \\ &- \sup_{\sigma^2>0}\Big\{ \sup_{\mu\in\mathbb{R}} \mathcal{L}_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \sup_{\mu\in\mathbb{R}} \mathcal{L}_{R_\tau}\left(\mu,\sigma^2 \mid X_{t\in R_\tau}\right)\Big\}. \end{aligned} \quad (2.9)$$

It is, of course, well-known that

$$\{\bar{x}_j, \hat{\sigma}_j^2\} = \arg\max_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}_j\left(\mu,\sigma^2 \mid X_{t\in j}\right),$$

where $\hat{\sigma}_j^2 = n_j^{-1}\sum_{t\in j}(x_t-\bar{x}_j)^2$ and $\bar{x}_j = n_j^{-1}\sum_{t\in j} x_t$, for $j = \{L_\tau, R_\tau\}$. The ML estimator over the entire interval $\mathcal{I}$ is obtained as the pooled variance of both subintervals weighted by the sample sizes, i.e., $\hat{\sigma}^{\circ 2}_{\mathcal{I}} = n_{\mathcal{I}}^{-1}(n_{L_\tau}\hat{\sigma}^2_{L_\tau} + n_{R_\tau}\hat{\sigma}^2_{R_\tau})$.

Thus, the test statistic for homogeneity in variance can be written as

$$\begin{aligned} T_{\mathcal{I},\tau} &= \mathcal{L}_{L_\tau}\left(\bar{x}_{L_\tau},\hat{\sigma}^2_{L_\tau} \mid X_{t\in L_\tau}\right) + \mathcal{L}_{R_\tau}\left(\bar{x}_{R_\tau},\hat{\sigma}^2_{R_\tau} \mid X_{t\in R_\tau}\right) \\ &\quad - \mathcal{L}_{L_\tau}\left(\bar{x}_{L_\tau},\hat{\sigma}^{\circ 2}_{\mathcal{I}} \mid X_{t\in L_\tau}\right) \\ &\quad - \mathcal{L}_{R_\tau}\left(\bar{x}_{R_\tau},\hat{\sigma}^{\circ 2}_{\mathcal{I}} \mid X_{t\in R_\tau}\right), \\ &= -\frac{n_{L_\tau}}{2}\log\hat{\sigma}^2_{L_\tau} - \frac{n_{R_\tau}}{2}\log\hat{\sigma}^2_{R_\tau} + \frac{n_{\mathcal{I}}}{2}\log\hat{\sigma}^{\circ 2}_{\mathcal{I}}. \end{aligned} \quad (2.10)$$

In the case of a single break point and under correct model specification, this corresponds to the classical likelihood ratio test, which follows a $\chi^2(1)$-distribution due to [28].

Additive bias correction. Following (2.3), the multiplier bootstrap log-likelihood function obtains as

$$\begin{aligned} \mathcal{L}^\flat_j\left(\mu,\sigma^2 \mid X_{t\in j}\right) &= -\frac{1}{2}\log(2\pi)\sum_{t=1}^{n_j} u_t - \frac{1}{2}\log\sigma^2\sum_{t=1}^{n_j} u_t \\ &\quad - \frac{1}{2\sigma^2}\sum_{t=1}^{n_j} u_t\,(x_t-\mu)^2 \,, \\ &= -\frac{n_j}{2}\log(2\pi) - \frac{n_j}{2}\log\sigma^2 - \frac{n_j}{2}\frac{\hat{\sigma}^{*2}_j}{\sigma^2} \\ &\quad - \frac{n_j}{2\sigma^2}(\bar{x}^*_j-\mu)^2. \end{aligned} \quad (2.11)$$

Now, $\hat{\sigma}^{*2}_j = \sum_{t\in j} w_t(x_t-\bar{x}^*_j)^2$ and $\bar{x}^*_j = \sum_{t\in j} w_t x_t$, for $j = \{L_\tau, R_\tau\}$, are the ML estimators of the variance and the mean in the bootstrap world, respectively. We define

$w_t = u_t n_j^{-1}$, where $\{u_t\}_{t\in j}$ are drawn as discussed in Section 2.2. Note that for the second line in (2.11) to hold, we assume that $\sum_{t\in j} u_t = n_j$, because it helps us keep the formulae more transparent. During simulation, this property can be ensured by normalizing $\{u_t\}_{t\in j}$ in each bootstrap iteration. While this standardization introduces a slight linear dependence within the sampled $\{u_t\}$, to our experience, it enables the test to better keep its size because the variance of the bootstrap estimators is reduced.[3]

The bootstrapped generalized LRT becomes

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &\sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}^\flat_{R_\tau}\left(\mu,\sigma^2 \mid X_{t\in R_\tau}\right) \\ &- \sup_{\sigma^2>0}\Bigg\{ \sup_{\mu\in\mathbb{R}} \mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &\quad + \sup_{\mu\in\mathbb{R}} \mathcal{L}^\flat_{R_\tau}\left(\mu,\sigma^2+\hat\sigma_a^2 \mid X_{t\in R_\tau}\right)\Bigg\}, \end{aligned} \tag{2.12}$$

where $\hat\sigma_a^2 = \hat\sigma^2_{R_\tau} - \hat\sigma^2_{L_\tau}$ is the additive bias correction introduced in (2.6).

It is straightforward that

$$\{\bar x_j^*, \hat\sigma_j^{*2}\} = \arg\max_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}^\flat_j\left(\mu,\sigma^2 \mid X_{t\in j}\right),$$

for $j = \{L_\tau, R_\tau\}$. Thus, the first two summands in (2.12) are

$$\begin{aligned} \sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}^\flat_j\left(\mu,\sigma^2 \mid X_{t\in j}\right) = &-\frac{n_j}{2}\log(2\pi) \\ &-\frac{n_j}{2}\log\hat\sigma_j^{*2} - \frac{n_j}{2}. \end{aligned} \tag{2.13}$$

In contrast, the third term in (2.12) is more involved. We can still concentrate out the mean, after which we obtain the following function involving σ^2 only:

$$\begin{aligned} f^\flat_{HVa}(\sigma^2) = &\mathcal{L}^\flat_{L_\tau}\left(\bar x^*_{L_\tau},\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}^\flat_{R_\tau}\left(\bar x^*_{R_\tau},\sigma^2+\hat\sigma_a^2 \mid X_{t\in R_\tau}\right) \\ = &-\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{L_\tau}}{2}\log\sigma^2 - \frac{n_{L_\tau}}{2}\frac{\hat\sigma^{*2}_{L_\tau}}{\sigma^2} \\ &-\frac{n_{R_\tau}}{2}\log\left(\sigma^2+\hat\sigma_a^2\right) - \frac{n_{R_\tau}}{2}\frac{\hat\sigma^{*2}_{R_\tau}}{\sigma^2+\hat\sigma_a^2}. \end{aligned} \tag{2.14}$$

As we discuss in Appendix A.1, we find the maximizer of $f^\flat_{HVa}$ numerically as a root to a cubic polynomial in σ^2. Our simulations show that for a sufficiently large sample size the discriminant of the cubic polynomial is always negative, implying the existence of a single real root. However, the discriminant can sometimes be positive for very small sample sizes. In such instances, we skip the iteration and draw a new set of weights u_t until the desired number of bootstrap samples is obtained.

Denote the maximizer of $f^\flat_{HVa}$ by $\hat\sigma^{\flat*2}_{HVa}$. Then, the bootstrapped test statistic is given by

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &-\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{\mathcal{I}}}{2} - \frac{n_{L_\tau}}{2}\log\hat\sigma^{*2}_{L_\tau} - \frac{n_{R_\tau}}{2}\log\hat\sigma^{*2}_{R_\tau} \\ &- f^\flat_{HVa}\left(\hat\sigma^{\flat*2}_{HVa}\right). \end{aligned} \tag{2.15}$$

Multiplicative bias correction. We now consider the proposed multiplicative bias correction, which takes the form $\hat\sigma_m^2 = \hat\sigma^2_{R\tau}/\hat\sigma^2_{L_\tau}$. Instead of (2.12), we utilize

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &\sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}^\flat_{R_\tau}\left(\mu,\sigma^2 \mid X_{t\in R_\tau}\right) \\ &- \sup_{\sigma^2>0}\Bigg\{ \sup_{\mu\in\mathbb{R}} \mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &\quad + \sup_{\mu\in\mathbb{R}} \mathcal{L}^\flat_{R_\tau}\left(\mu,\sigma^2\hat\sigma_m^2 \mid X_{t\in R_\tau}\right)\Bigg\}. \end{aligned} \tag{2.16}$$

As before, we have $\{\bar x_j^*, \hat\sigma_j^{*2}\} = \arg\max_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}^\flat_j\left(\mu,\sigma^2 \mid X_{t\in j}\right)$, for $j = \{L_\tau, R_\tau\}$. Thus, the first two summands in (2.16) are as in (2.13). Proceeding as above, we can rewrite the third term in (2.16) as

$$\begin{aligned} f^\flat_{HVm}(\sigma^2) = &\mathcal{L}^\flat_{L_\tau}\left(\bar x^*_{L_\tau},\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}^\flat_{R_\tau}\left(\bar x^*_{R_\tau},\sigma^2\hat\sigma_m^2 \mid X_{t\in R_\tau}\right) \\ = &-\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{L_\tau}}{2}\log\sigma^2 - \frac{n_{L_\tau}}{2}\frac{\hat\sigma^{*2}_{L_\tau}}{\sigma^2} \\ &-\frac{n_{R_\tau}}{2}\log\left(\sigma^2\hat\sigma_m^2\right) - \frac{n_{R_\tau}}{2}\frac{\hat\sigma^{*2}_{R_\tau}}{\sigma^2\hat\sigma_m^2}, \end{aligned} \tag{2.17}$$

which notably has the unique maximizer

$$\hat\sigma^{\flat*2}_{HVm} = \frac{1}{n_{\mathcal{I}}}\left(n_{L_\tau}\hat\sigma^{*2}_{L_\tau} + n_{R_\tau}\hat\sigma^{*2}_{R_\tau}\frac{\hat\sigma^2_{L_\tau}}{\hat\sigma^2_{R_\tau}}\right). \tag{2.18}$$

Unlike the additive bias correction, which lacks a closed-form solution, this estimator offers a compelling and intuitive interpretation: it is a pooled variance estimator incorporating the bias correction. Thus, the multiplicative bias correction provides significant computational advantages compared to the additive bias correction, which requires numerical root-finding methods (see Appendix A.1 for details).

Summarizing, the bootstrapped test statistic is

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &-\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{\mathcal{I}}}{2} - \frac{n_{L_\tau}}{2}\log\hat\sigma^{*2}_{L_\tau} - \frac{n_{R_\tau}}{2}\log\hat\sigma^{*2}_{R_\tau} \\ &- f^\flat_{HVm}\left(\hat\sigma^{\flat*2}_{HVm}\right). \end{aligned} \tag{2.19}$$

[3] In Appendix B, we report the general formulae for all tests that are valid for samples of $\{u_t\}$ without standardization, and present simulations comparing the size properties of both approaches.

2.4 Testing for Complete Homogeneity

We now discuss testing for complete homogeneity. More specifically, we test whether $X_{t\in\mathcal{I}} \sim \mathcal{N}\left(\mu, \sigma^2\right)$ against the alternative that $X_{t\in L_\tau} \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and $X_{t\in R_\tau} \sim \mathcal{N}\left(\breve{\mu}, \breve{\sigma}^2\right)$ where $\mu \neq \breve{\mu}$ and/or $\sigma^2 \neq \breve{\sigma}^2$. Of course, this test is closely related to the test for homogeneity in variance. Because (2.8) represents the log-likelihood function for all $j = \{L_\tau, R_\tau, \mathcal{I}\}$, the test statistic for complete homogeneity is

$$T_{\mathcal{I},\tau} = -\frac{n_{L_\tau}}{2}\log\hat{\sigma}^2_{L_\tau} - \frac{n_{R_\tau}}{2}\log\hat{\sigma}^2_{R_\tau} + \frac{n_{\mathcal{I}}}{2}\log\hat{\sigma}^2_{\mathcal{I}}, \quad (2.20)$$

where $\hat{\sigma}^2_{\mathcal{I}}$ is the usual ML estimator of the variance on $\mathcal{I}$. Under correct model specification, the test with a single break point corresponds to the classical likelihood ratio test which follows a $\chi^2(2)$-distribution.

Additive bias correction. The bootstrapped generalized LRT now takes the form

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &\sup_{\mu\in\mathbb{R},\sigma^2>0} \mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \sup_{\mu\in\mathbb{R},\sigma^2>0} \mathcal{L}^\flat_{R_\tau}\left(\mu,\sigma^2 \mid X_{t\in R_\tau}\right) \\ &- \sup_{\mu\in\mathbb{R},\sigma^2>0}\Big\{\mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}^\flat_{R_\tau}\left(\mu+\hat{\mu}_a,\sigma^2+\hat{\sigma}^2_a \mid X_{t\in R_\tau}\right)\Big\}, \end{aligned} \quad (2.21)$$

where $\hat{\sigma}^2_a = \hat{\sigma}^2_{R_\tau} - \hat{\sigma}^2_{L_\tau}$ and $\hat{\mu}_a = \bar{x}_{R_\tau} - \bar{x}_{L_\tau}$ are the additive bias corrections.

As discussed in the context of (2.11), we have $\left\{\bar{x}^*_j, \hat{\sigma}^{*2}_j\right\} = \arg\max_{\mu\in\mathbb{R},\sigma^2>0} \mathcal{L}^\flat_j\left(\mu,\sigma^2 \mid X_{t\in j}\right)$, for $j = \{L_\tau, R_\tau\}$. Hence, as in the test for homogeneity in variance, the first two summands in (2.21) are given by

$$\begin{aligned} \sup_{\mu\in\mathbb{R},\sigma^2>0} \mathcal{L}^\flat_j\left(\mu,\sigma^2 \mid X_{t\in j}\right) = &-\frac{n_j}{2}\log(2\pi) \\ &- \frac{n_j}{2}\log\hat{\sigma}^{*2}_j - \frac{n_j}{2}. \end{aligned} \quad (2.22)$$

The third term in (2.21) is now more complicated. To tackle it, introduce the mean-corrected right-hand sample $\tilde{x}_{t\in R_\tau} = \{x_t - \hat{\mu}_a\}_{t\in R_\tau}$, and let $\hat{\tilde{\sigma}}^{*2}_{R_\tau}$ and $\bar{\tilde{x}}^*_{R_\tau}$ be the ML estimators of the respective subsample in the bootstrap world. Maximizing (2.11) with respect to μ, we obtain

$$\mu(\sigma^2) = \frac{n_{L_\tau}\left(\sigma^2+\hat{\sigma}^2_a\right)\bar{x}^*_{L_\tau} + n_{R_\tau}\sigma^2\bar{\tilde{x}}^*_{R_\tau}}{n_{L\tau}\left(\sigma^2+\hat{\sigma}^2_a\right) + n_{R_\tau}\sigma^2}. \quad (2.23)$$

Note that $\mu(\sigma^2)$ is a function in σ^2.

Inserting (2.23) into (2.11), we can reformulate the third term in (2.21) involving only σ^2. This yields

$$\begin{aligned} f^\flat_{CHa}(\sigma^2) = &\mathcal{L}^\flat_{L_\tau}\left\{\mu(\sigma^2),\sigma^2 \mid X_{t\in L_\tau}\right\} \\ &+ \mathcal{L}^\flat_{R_\tau}\left\{\mu(\sigma^2)+\hat{\mu}_a,\sigma^2+\hat{\sigma}^2_a \mid X_{t\in R_\tau}\right\} \\ = &-\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{L_\tau}}{2}\log\sigma^2 - \frac{n_{L_\tau}}{2}\frac{\hat{\sigma}^{*2}_{L_\tau}}{\sigma^2} \\ &- \frac{n_{R_\tau}}{2}\log\left(\sigma^2+\hat{\sigma}^2_a\right) - \frac{n_{R_\tau}}{2}\frac{\hat{\tilde{\sigma}}^{*2}_{R_\tau}}{\sigma^2+\hat{\sigma}^2_a} \\ &- \frac{n_{L_\tau}n^2_{R_\tau}}{2}\frac{\sigma^2\left(\bar{x}^*_{L_\tau} - \bar{\tilde{x}}^*_{R_\tau}\right)^2}{\left[n_{L_\tau}\left(\sigma^2+\hat{\sigma}^2_a\right) + n_{R_\tau}\sigma^2\right]^2} \\ &- \frac{n_{R_\tau}n^2_{L_\tau}}{2}\frac{\left(\sigma^2+\hat{\sigma}^2_a\right)\left(\bar{\tilde{x}}^*_{R_\tau} - \bar{x}^*_{L_\tau}\right)^2}{\left[n_{L_\tau}\left(\sigma^2+\hat{\sigma}^2_a\right) + n_{R_\tau}\sigma^2\right]^2}. \end{aligned} \quad (2.24)$$

In Appendix A.2, we maximize $f^\flat_{CHa}$, which leads to a quintic polynomial in σ^2. Our simulations indicate that the quintic polynomial's discriminant is positive with adequately large samples, implying a single real root. If the discriminant is negative during a bootstrap draw, we redraw u_t until we have the desired number of bootstrap samples. We denote the solution by $\hat{\sigma}^{\flat*2}_{CHa}$.

In summary, the bootstrapped test statistic is given by

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &-\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{\mathcal{I}}}{2} - \frac{n_{L_\tau}}{2}\log\hat{\sigma}^{*2}_{L_\tau} - \frac{n_{R_\tau}}{2}\log\hat{\sigma}^{*2}_{R_\tau} \\ &- f^\flat_{CHa}(\hat{\sigma}^{\flat*2}_{CHa}). \end{aligned} \quad (2.25)$$

Multiplicative bias correction. Alternatively, we propose using a multiplicative bias correction for the variance, which results in a more efficient procedure. Nonetheless, we continue to use the additive correction for the mean. Thus, we consider

$$\begin{aligned} T^\flat_{\mathcal{I},\tau} = &\sup_{\mu\in\mathbb{R},\sigma^2>0} \mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \sup_{\mu\in\mathbb{R},\sigma^2>0} \mathcal{L}^\flat_{R_\tau}\left(\mu,\sigma^2 \mid X_{t\in R_\tau}\right) \\ &- \sup_{\mu\in\mathbb{R},\sigma^2>0}\Big\{\mathcal{L}^\flat_{L_\tau}\left(\mu,\sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}^\flat_{R_\tau}\left(\mu+\hat{\mu}_m,\sigma^2\hat{\sigma}^2_m \mid X_{t\in R_\tau}\right)\Big\}, \end{aligned} \quad (2.26)$$

where $\hat{\sigma}_m = \hat{\sigma}^2_{R\tau}/\hat{\sigma}^2_{L_\tau}$ is the proposed multiplicative bias correction for the variance, while $\hat{\mu}_a = \bar{x}_{R_\tau} - \bar{x}_{L_\tau}$ is the additive correction for the mean, used already in (2.21).

We have $\left\{\bar{x}^*_j, \hat{\sigma}^{*2}_j\right\} = \arg\max_{\mu\in\mathbb{R},\sigma^2>0} \mathcal{L}^\flat_j\left(\mu,\sigma^2 \mid X_{t\in j}\right)$ for $j = \{L_\tau, R_\tau\}$ as before. As a benefit of the multiplicative bias correction for the variance, we can explicitly solve the maximization problem implied by the third term in (2.26). This yields

$$\hat{\mu} = \frac{n_{L_\tau}\hat{\sigma}^2_{R_\tau}\bar{x}^*_{L_\tau} + n_{R_\tau}\hat{\sigma}^2_{L_\tau}\bar{\tilde{x}}^*_{R_\tau}}{n_{L_\tau}\hat{\sigma}^2_{R_\tau} + n_{R_\tau}\hat{\sigma}^2_{L_\tau}}, \quad (2.27)$$

which can be viewed as a weighted estimator for the mean, with weights applied across the two subsamples.

Together with (2.27), we receive a maximization problem solely in σ^2:

$$\begin{aligned} f^{\flat}_{CHm}(\sigma^2) =& \mathcal{L}^{\flat}_{L_\tau}\left(\hat{\mu}, \sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}^{\flat}_{R_\tau}\left(\hat{\mu}+\hat{\mu}_a, \sigma^2\hat{\sigma}^2_m \mid X_{t\in R_\tau}\right) \\ =& -\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{L_\tau}}{2}\log\sigma^2 - \frac{n_{L_\tau}}{2}\frac{\hat{\sigma}^{*2}_{L_\tau}}{\sigma^2} \\ &- \frac{n_{R_\tau}}{2}\log\left(\sigma^2\hat{\sigma}^2_m\right) - \frac{n_{R_\tau}}{2}\frac{\hat{\bar{\sigma}}^{*2}_{R_\tau}}{\sigma^2\hat{\sigma}^2_m} \\ &- \frac{n_{L_\tau}n^2_{R_\tau}}{2}\frac{\sigma^2\left(\bar{x}^*_{L_\tau} - \bar{\bar{x}}^*_{R_\tau}\right)^2}{\left[n_{L_\tau}\left(\sigma^2\hat{\sigma}^2_m\right) + n_{R_\tau}\sigma^2\right]^2} \\ &- \frac{n_{R_\tau}n^2_{L_\tau}}{2}\frac{\left(\sigma^2\hat{\sigma}^2_m\right)\left(\bar{\bar{x}}^*_{R_\tau} - \bar{x}^*_{L_\tau}\right)^2}{\left[n_{L_\tau}\left(\sigma^2\hat{\sigma}^2_m\right) + n_{R_\tau}\sigma^2\right]^2}. \end{aligned} \tag{2.28}$$

Let $n_{L_\tau}\hat{\sigma}^{**2}_{L_\tau} = \sum_{t\in L_\tau}\left[u_t\left(x_t-\hat{\mu}\right)^2\right]$ and $n_{R_\tau}\hat{\bar{\sigma}}^{**2}_{R_\tau} = \sum_{t\in R_\tau}\left[u_t\left(\tilde{x}_t-\hat{\mu}\right)^2\right]$ be the sum of the weighted squared deviations from $\hat{\mu}$, where we recall that $\tilde{x}_{t\in R_\tau} = \{x_t - \hat{\mu}_a\}_{t\in R_\tau}$ stands for the mean-corrected sample.

Surprisingly, (2.28) has a unique maximizer in σ^2 given by

$$\hat{\sigma}^{\flat *2}_{CHm} = \frac{1}{n_{\mathcal{I}}}\left(n_{L_\tau}\hat{\sigma}^{**2}_{L_\tau} + n_{R_\tau}\hat{\bar{\sigma}}^{**2}_{R_\tau}\frac{\hat{\sigma}^2_{L_\tau}}{\hat{\sigma}^2_{R_\tau}}\right). \tag{2.29}$$

Therefore, in the bootstrap context, the variance estimator resembles (2.18) seen in the homogeneity in variance test. It represents as a pooled variance estimator, but corrects the bias in variance, accommodates the bias-induced mean correction of the right-hand sample, and incorporates the weighted mean from (2.27). Consequently, by utilizing closed-form estimates, the bootstrap with multiplicative bias correction secures a substantial computational advantage compared to the case with additive bias correction, which requires numerical solutions (see Appendix A.2 for details).

In summary, the bootstrapped test statistic for complete homogeneity reads as

$$\begin{aligned} T^{\flat}_{\mathcal{I},\tau} =& -\frac{n_{\mathcal{I}}}{2}\log(2\pi) - \frac{n_{\mathcal{I}}}{2} - \frac{n_{L_\tau}}{2}\log\hat{\sigma}^{*2}_{L_\tau} - \frac{n_{R_\tau}}{2}\log\hat{\sigma}^{*2}_{R_\tau} \\ &- f^{\flat}_{CHm}\left(\hat{\sigma}^{\flat *2}_{CHm}\right). \end{aligned} \tag{2.30}$$

2.5 Local Change Point Detection

Central to LCP detection is the notion that there exists an interval, called a 'homogeneous interval' or an 'interval of homogeneity', on which the process is governed by a fixed probability measure P_0. In line with this understanding, for a given point in time t, one tries to find—backward looking—the longest interval of homogeneity for local estimation [15, 20, 2, 24, 13].

To find the longest possible homogeneous interval, we sequentially test for homogeneity on $K+1$ nested intervals, where $\mathcal{I}_0 \subset \mathcal{I}_1 \ldots \subset \mathcal{I}_K$. We choose the interval length to increase arithmetically, i.e., $N_k = \lceil N_0 + ck \rceil$, where $c \in \mathbb{N}$, with c representing the increment per iteration. Denoting the most recent index by t (point of estimation), we set $\mathcal{I}_k = [t - N_k, t]$, for $k = 0, \ldots, K$ and $k \in \mathbb{N}$. The initial interval $\mathcal{I}_0$ is assumed to be homogeneous, which requires its length N_0 to be sufficiently large for reliable estimation. At the same time, N_0 should not be so long that the homogeneity assumption becomes implausible. The collection of nested intervals is referred to as the grid $\mathcal{G}$.

Figure 1 illustrates the intervals involved in the LCP procedure for a given iteration k. In more detail, for each k, every candidate point τ of the interval $\mathcal{J}_k = \mathcal{I}_k \setminus \mathcal{I}_{k-1}$ is regarded as a potential break point. In other words, we test on $\mathcal{I}_k$ for homogeneity against the alternative of a break-point in $\mathcal{I}_k$ at an unknown point $\tau \in \mathcal{J}_k \subset \mathcal{I}_k$. Define the two intervals $L_{k,\tau} = [t - N_{k+1}, \tau]$ and $R_{k,\tau} = (\tau, t]$, which are subintervals of $\mathcal{I}_{k+1}$. Given k, we evaluate either (2.15) and (2.19), or (2.25) and (2.30) on $L_{k,\tau}$ and $R_{k,\tau}$, for all $\tau \in \mathcal{J}_k$, depending on the test case as well as the version of bias correction employed.

The kth interval is rejected if $T_{\mathcal{I}_k} = \max_{\tau\in\mathcal{J}_k} T^{(k)}_\tau > \mathfrak{z}_{\mathcal{I}_k,\alpha}$, where $\mathfrak{z}_{\mathcal{I}_k,\alpha}$ is found using the multiplier bootstrap as described in Section 2.2. Note that to reduce the computational costs, one can run τ only on a subset of $\mathcal{J}_k$, e.g., on every second observation.

If homogeneity cannot be rejected, we increment k and continue with $\mathcal{I}_{k+1}$ until a rejection occurs, or the largest possible interval $\mathcal{I}_{K-1}$ is accepted. As the final estimate of the adaptive estimation procedure, we choose

$$\hat{\theta} := \hat{\theta}_{\mathcal{I}_{\hat{k}}}, \qquad \hat{k} = \max\{k \le K-1 : T_{\mathcal{I}_k} \le \mathfrak{z}^{\flat}_{\mathcal{I}_k,\alpha}\}, \tag{2.31}$$

where $\mathcal{I}_{\hat{k}}$ corresponds to the largest non-rejected interval of homogeneity with $\hat{\theta}_{\mathcal{I}_{\hat{k}}}$ as its estimate. The procedure is repeated for all t in the sample, such that $N_K < t \le T$.

3. SIMULATIONS

In this section, we investigate the size and power of the homogeneity tests under consideration, as well as their ability to perform local change point detection, utilizing both bias corrections. We examine data-generating processes (DGPs) with unconditional heteroskedasticity. Given that most empirical research now employs models of conditional heteroskedasticity, we also apply our tests within that framework. Furthermore, we account for deviations from the strict *i.i.d.* framework by allowing for light serial dependence in the DGP.

We sample the multiplier bootstrap weights $\{u_t\}$ from a Poisson distribution with $\lambda = 1$. We also experimented with alternative distributions, including an exponential distribution and a mixture of uniform distributions. As the results are very similar, we only report those associated with the Poisson distribution. As mentioned in Section 2, we always normalize the weights such that the empirical mean and

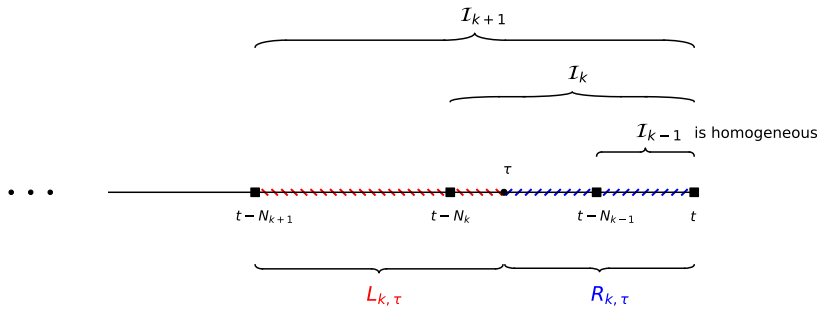


Figure 1: Intervals in the sequential LCP detection procedure. For a given iteration k, we test for parameter homogeneity in $\mathcal{I}_k$ with length N_k and a fixed end point t. In each iteration, we evaluate a set of candidate change points contained in $\mathcal{J}_k = \mathcal{I}_k \setminus \mathcal{I}_{k-1}$; the test decision is based on the maximum statistic over all available candidate change points.

variance are equal to one. We do this to reduce the variance of the bootstrap estimates and to lower size distortions. The number of bootstrap samples is $B = 1000$.

3.1 Evaluation of the Multiplier Bootstrap Tests

Simulation framework. To evaluate the size and power of the proposed homogeneity tests, we consider two DGPs. In case (i), we sample from a normal distribution $\mathcal{N}(0,1)$; in case (ii), we draw from a standardized $t(5)$-distribution. The standardization helps maintain comparability across both DGPs. Case (i) represents our baseline setting, where the parametric assumptions are correctly specified, while case (ii) represents a case, where the model is misspecified. We benchmark our tests against the χ^2-test with one (homogeneity in variance) or two degrees of freedom (complete homogeneity). We expect similar results in case (i), especially for moderately large sample sizes, whereas the inference based on the χ^2-distribution will be invalid in case (ii).

We consider three setups with sample sizes $n_{L_\tau} = n_{R_\tau} = \{5, 25, 50\}$, which further results in total interval lengths of $n_{\mathcal{I}} = \{10, 50, 100\}$. The tests are evaluated on three significance levels, $\alpha = \{0.025, 0.05, 0.1\}$. To compute the coverage probabilities for assessing the size and power of the test, we repeat each test case $M = 1000$ times across all scenarios.

To study the power, we use a parameter $\lambda \in (-0.8, 2)$ to control the magnitude of a breakpoint. We start by assessing the power of each test given a change in mean, where the samples are generated from:

$$X_t = \begin{cases} \mathcal{Z}, & t \in L_\tau \\ \mathcal{Z} + \lambda, & t \in R_\tau \end{cases},$$

where $\mathcal{Z}$ is drawn independently either from (i) $\mathcal{N}(0,1)$ or (ii) a standardized $t(5)$-distribution. H_0 is obtained for $\lambda = 0$.

To examine power, when there is a break in the variance, the samples are generated according to:

$$X_t = \begin{cases} \mathcal{Z}, & t \in L_\tau \\ \sqrt{1+\lambda}\mathcal{Z}, & t \in R_\tau \end{cases}.$$

When generating the alternative by multiplying by $\sqrt{1+\lambda}$, note that the sample mean inflates. To ensure comparability across samples in terms of their empirical means, we first demean each sample, apply the multiplication, and then add back its empirical sample mean. This prevents artificial inflation of the empirical means as we range over λ. When $\lambda = 0$, we simulate under H_0.

Size. Figure 2 illustrates the size discrepancies, which represent the differences between the actual coverage probabilities and theoretical target values (type I error). We find very similar results for the χ^2-test relative to the multiplier bootstrap, when we draw $\mathcal{N}(0,1)$ sequences, no matter which bias correction is used. For standardized $t(5)$ sequences, the χ^2-test fails. In comparing the two bias corrections, we observe that the additive version performs slightly more accurately under model misspecification, given a moderate sample size $n_{\mathcal{I}} \in \{50, 100\}$. However, the multiplicative correction is markedly superior in the extreme case of $n_{\mathcal{I}} = 10$. Generally, the discrepancies in size are larger under misspecification, as one may expect. Overall, the results indicate that the multiplicative bias correction is a suitable substitute while providing significant computational advantages as pointed out in Sections 2.3 and 2.4.

Power: break in μ. In Figure 3, we investigate breaks in the mean parameter. Of course, the test of homogeneity in variance is expected to display a constant coverage probability, independent of $\lambda \in (-0.8, 2)$, which we confirm. In contrast, the test for complete homogeneity detects the changes in the mean, and panels (c) and (d) exhibit the power (type II error). As with size, power decreases under model misspecification and smaller sample sizes. Comparing both bias correction methods, we find similar power across both settings, except for small samples, where the multiplicative correction is superior.

Power: break in σ^2. For changes in the variance parameter, the power is illustrated in Figure 4. The results suggest that the test for homogeneity in variance consistently achieves a marginally better performance in terms of power than the test for complete homogeneity across both

DGPs. This is evidenced by lower frequencies of type II errors across all test configurations, including bias correction, sample size, and significance level. Once again, lower power is observed under the misspecified model and with small sample sizes. Consistent with previous findings, the simulations suggest that the multiplicative bias correction is a suitable alternative, particularly in tiny samples.

3.2 Local Change Point Detection

Unconditional changes in the mean and variance. We continue to consider two different underlying families of distributions for drawing segments of non-stationary sequences: (*i*) *i.i.d.* $\mathcal{N}(0,1)$ sequences and (*ii*) *i.i.d.* standardized $t(5)$ ones. We draw a cross-section of $M = 1000$ processes in both settings, comprising 1000 realizations each. The processes are modeled as follows:

$$X_t = \begin{cases} \mathcal{Z}, & 1 \le t \le 340 \\ \mathcal{Z} + 0.75, & 341 \le t \le 670 \\ \sqrt{1+1.5}(\mathcal{Z}+0.75), & 671 \le t \le 1000 \end{cases},$$

where $\mathcal{Z}$ is drawn as discussed above. Furthermore, as described in Section 3.1, we ensure a pure change in variance scenario for $t > 670$, consistent with the mean-preserving transformation outlined earlier.

We employ $\mathcal{G} = \{50, 100, 150, 200, 250, 300\}$ as the grid of interval sizes for the sequential LCP detection algorithm. To ease the computational burden, we let τ run on every second observation, which doubles the speed of the simulations. When testing, we consider a significance level of $\alpha = 0.025$. Figure 5 presents the results for each test configuration. At each time point t, for a given test and DGP, the plot displays the cross-sectional mean and median of the estimated index $\hat{k}$, computed using (2.31), where the corresponding interval length is given by $\mathcal{I}_{\hat{k}}$.

The test for homogeneity in variance excels the one for complete homogeneity in detecting changes in the second moment. Moreover, under model misspecification, the tests incorporating the multiplicative adjustment detect changes in more instances in the cross-section of simulated data sequences. For both DGPs, in the case of testing for homogeneity in variance, we find a clustering of false rejections due to changes in the mean. The reasons for these false alarms are heterogeneous subintervals: Whenever a subinterval (left or right) is rolled over a change point in the mean, the estimated variance in the respective subinterval can be affected, which leads to rejections of the hypotheses. Notably, the additive bias correction mitigates this behavior in the presence of misspecification while providing lower detection power.

Because empirical practice often raises concerns about potential misspecification regarding serial dependence, we also consider sequences generated by the autoregressive moving average (ARMA) process

$$X_t = \begin{cases} \phi_1 X_{t-1} + \gamma_1 \varepsilon_{t-1} + \varepsilon_t, \\ \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_1^2), \quad 1 \le t \le 340 \\ \phi_1 X_{t-1} + \gamma_1 \varepsilon_{t-1} + \varepsilon_t, \\ \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_2^2), \quad 341 \le t \le 670 \\ \phi_2 X_{t-1} + \gamma_2 \varepsilon_{t-1} + \varepsilon_t, \\ \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_2^2), \quad 671 \le t \le 1000 \end{cases}$$

where we have $\phi_1 = 0.2$, $\gamma_1 = 0.2$ and $\sigma_1^2 = 1$ initially, and $\phi_2 = 0.6$, $\gamma_2 = 0.6$ and $\sigma_2^2 = 2.5$ after the change. This setup enables us to examine two distinct jumps in the unconditional variance of the ARMA$(1,1)$ sequences: one driven by changes in the residual variance and the other caused by shifts in the ARMA coefficients. Notably, changes in γ and ϕ do not affect the first moments.

We use the LCP configuration described above, including the values for $\mathcal{G}$, τ, and α. Figure 6 presents the results for both tests. Again, at each time point t, the plot shows the cross-sectional mean and median of the estimated index $\hat{k}$.

In line with previous findings on variance changes, our results demonstrate that the test for homogeneity in variance exhibits superior detection power compared to the test for complete homogeneity. Moreover, it offers an additional advantage by showing greater robustness to time-dependent dynamics in the mean. This is particularly evident in the final segment, where we have increased magnitudes for both ARMA coefficients, ϕ and γ. In such cases, the test for complete homogeneity frequently produces rejections, whereas the test for homogeneity in variance more effectively mitigates the time-variation in the conditional mean.

Conditional changes in the variance. So far, we have examined the tests' ability to detect abrupt transitions in the DGPs, including both independent and serially dependent sequences. Next, we evaluate the performance of our tests in capturing conditional variance. A widely used model for this purpose is the generalized autoregressive conditional heteroskedasticity (GARCH) model, which allows the conditional variance to evolve over time as a function of past errors and past conditional variances [1]. This exercise compares the LCP detection algorithm, which uses data-driven interval lengths, with a naive rolling window approach employing a fixed window width for variance estimation, under a GARCH process. Rolling window estimators for variance are commonly used when no specific model is available.

We sample $M = 1000$ GARCH(1,1) processes for six different parametric configurations. Each process comprises 1000 realizations and is given by:

$$X_t = \sigma_t \varepsilon_t, \tag{3.1}$$

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{3.2}$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$. We set $\omega = 10^{-5}$ in each configuration. To work with meaningful magnitudes for α and β, we

Table 1. Summary of GARCH(1,1) fits, derived by their closeness to the minimum (Min), first quartile (Q1), median (Median), third quartile (Q3), and maximum (Max) across the 503 persistence parameters (ρ). Besides the coefficients, we also display prevailing ticker symbols as well as the number of observations comprised by the fit. The underlying data is sourced from Yahoo Finance.

	α	β	ρ
Min	0.0623	0.7390	0.8013
Q1	0.1223	0.8548	0.9771
Median	0.0600	0.9287	0.9887
Q3	0.0896	0.9054	0.9950
Max	0.0197	0.9795	0.9992

collect daily closing prices for each 503 constituents of the S&P 500 as of May 3, 2023, where the earliest starting date is January 1, 1970. We fit the GARCH(1,1) model to the log return sequences of each asset and record the estimated coefficients—see Table 1 for the descriptive statistics on the estimates of α and β across all 503 time series.

The persistence of the conditional variance process is $\rho = \alpha + \beta$. If $\rho < 1$, the process is strictly stationary [6]. As reported in Table 1, U.S. large-cap equity returns exhibit notable persistence. For our simulations, we select five configurations from the 503 candidates based on their proximity to the minimum (Min), first quartile (Q1), median (Median), third quartile (Q3), and maximum (Max) of all persistence parameters estimated.

We 'forecast' the 1-day ahead conditional variance of the GARCH(1,1) sequence by estimating the realized variance over varying interval lengths. We examine the forecasting accuracy in terms of the mean squared forecasting error (MSFE) and calculate this measure for (i) a locally adaptive estimation procedure incorporating the proposed test for homogeneity in variance; and (ii) various rolling window estimators relying on a fixed interval length. We use $\mathcal{G} = \{25, 50, 75, 100, 125, 150\}$ as the grid, run τ on each observation in $\mathcal{J}_k = \mathcal{I}_k \backslash \mathcal{I}_{k-1}$, and set a significance level of $\alpha = 0.025$ for testing.

In Figure 7, we present the ratios of the MSFE for the LCP algorithm relative to a specific fixed window estimator. A ratio below one indicates that the LCP method outperforms the fixed window approach. As is visible, there is no clear relationship between the persistence of the process and the predictive performance of the models considered. Longer interval lengths improve accuracy in environments with either very low or very high persistence ("Min" and "Max"), while shorter ones excel in "moderate" persistence settings ("Q1", "Median" and "Q3"). Interestingly, in the "Max" scenario, which features the highest persistence and the largest magnitude of β, neither the estimators with the smallest $(k = 1)$ nor those with the largest window lengths $(k = 5)$ rank among the most accurate predictive models. Although our locally adaptive approach never outperforms all benchmarks (based on the median, it beats at least three out of five benchmarks in four out of five cases), no benchmark constantly outperforms the LCP approach across all scenarios. Because persistence in practice is unknown or may even be subject to structural breaks, these findings underscore the benefit of locally adaptive estimation over fixed-window methods.

4. APPLICATIONS

We evaluate the ability of our test for homogeneity in variance to detect potential breaks in economic and financial time series data. Specifically, we examine (i) the month-over-month log inflation in the U.S.; (ii) the monthly log growth of U.S. industrial production; and (iii) the daily log returns from Bitcoin (BTC). Regarding BTC, we also study change point detection in the mean and the variance of the logarithm of the absolute values of daily returns, following ideas in [23].

In line with our simulation studies, we employ Poisson weights in the multiplier bootstrap, set the bootstrap samples to $B = 1000$ and the significance level to $\alpha = 0.025$. Moreover, throughout all procedures, we let τ run on each observation and employ $\mathcal{G} = \{25, 50, 75, 100, 125, 150\}$ as the grid.

4.1 U.S. Month-over-Month Inflation Rates

In recent years, the variance of consumer prices changes has made price stability a central topic. Therefore, our first analysis focuses on month-over-month inflation rates in the U.S., computed from the seasonally adjusted core consumer price index (CPI_t), excluding energy and food. The data spans from January 1957 to September 2023, comprising 802 observations.

We fit ARIMA(p, d, q) models to monthly log inflation rates r_t and extract the residuals ε_t. Both the original series and the residuals are treated as data sequences for the application of our LCP procedure. Model selection based on various information criteria (AIC, AICc, and BIC) suggests an ARIMA$(0, 1, 1)$ specification. The presence of a unit root is further supported by an augmented Dickey–Fuller test.

Figure 8 displays the results. For both sequences (r_t and ε_t), we present the estimated index $\hat{k}_t$ and the corresponding locally adaptive estimates for the log variance. The LCP identifies a break in the late 70s and early 80s for both series, associated with high inflation levels. Since then, the variance has appeared fairly constant. Additionally, we observe a break in variance between 2016 and 2018, coinciding with a reduction in variance. Another break in the variance is detected around the start of the COVID-19 pandemic. Comparing the results across both sequences, the LCP procedure on r_t identifies an additional break in variance at the beginning of our sample. The results demonstrate the

effectiveness of our LCP detection procedure, whether applied to the raw signal or the residuals of the ARIMA fit, in identifying changes in the variance of month-over-month inflation rates.

4.2 Monthly Growth Rates in U.S. Industrial Production

The Great Moderation, characterized by reduced variance in the growth rates of industrial production from the mid-1980s until at least 2007, is a widely discussed phenomenon in macroeconomic literature [22]. In our second analysis, we apply the LCP to monthly growth rates of U.S. industrial production (monthly seasonally adjusted index INDPRO), covering the period from February 1919 to September 2023, with 1,257 observations. We calculate monthly log growth rates, again denoted here by r_t, and also extract, as previously, the residuals from an ARIMA$(2, 0, 3)$ model, following a same search methodology as above.

Figure 9 presents the results for both data sequences, which are qualitatively very similar. The LCP procedure identifies numerous structural breaks in the variance, particularly before 1960. After 1965, another break in variance is observed, followed by a prolonged period of homogeneity with historically low variance lasting until the mid-80s. At this point, the procedure detects a new change. The subsequent period, characterized by even lower variance, extends until 2007. Thus, the locally adaptive approach not only delivers results that align precisely with the Great Moderation but, remarkably, also identifies the period from 1970 to the mid-1980s as a time of relatively low variance.

4.3 Daily Bitcoin Returns

In our last application, we analyze the historical variance of Bitcoin (BTC) using daily closing prices from January 1, 2016, to November 30, 2023 (2,891 observations). For this purpose, we calculate the daily log returns r_t of BTC.

Figure 10 presents the outcomes for r_t. Initially, we observe a high number of breaks in variance in the first half of the sample, whereas breaks become considerably less frequent in the latter half. Several variance breaks are detected during 2017/18, coinciding with sharp price fluctuations in BTC. Additional variance breaks are noted in 2018/19. Further, the LCP procedure indicates a growing period of homogeneity until the rapid price decline triggered by the recession induced by the COVID-19 pandemic. Since 2022, the variance has been fairly constant. As Figure 10 shows, breaks in variance tend to occur during extreme price movements, both positive and negative. The fewer change points over time possibly indicate the maturation of the cryptocurrency market.

Figure 11 provides another perspective, focusing on the variance of BTC and its volatility. To this end, we compute $\log|r_t|$ as a (rough) measure of daily volatility. Consequently, the mean estimator, applied to $\{\log|r_t|\}$ becomes an estimator of variance, while the variance estimate is interpretable as a measure of volatility of variance. Therefore, it is now of interest to apply both tests, the test for homogeneity in variance as well as the test for complete homogeneity.

The comparison of the estimated intervals of homogeneity offers telling insights: Until mid-2017, our procedures detected changes only in the mean (here meaning: the variance), as the test for homogeneity in variance (here meaning: volatility of variance) does not identify any breaks. Likewise, another period of instability occurred in 2019. These findings align well with the previous analysis of the returns. Besides these two periods, the sequences of the estimated intervals of homogeneity are quite similar, suggesting that changes in the mean (here: variance) might also correspond to breaks in the variance (here: volatility of variance), see for instance, at the end of 2017, the beginning of 2020 (COVID-19), and the end of 2022. When comparing the results across both characteristics, we see that the volatility of variance tends to have fewer structural breaks than the variance. Overall, these observations suggest that a thorough understanding of both the variance and the volatility of variance is essential for effectively modeling the rapidly evolving BTC market.

5. CONCLUSION

We propose a locally adaptive testing procedure for homogeneity in variance (allowing for an arbitrary mean) and for complete homogeneity, where both the mean and variance are tested jointly. This procedure utilizes a maximum statistic derived from a generalized likelihood ratio test, relying on critical values from a multiplier bootstrap.

To enhance the efficacy of the simulation, we introduce a multiplicative bias correction in the bootstrap. Unlike the traditional additive correction, this multiplicative approach not only offers a more intuitive interpretation for the bootstrap estimators, but also reduces computational costs by allowing all necessary quantities for the bootstrap to be computed in closed form. Our simulations demonstrate that tests incorporating the multiplicative bias correction exhibit size and power properties comparable to those using the additive correction. Additional simulations show that the locally adaptive testing procedure competes well with rolling window estimators based on a fixed window length, particularly when the persistence of variance is unknown.

We apply the procedure to several data sequences to detect breaks in the variance in the growth rates of inflation and industrial production, as well as in cryptocurrency returns. The results suggest that this LCP procedure is a valuable addition to the empiricist's toolkit for analyzing breaks in cases of unconditional heteroskedasticity. Ongoing research will adapt the methodology to the multivariate use case.

APPENDIX A

A.1 Homogeneity in Variance: Additive Bias Correction

As shown in (2.14), using the additive bias correction, the third term in (2.12) results in a function of σ^2, which we denote by $f^{\flat}_{HVa}\left(\sigma^2\right)$. Taking the derivative with respect to σ^2 and setting the resulting expression equal to zero, we obtain

$$0 = -\frac{n_{L_\tau}}{\sigma^2} - \frac{n_{R_\tau}}{(\sigma^2+\hat{\sigma}_a^2)} + \frac{n_{L_\tau}\hat{\sigma}^{*2}_{L_\tau}}{(\sigma^2)^2} + \frac{n_{R_\tau}\hat{\sigma}^{*2}_{R_\tau}}{(\sigma^2+\hat{\sigma}_a^2)^2},$$

which can be simplified to the following cubic polynomial in σ^2:

$$0 = a\sigma^6 + b\sigma^4 + c\sigma^2 + d,$$

where

$$\begin{aligned}
a &= -n_{\mathcal{I}},\\
b &= n_{L_\tau}\hat{\sigma}^{*2}_{L_\tau} - 2n_{L_\tau}\hat{\sigma}_a^2 - n_{R_\tau}\hat{\sigma}_a^2 + n_{R_\tau}\hat{\sigma}^{*2}_{R_\tau},\\
c &= n_{L_\tau}\hat{\sigma}_a^2(2\hat{\sigma}^{*2}_{L_\tau} - \hat{\sigma}_a^2),\\
d &= n_{L_\tau}\hat{\sigma}^{*2}_{L_\tau}\hat{\sigma}_a^4.
\end{aligned}$$

As discussed in Section 2.3, we find the corresponding real root ($\hat{\sigma}^{\flat *2}_{HVa}$) numerically. Again, our simulations show that for a sufficiently large sample size, the discriminant of the cubic polynomial is negative, which translates into a single real root.

A.2 Complete Homogeneity: Additive Bias Correction

Testing for complete homogeneity with the additive bias correction, the third term in (2.21) is a function denoted by $f^{\flat}_{CHa}\left(\sigma^2\right)$. On differentiating wrt σ^2, we find

$$\begin{aligned}
0 = &-\frac{n_{L_\tau}}{\sigma^2} - \frac{n_{R_\tau}}{(\sigma^2+\hat{\sigma}_a^2)} + \frac{n_{L_\tau}\hat{\sigma}^{*2}_{L_\tau}}{(\sigma^2)^2} + \frac{n_{R_\tau}\hat{\tilde{\sigma}}^{*2}_{R_\tau}}{(\sigma^2+\hat{\sigma}_a^2)^2}\\
&- \left[n_{L_\tau}n^2_{R_\tau}(\bar{x}^*_{L_\tau} - \bar{\tilde{x}}^*_{R_\tau})^2\right]\\
&\left\{\frac{1}{[n_{L_\tau}(\sigma^2+\hat{\sigma}_a^2)+n_{R_\tau}\sigma^2]^2}\right.\\
&\left. - \frac{2\sigma^2(n_{L_\tau}+n_{R_\tau})}{[n_{L_\tau}(\sigma^2+\hat{\sigma}_a^2)+n_{R_\tau}\sigma^2]^3}\right\}\\
&- \left[n_{R_\tau}n^2_{L_\tau}(\bar{x}^*_{L_\tau} - \bar{\tilde{x}}^*_{R_\tau})^2\right]\\
&\left\{\frac{1}{[n_{L_\tau}(\sigma^2+\hat{\sigma}_a^2)+n_{R_\tau}\sigma^2]^2}\right.\\
&\left. - \frac{2(\sigma^2+\hat{\sigma}_a^2)(n_{L_\tau}+n_{R_\tau})}{[n_{L_\tau}(\sigma^2+\hat{\sigma}_a^2)+n_{R_\tau}\sigma^2]^3}\right\},
\end{aligned}$$

which can be cast as the following quintic polynomial in σ^2:

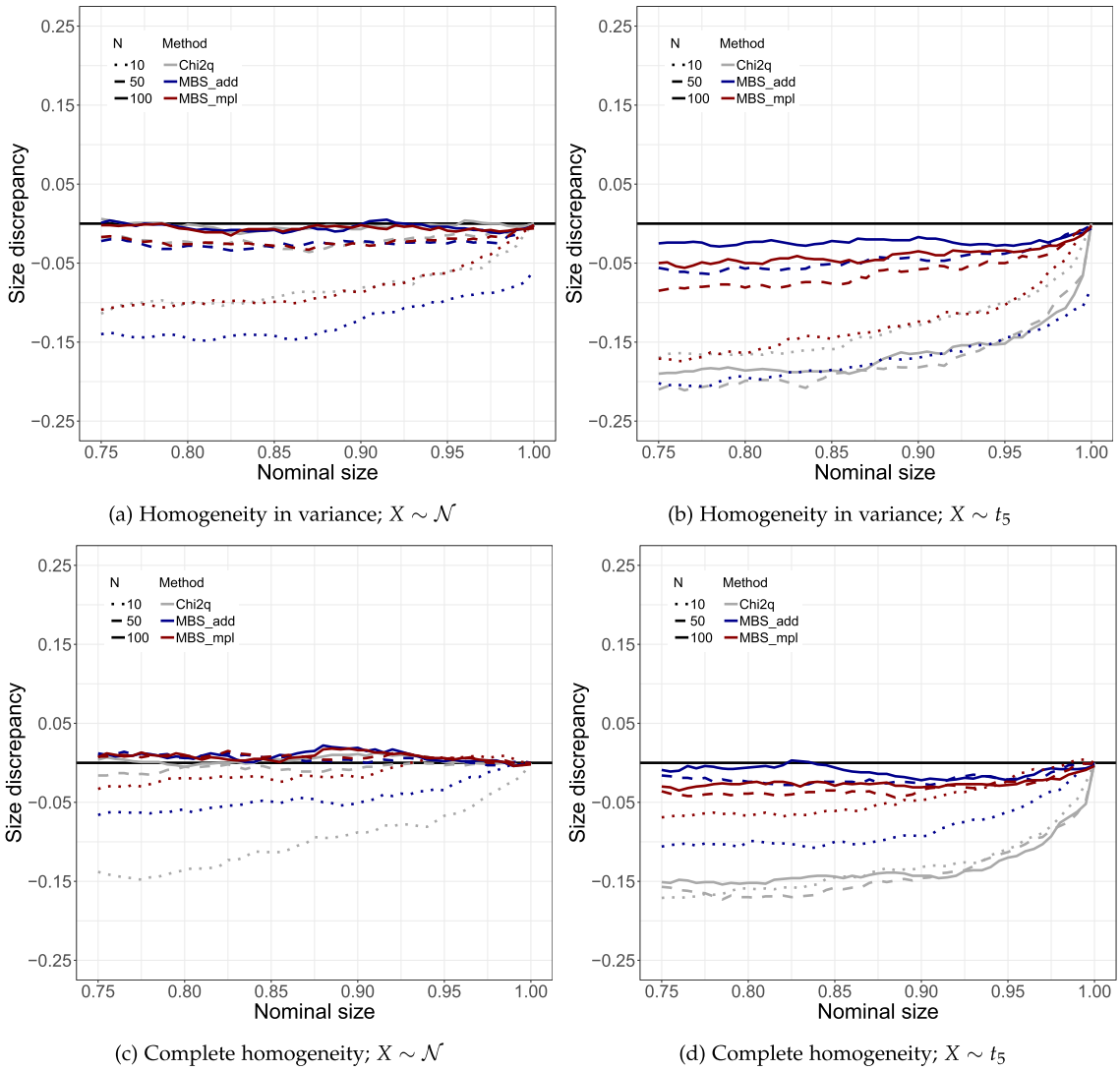
$$0 = a\sigma^{10} + b\sigma^8 + c\sigma^6 + d\sigma^4 + e\sigma^2 + f,$$

where

$$\begin{aligned}
a &= n^3_{\mathcal{I}},\\
b &= n_{\mathcal{I}}\Big\{n^2_{L_\tau}(4\hat{\sigma}_a^2 - \hat{\sigma}^{*2}_{L_\tau}) + n^2_{R_\tau}(\hat{\sigma}_a^2 - \hat{\sigma}^{*2}_{R_\tau})\\
&\quad + n_{L_\tau}n_{R_\tau}\left[-(\bar{x}^*_{L_\tau} - \bar{\tilde{x}}^*_{R_\tau})^2 + 5\hat{\sigma}_a^2 - \hat{\sigma}^{*2}_{L_\tau} - \hat{\sigma}^{*2}_{R_\tau}\right]\Big\},\\
c &= n_{L_\tau}n_{\mathcal{I}}\hat{\sigma}_a^2\Big\{3n_{R_\tau}\hat{\sigma}_a^2 + n_{L_\tau}(6\hat{\sigma}_a^2 - 4\hat{\sigma}^{*2}_{L_\tau})\\
&\quad - 2n_{R_\tau}\left[(\bar{x}^*_{L_\tau} - \bar{\tilde{x}}^*_{R_\tau})^2 + \hat{\sigma}^{*2}_{L_\tau} + \hat{\sigma}^{*2}_{R_\tau}\right]\Big\},\\
d &= n_{L_\tau}\hat{\sigma}_a^4\Big\{n^2_{L_\tau}(4\hat{\sigma}_a^2 - 6\hat{\sigma}^{*2}_{L_\tau}) - n^2_{R_\tau}\left[(\bar{x}^*_{L_\tau} - \bar{\tilde{x}}^*_{R_\tau})^2 + \hat{\sigma}^{*2}_{L_\tau}\right]\\
&\quad + n_{L_\tau}n_{R_\tau}\left[-(\bar{x}^*_{L_\tau} - \bar{\tilde{x}}^*_{R_\tau})^2 + 3\hat{\sigma}_a^2 - 6\hat{\sigma}^{*2}_{L_\tau} - \hat{\sigma}^{*2}_{R_\tau}\right]\Big\},\\
e &= n^2_{L_\tau}\hat{\sigma}_a^6\left[n_{L_\tau}\hat{\sigma}_a^2 - 2\hat{\sigma}^{*2}_{L_\tau}(2n_{L_\tau}+n_{R_\tau})\right],\\
f &= -n^3_{L_\tau}\hat{\sigma}^{*2}_{L_\tau}\hat{\sigma}_a^8.
\end{aligned}$$

As discussed in Section 2.4, we find the corresponding real root ($\hat{\sigma}^{\flat *2}_{CH}$) numerically. Again, our simulations indicate that the quintic polynomial's discriminant for adequately large samples is always positive; hence, a single real root exists.

A.3 Size Discrepancy



(a) Homogeneity in variance; $X \sim \mathcal{N}$

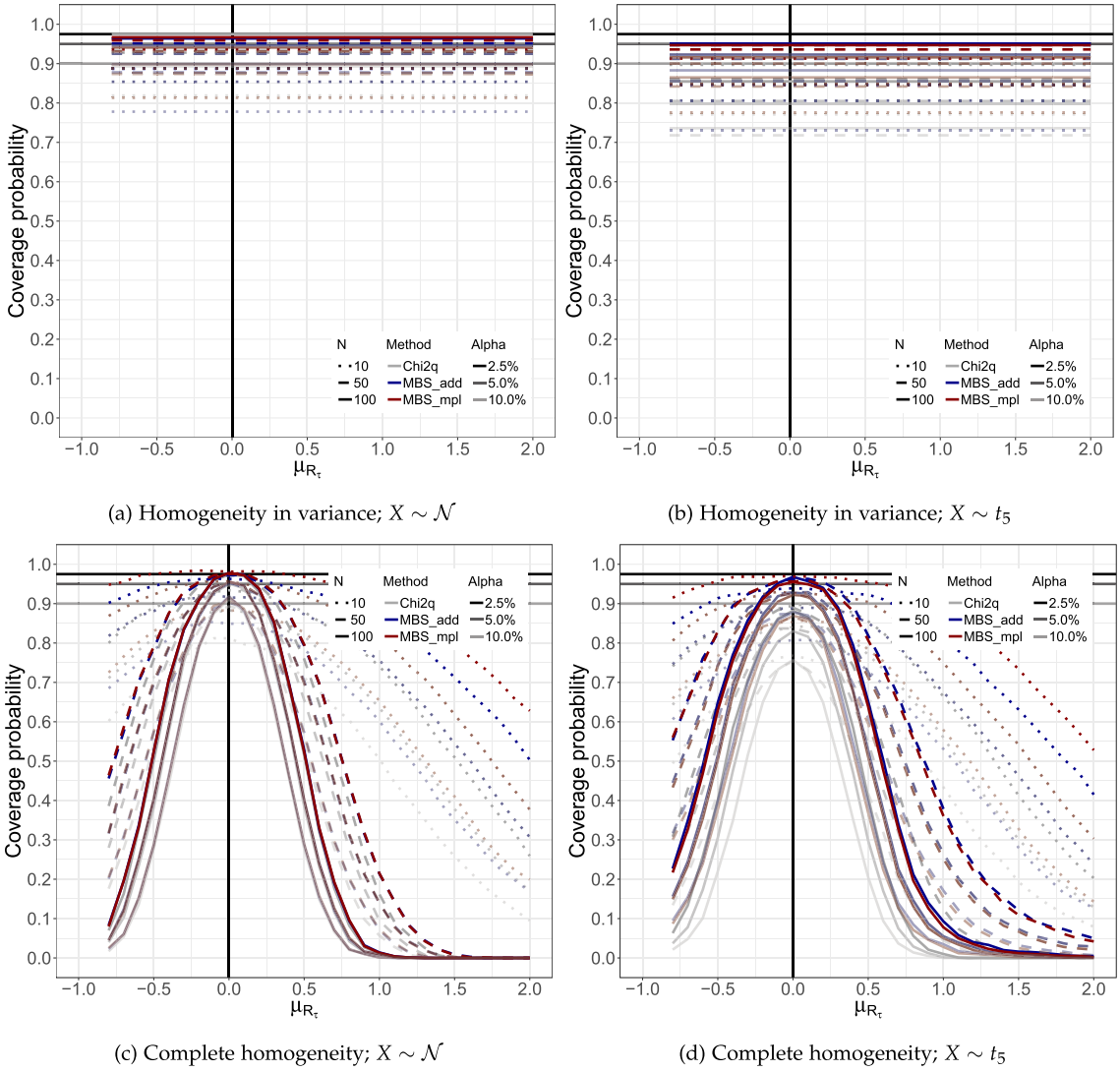
(b) Homogeneity in variance; $X \sim t_5$

(c) Complete homogeneity; $X \sim \mathcal{N}$

(d) Complete homogeneity; $X \sim t_5$

Figure 2: Size discrepancies (under H_0) for both tests and each of the two *i.i.d.* DGPs. Red lines refer to tests utilizing the multiplicative bias correction in the multiplier bootstrap, while blue ones refer to additive versions. Results from the χ^2-test are illustrated by grey lines. Sample sizes are depicted in dotted, dashed, or solid lines for $n_{\mathcal{I}} = \{10, 50, 100\}$, respectively. Configuration of LCP: $M = 1000$, $B = 1000$ and $u \sim Poi(1)$.

A.4 Power: Break in μ



(a) Homogeneity in variance; $X \sim \mathcal{N}$

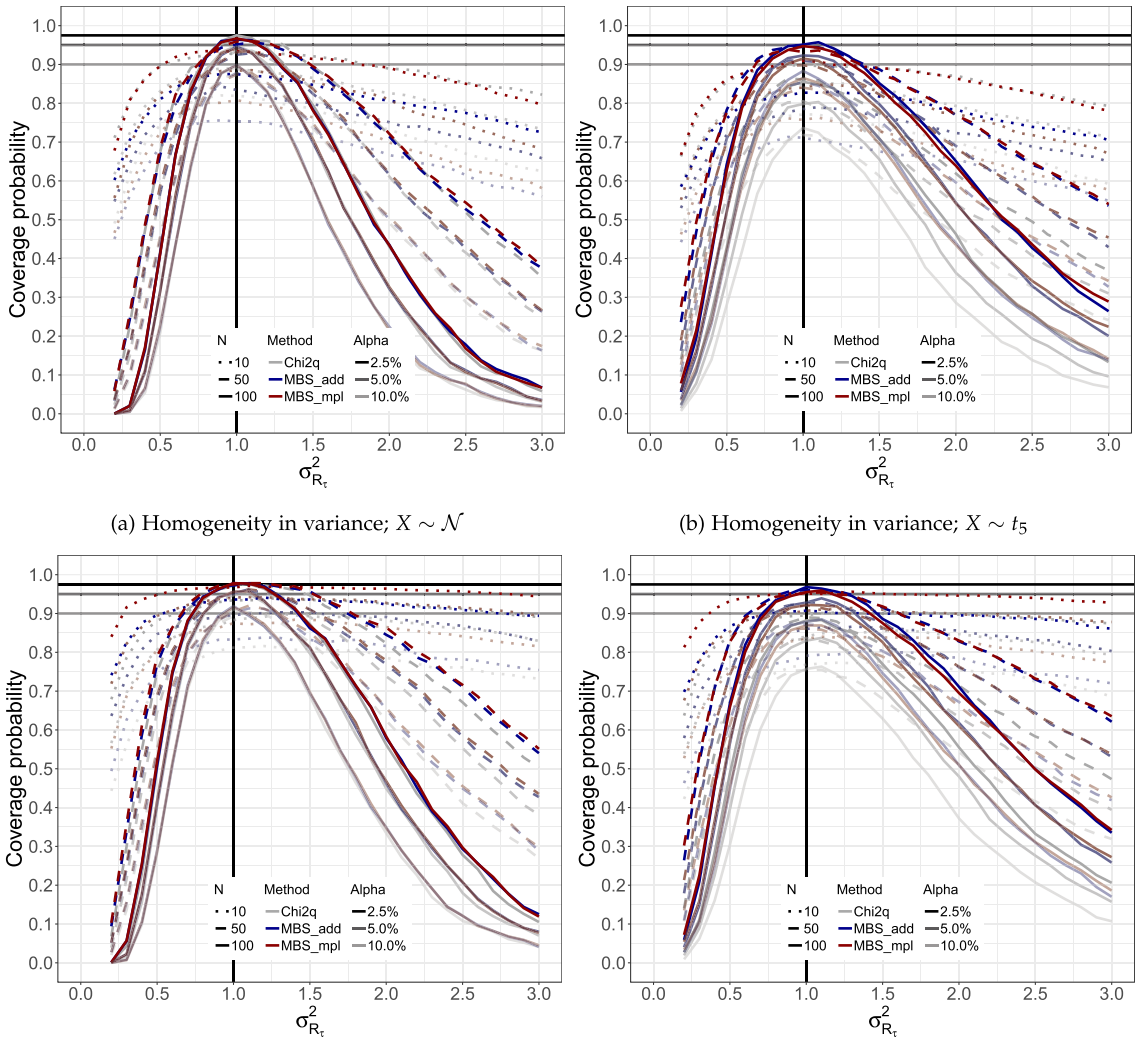
(b) Homogeneity in variance; $X \sim t_5$

(c) Complete homogeneity; $X \sim \mathcal{N}$

(d) Complete homogeneity; $X \sim t_5$

Figure 3: Local power plots in the scenario of a change in mean for both tests and each of the two *i.i.d.* DGPs. Red lines refer to tests utilizing the multiplicative bias correction in the multiplier bootstrap, while blue ones refer to additive versions. Results from the χ^2-test are illustrated by grey lines. Sample sizes are depicted in dotted, dashed, or solid lines for $n_{\mathcal{I}} = \{10, 50, 100\}$, respectively. Configuration of LCP: $M = 1000$, $B = 1000$ and $u \sim Poi(1)$.

A.5 Power: Break in σ^2



(a) Homogeneity in variance; $X \sim \mathcal{N}$

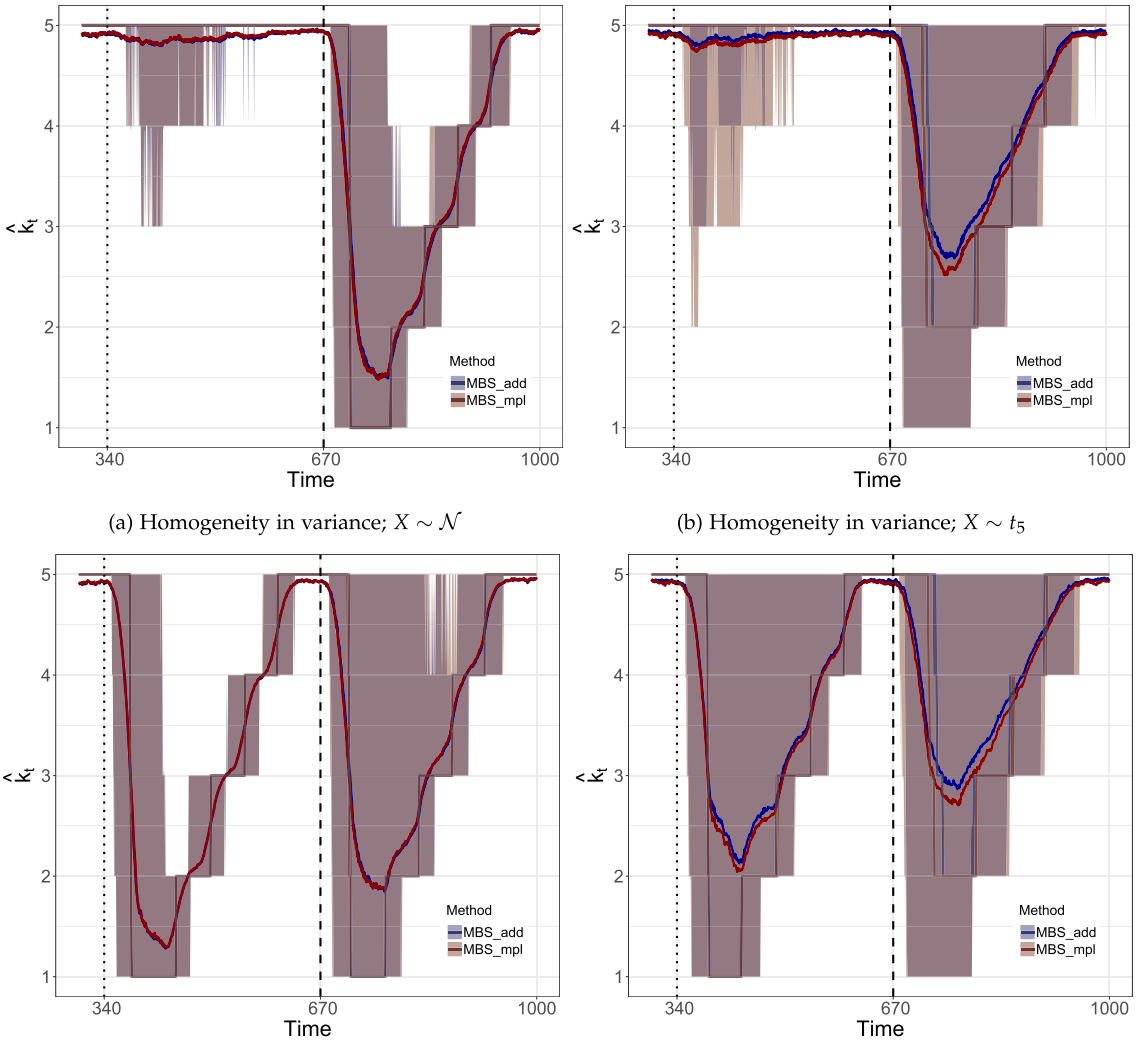
(b) Homogeneity in variance; $X \sim t_5$

(c) Complete homogeneity; $X \sim \mathcal{N}$

(d) Complete homogeneity; $X \sim t_5$

Figure 4: Local power plots in the scenario of a change in variance for both tests and each of the two *i.i.d.* DGPs. Red lines refer to tests utilizing the multiplicative bias correction in the multiplier bootstrap, while blue ones refer to additive versions. Results from the χ^2-test are illustrated by grey lines. Sample sizes are depicted in dotted, dashed, or solid lines for $n_{\mathcal{I}} = \{10, 50, 100\}$, respectively. Configuration of LCP: $M = 1000$, $B = 1000$ and $u \sim Poi(1)$.

A.6 LCP: Unconditional Changes in Parameters



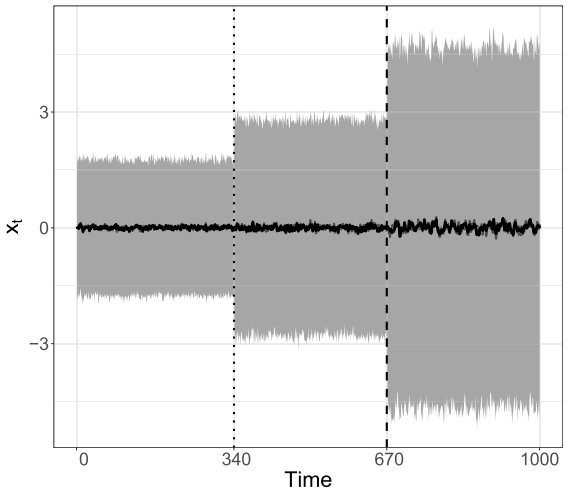
(a) Homogeneity in variance; $X \sim \mathcal{N}$

(b) Homogeneity in variance; $X \sim t_5$

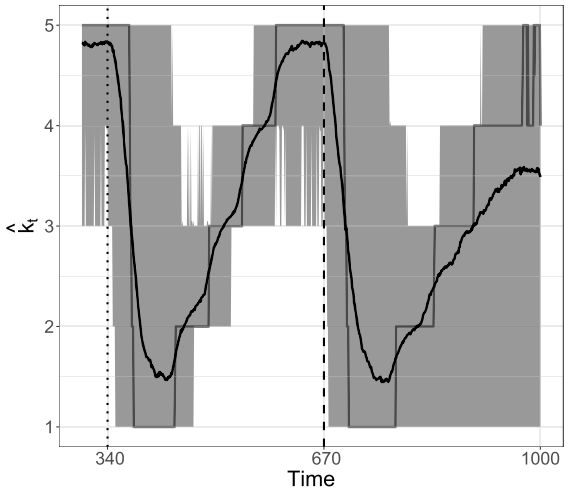
(c) Complete homogeneity; $X \sim \mathcal{N}$

(d) Complete homogeneity; $X \sim t_5$

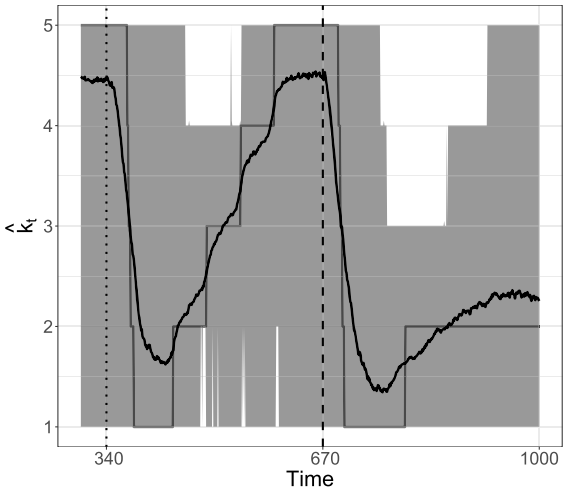
Figure 5: Summary of local change point detection procedures for both tests and each of the two *i.i.d.* DGPs. Red lines refer to tests utilizing the multiplicative bias correction in the multiplier bootstrap, while blue ones refer to additive versions. The dotted and dashed vertical line either depicts a change in the mean or variance. Dark and light colors refer to cross-sectional means and medians. The shaded area is bounded by the cross-sectional 0.05 and 0.95 percentile. Configuration: $u \sim Poi(1)$, $M = 1000$, $B = 1000$, $\alpha = 0.025$, $\mathcal{G} = \{50, 100, 150, 200, 250, 300\}$, $\tau = 2$.



(a) Realizations



(b) Homogeneity in variance



(c) Complete homogeneity

Figure 6: Summary of local change point detection procedures for both tests for DGPs of the ARMA type, using the multiplicative bias correction in the multiplier bootstrap. The dotted and dashed vertical line either depicts a change in the residual variance or both ARMA(1, 1) coefficients, both scenarios affecting the unconditional variance. Dark and light colors refer to cross-sectional means and medians. The shaded area is bounded by the cross-sectional 0.05 and 0.95 percentile. Configuration: $u \sim Poi(1)$, $M = 1000$, $B = 1000$, $\alpha = 0.025$, $\mathcal{G} = \{50, 100, 150, 200, 250, 300\}$, $\tau = 2$.

A.7 LCP: Conditional Changes in Variance

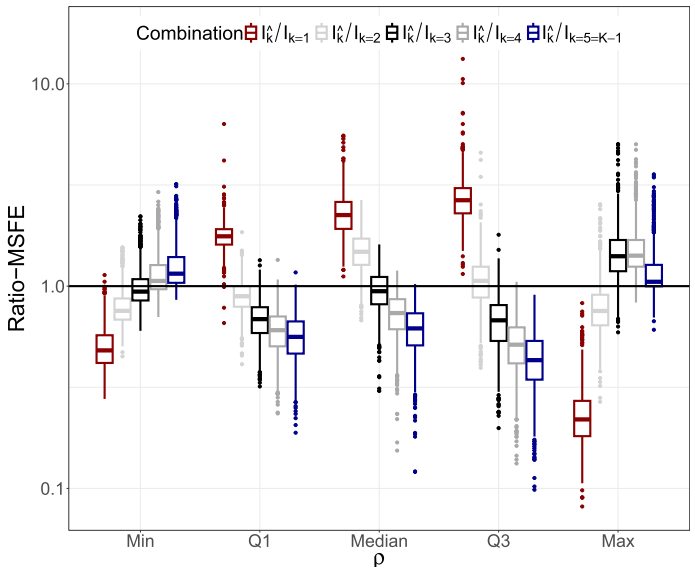


Figure 7: Summary of forecasting performances for each considered GARCH(1,1) configuration (see also Table 1), which were selected to span a range of persistence levels ρ, from minimum to maximum, as detailed in the main text. The scatter plots depict ratios in mean squared forecasting errors (Ratio-MSFE) between the locally adaptive estimator in the numerator and the naive benchmarks in the denominator. The set of naive benchmarks comprises rolling window estimators relying on a fixed interval length, which we derive from $\mathcal{G}$ excluding its last element. For example, $\mathcal{I}_{k=3}$ refers to the third element in $\mathcal{G}$, corresponding to a sample of length 75. Configuration: $u \sim Poi(1)$, $M = 1000$, $B = 1000$, $\alpha = 0.025$, $\mathcal{G} = \{25, 50, 75, 100, 125, 150\}$, $\tau = 1$.

A.8 U.S. Month-over-Month Inflation Rates

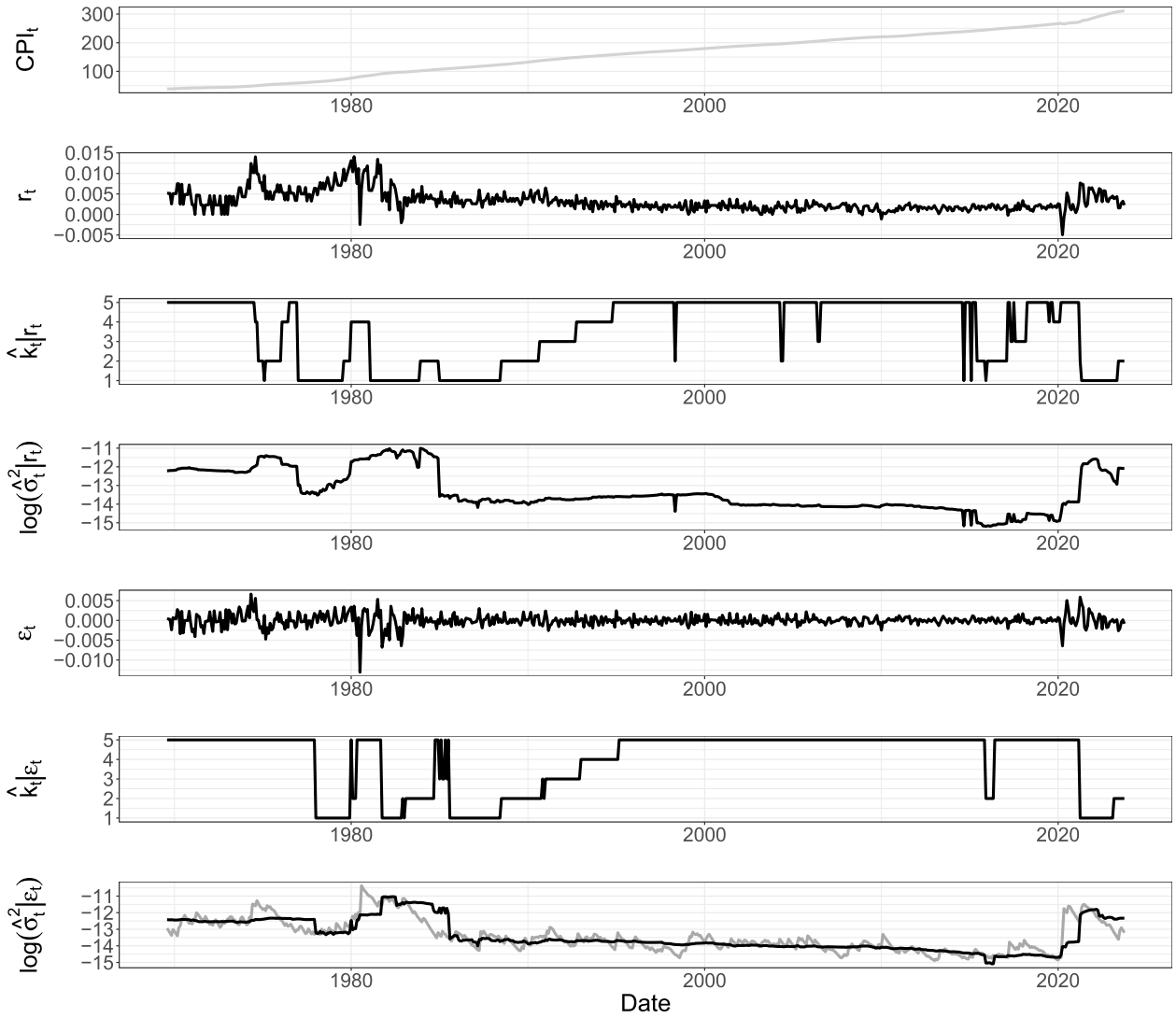


Figure 8: The first panel presents monthly values for the seasonally adjusted U.S. core consumer price index, sourced via the FRED database. The second and fifth panels show the monthly log inflation rates and the residuals filtered from an ARIMA$(0, 1, 1)$ model. The third and sixth panels depict the estimated index $\hat{k}$ of the interval of homogeneity at t for both signals individually. The fourth and seventh panels illustrate the respective log variance estimates at each point t, for (i) our locally adaptive approach (in black) and (ii) an in-sample fitted GARCH$(1, 1)$ model in the case of ε_t (in grey). Configuration: $u \sim Poi(1)$, $B = 1000$, $\alpha = 0.025$, $\mathcal{G} = \{25, 50, 75, 100, 125, 150\}$, $\tau = 1$.

A.9 Monthly Growth Rates in U.S. Industrial Production

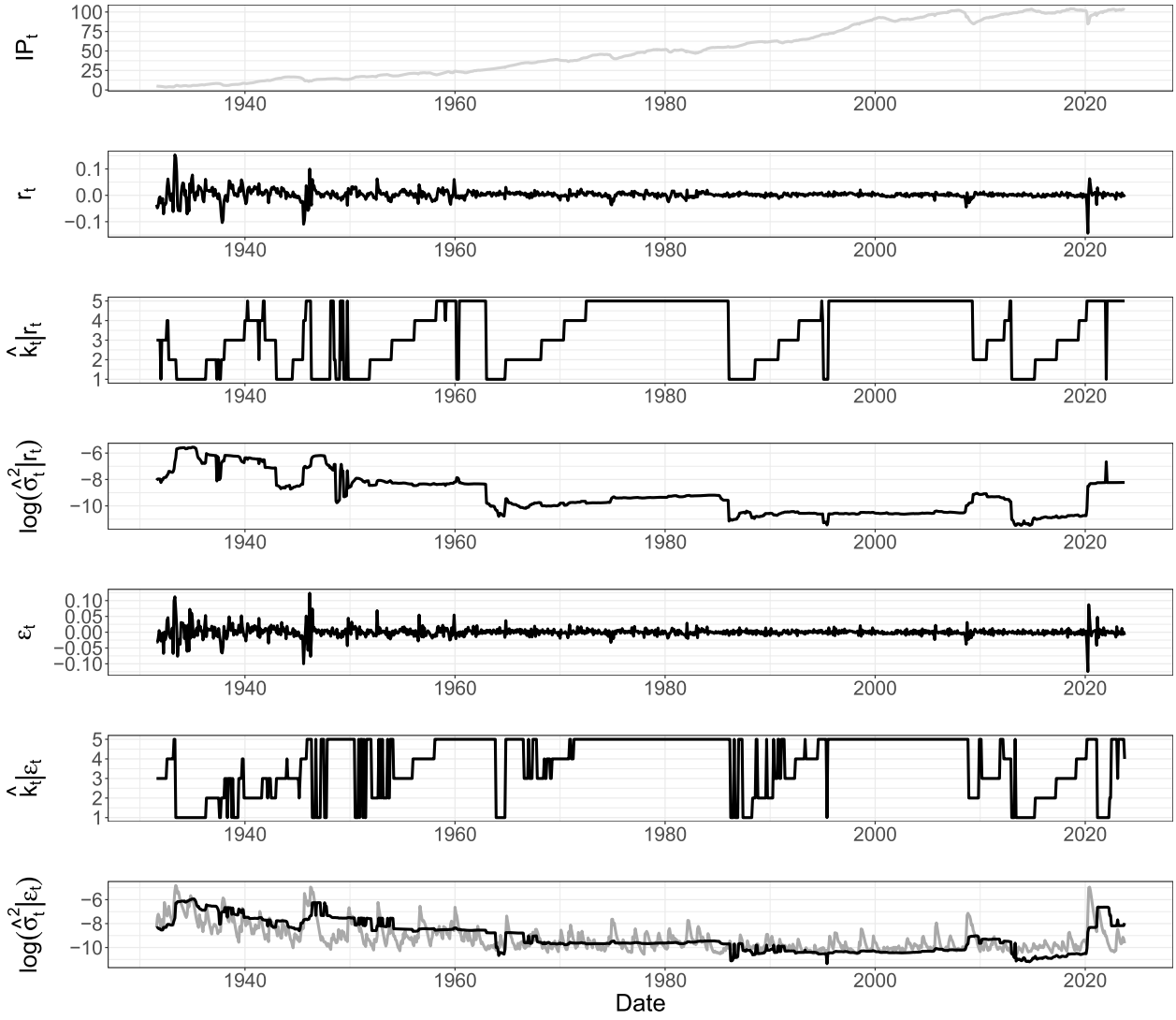


Figure 9: The first panel presents monthly values for the seasonally adjusted U.S. industrial production index, sourced via the FRED database. The second and fifth panels show the log growth rates and the residuals filtered from an ARIMA$(2, 0, 3)$ model. The third and sixth panels depict the estimated index $\hat{k}$ of the interval of homogeneity at t for both signals individually. The fourth and seventh panels illustrate the respective log variance estimates at each point t, for (i) our locally adaptive approach (in black) and (ii) an in-sample fitted GARCH$(1, 1)$ model in the case of ε_t (in grey). Configuration: $u \sim Poi(1)$, $B = 1000$, $\alpha = 0.025$, $\mathcal{G} = \{25, 50, 75, 100, 125, 150\}$, $\tau = 1$.

A.10 Daily Bitcoin Returns

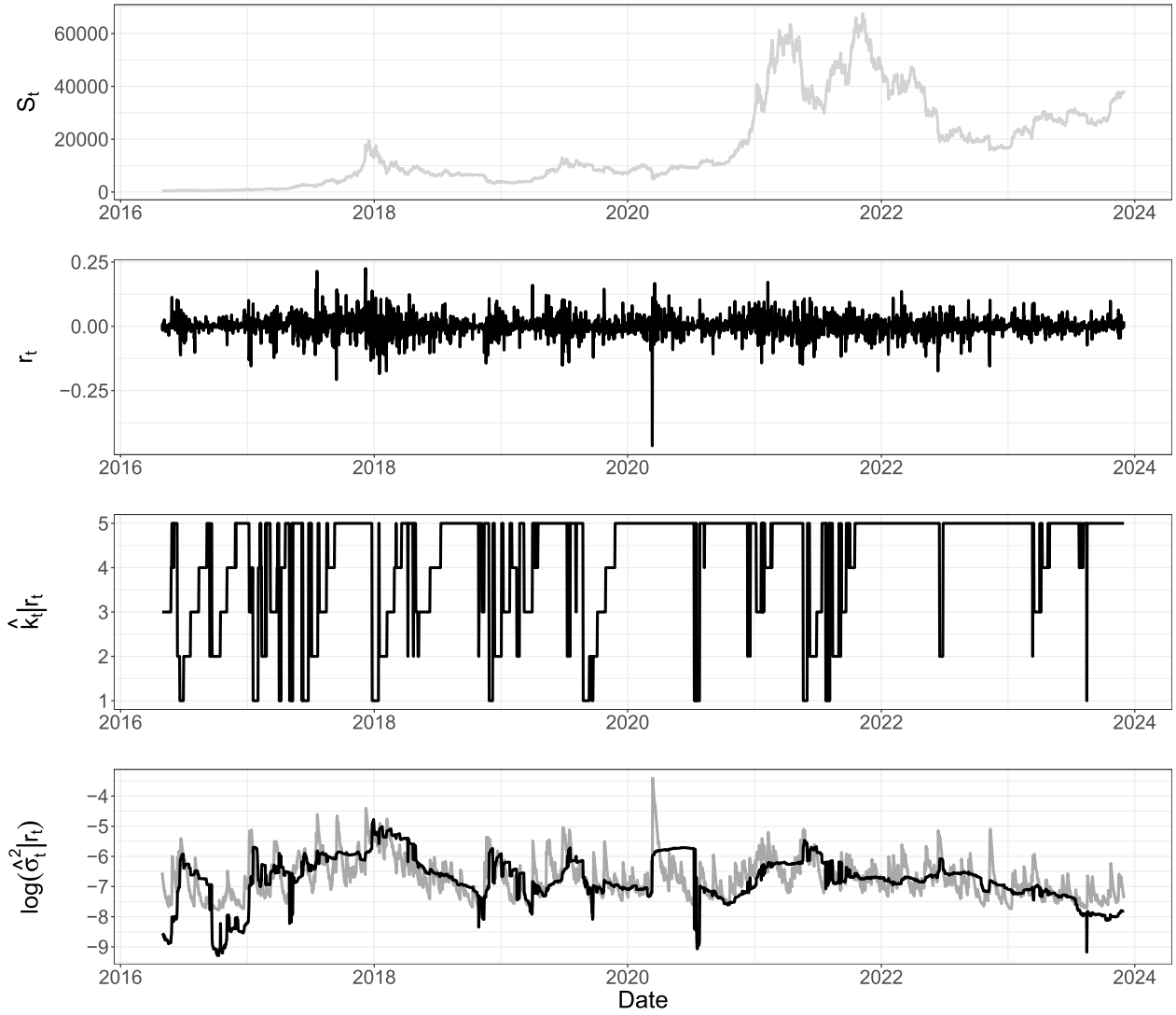


Figure 10: The first panel presents daily closing prices for Bitcoin, sourced via Yahoo Finance. The second panel depicts daily log returns r_t, and the third one shows the estimated index $\hat{k}$ of the interval of homogeneity at t. The fourth panel illustrates the log variance estimates at each point t, for (i) our locally adaptive approach (in black) and (ii) an in-sample fitted GARCH$(1,1)$ model (in grey). Configuration: $u \sim Poi(1)$, $B = 1000$, $\alpha = 0.025$, $\mathcal{G} = \{25, 50, 75, 100, 125, 150\}$, $\tau = 1$.

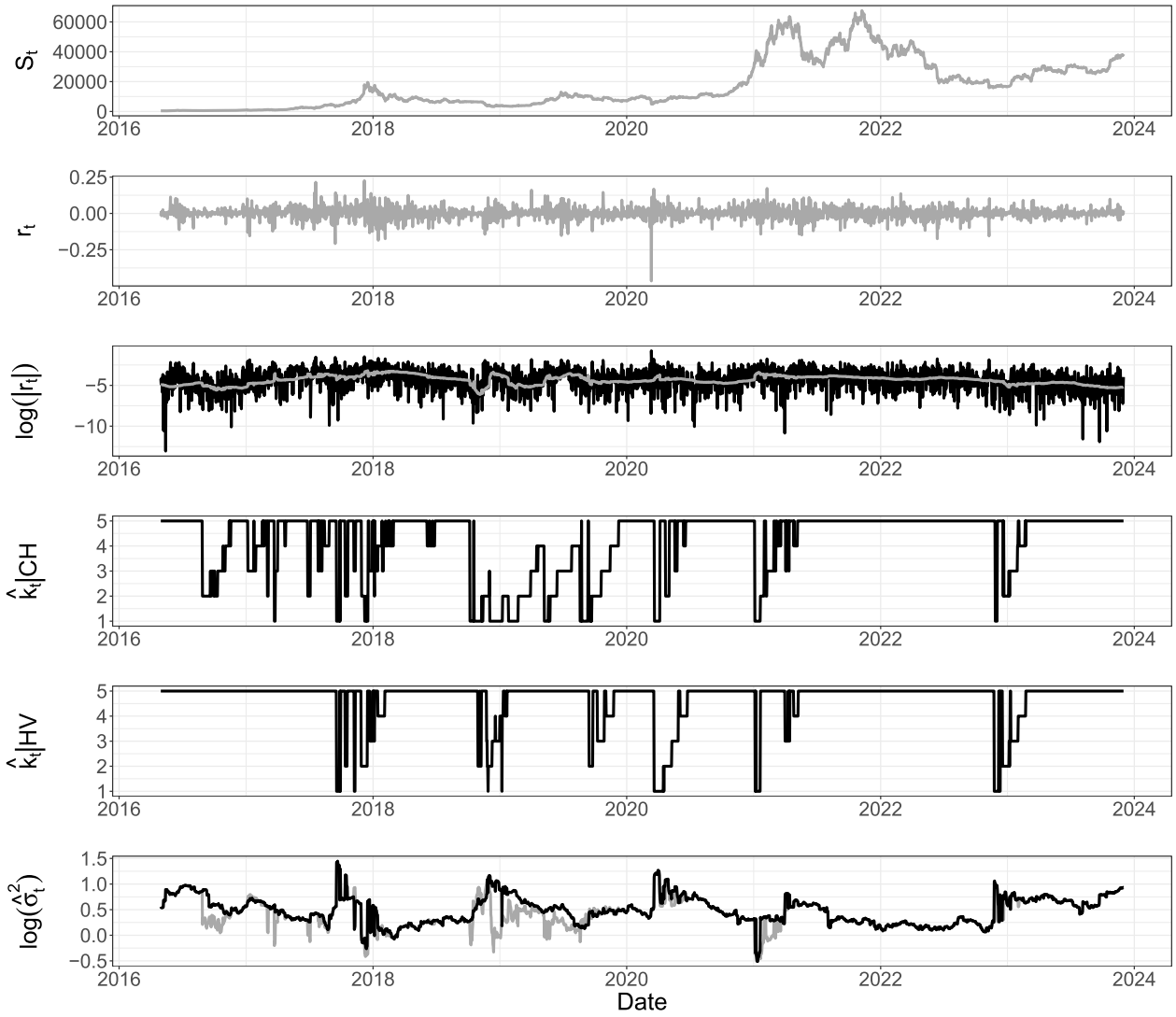


Figure 11: The first and second panels present daily closing prices for Bitcoin and the daily log returns r_t, sourced via Yahoo Finance. The third panel shows the logarithm of the absolute values of the returns $\log(|r_t|)$, on which we perform the LCP procedures. We add the mean estimated in the locally adaptive manner (in grey), relying on our test for complete homogeneity. The fourth and fifth panels exhibit the estimated index $\hat{k}$ of the interval of homogeneity at t, either for (i) complete homogeneity or (ii) for homogeneity in variance, respectively. The sixth panel illustrates locally adaptive estimates of log volatility at each point t (complete homogeneity in grey and homogeneity in variance in black). Configuration: $u \sim Poi(1)$, $B = 1000$, $\alpha = 0.025$, $\mathcal{G} = \{25, 50, 75, 100, 125, 150\}$, $\tau = 1$.

APPENDIX B

B.1 Using Unscaled Weights in the MBS

As discussed in Section 2.3, in our implementation across all test versions, we ensure that $\sum_{t\in j} u_t = n_j$, for $j = L_\tau, R_\tau$. This is achieved by normalizing $\{u_t\}_{t\in j}$ during each MBS iteration. The normalization offers two advantages: (i) it allows us to maintain more compact formulae in the main text; and (ii) it helps maintain lower size discrepancies, especially under misspecification. It must be noted, however, that normalizing introduces linear dependence among the weights, which diminishes the larger the sample.

Without normalization, and in contrast to (2.11), the bootstrapped log-likelihood function is given by:

$$\begin{aligned}\mathcal{L}_j^\flat\left(\mu,\sigma^2 \mid X_{t\in j}\right) =& -\frac{1}{2}\log(2\pi)\sum_{t=1}^{n_j} u_t - \frac{1}{2}\log\sigma^2\sum_{t=1}^{n_j} u_t \\ &- \frac{1}{2\sigma^2}\sum_{t=1}^{n_j} u_t\left(x_t-\mu\right)^2, \\ =& -\frac{\mathcal{U}_j}{2}\log(2\pi) - \frac{\mathcal{U}_j}{2}\log\sigma^2 - \frac{\mathcal{U}_j}{2}\frac{\hat\sigma_j^{*2}}{\sigma^2} \\ &- \frac{\mathcal{U}_j}{2\sigma^2}(\bar x_j^* - \mu)^2, \end{aligned} \tag{B.1}$$

where we have $\mathcal{U}_j = \sum_{t=1}^{n_j} u_t$, $w_t = u_t/\mathcal{U}_j$, $\hat\sigma_j^{*2} = \sum_{t\in j} w_t(x_t - \bar x_j^*)^2$ and $\bar x_j^* = \sum_{t\in j} w_t x_t$, for $j = \{L_\tau, R_\tau\}$. Note that $\sum_{t=1}^{n_j} u_t/n_j \xrightarrow{p} 1$ and $\sum_{t=1}^{n_j} u_t^2/n_j \xrightarrow{p} 1$.

Thus, for $j = \{L_\tau, R_\tau\}$, the first two summands in (2.12) now obtain as

$$\begin{aligned}\sup_{\mu\in\mathbb{R},\,\sigma^2>0} \mathcal{L}_j^\flat\left(\mu,\sigma^2 \mid X_{t\in j}\right) =& -\frac{\mathcal{U}_j}{2}\log(2\pi) \\ &- \frac{\mathcal{U}_j}{2}\log\hat\sigma_j^{*2} - \frac{\mathcal{U}_j}{2}. \end{aligned} \tag{B.2}$$

B.2 Unscaled Weights: Homogeneity in Variance

Additive bias correction. The bootstrapped log-likelihood function over both intervals, without normalization and combined with the additive bias correction, is:

$$\begin{aligned} f_{HVa}^\flat(\sigma^2) =& \mathcal{L}_{L_\tau}^\flat\left(\bar x_{L_\tau}^*, \sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}_{R_\tau}^\flat\left(\bar x_{R_\tau}^*, \sigma^2+\hat\sigma_a^2 \mid X_{t\in R_\tau}\right) \\ =& -\frac{\mathcal{U}_\mathcal{I}}{2}\log(2\pi) - \frac{\mathcal{U}_{L_\tau}}{2}\log\sigma^2 - \frac{\mathcal{U}_{L_\tau}}{2}\frac{\hat\sigma_{L_\tau}^{*2}}{\sigma^2} \\ &- \frac{\mathcal{U}_{R_\tau}}{2}\log\left(\sigma^2+\hat\sigma_a^2\right) - \frac{\mathcal{U}_{R_\tau}}{2}\frac{\hat\sigma_{R_\tau}^{*2}}{\sigma^2+\hat\sigma_a^2}, \end{aligned} \tag{B.3}$$

where we find the maximizer for (B.3) numerically, as discussed in Appendix A.1.

In summary, the bootstrapped test statistic is

$$\begin{aligned} T_{\mathcal{I},\tau}^\flat =& -\frac{\mathcal{U}_\mathcal{I}}{2}\log(2\pi) - \frac{\mathcal{U}_\mathcal{I}}{2} - \frac{\mathcal{U}_{L_\tau}}{2}\log\hat\sigma_{L_\tau}^{*2} - \frac{\mathcal{U}_{R_\tau}}{2}\log\hat\sigma_{R_\tau}^{*2} \\ &- f_{HVa}^\flat\left(\hat\sigma_{HVa}^{\flat *2}\right). \end{aligned} \tag{B.4}$$

Multiplicative bias correction. Using the multiplicative bias correction, we have:

$$\begin{aligned} f_{HVm}^\flat(\sigma^2) =& \mathcal{L}_{L_\tau}^\flat\left(\bar x_{L_\tau}^*, \sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}_{R_\tau}^\flat\left(\bar x_{R_\tau}^*, \sigma^2\hat\sigma_m^2 \mid X_{t\in R_\tau}\right) \\ =& -\frac{\mathcal{U}_\mathcal{I}}{2}\log(2\pi) - \frac{\mathcal{U}_{L_\tau}}{2}\log\sigma^2 - \frac{\mathcal{U}_{L_\tau}}{2}\frac{\hat\sigma_{L_\tau}^{*2}}{\sigma^2} \\ &- \frac{\mathcal{U}_{R_\tau}}{2}\log\left(\sigma^2\hat\sigma_m^2\right) - \frac{\mathcal{U}_{R_\tau}}{2}\frac{\hat\sigma_{R_\tau}^{*2}}{\sigma^2\hat\sigma_m^2}, \end{aligned} \tag{B.5}$$

for which, as in (2.17) and (2.18), we can find the unique maximizer

$$\hat\sigma_{HVm}^{\flat *2} = \frac{1}{\mathcal{U}_\mathcal{I}}\left(\mathcal{U}_{L_\tau}\hat\sigma_{L_\tau}^{*2} + \mathcal{U}_{R_\tau}\hat\sigma_{R_\tau}^{*2}\frac{\hat\sigma_{L_\tau}^2}{\hat\sigma_{R_\tau}^2}\right). \tag{B.6}$$

The bootstrapped test statistic is given by

$$\begin{aligned} T_{\mathcal{I},\tau}^\flat =& -\frac{\mathcal{U}_\mathcal{I}}{2}\log(2\pi) - \frac{\mathcal{U}_\mathcal{I}}{2} - \frac{\mathcal{U}_{L_\tau}}{2}\log\hat\sigma_{L_\tau}^{*2} - \frac{\mathcal{U}_{R_\tau}}{2}\log\hat\sigma_{R_\tau}^{*2} \\ &- f_{HVm}^\flat\left(\hat\sigma_{HVm}^{\flat *2}\right). \end{aligned} \tag{B.7}$$

B.3 Unscaled Weights: Complete Homogeneity

Additive bias correction. We follow the same logic as outlined in the main text. First, we derive μ as a function of σ^2, which equals

$$\mu(\sigma^2) = \frac{\mathcal{U}_{L_\tau}\left(\sigma^2+\hat\sigma_a^2\right)\bar x_{L_\tau}^* + \mathcal{U}_{R_\tau}\sigma^2\bar{\tilde x}_{R_\tau}^*}{\mathcal{U}_{L_\tau}\left(\sigma^2+\hat\sigma_a^2\right) + \mathcal{U}_{R_\tau}\sigma^2}. \tag{B.8}$$

Utilizing (B.8), we find the bootstrapped log-likelihood function over both intervals, using the additive bias corrections; it is

$$\begin{aligned} f_{CHa}^\flat(\sigma^2) =& \mathcal{L}_{L_\tau}^\flat\left\{\mu(\sigma^2), \sigma^2 \mid X_{t\in L_\tau}\right\} \\ &+ \mathcal{L}_{R_\tau}^\flat\left\{\mu(\sigma^2)+\hat\mu_a, \sigma^2+\hat\sigma_a^2 \mid X_{t\in R_\tau}\right\} \\ =& -\frac{\mathcal{U}_\mathcal{I}}{2}\log(2\pi) - \frac{\mathcal{U}_{L_\tau}}{2}\log\sigma^2 - \frac{\mathcal{U}_{L_\tau}}{2}\frac{\hat\sigma_{L_\tau}^{*2}}{\sigma^2} \\ &- \frac{\mathcal{U}_{R_\tau}}{2}\log\left(\sigma^2+\hat\sigma_a^2\right) - \frac{\mathcal{U}_{R_\tau}}{2}\frac{\hat\sigma_{R_\tau}^{*2}}{\sigma^2+\hat\sigma_a^2} \\ &- \frac{\mathcal{U}_{L_\tau}\mathcal{U}_{R_\tau}^2}{2}\frac{\sigma^2\left(\bar x_{L_\tau}^* - \bar{\tilde x}_{R_\tau}^*\right)^2}{\left[\mathcal{U}_{L_\tau}\left(\sigma^2+\hat\sigma_a^2\right) + \mathcal{U}_{R_\tau}\sigma^2\right]^2} \\ &- \frac{\mathcal{U}_{R_\tau}\mathcal{U}_{L_\tau}^2}{2}\frac{\left(\sigma^2+\hat\sigma_a^2\right)\left(\bar{\tilde x}_{R_\tau}^* - \bar x_{L_\tau}^*\right)^2}{\left[\mathcal{U}_{L_\tau}\left(\sigma^2+\hat\sigma_a^2\right) + \mathcal{U}_{R_\tau}\sigma^2\right]^2}. \end{aligned} \tag{B.9}$$

where, as in (2.24), we have $\tilde{x}_{t\in R_\tau} = \{x_t - \hat{\mu}_a\}_{t\in R_\tau}$, with $\hat{\tilde{\sigma}}^{*2}_{R_\tau}$ and $\bar{\tilde{x}}^{*}_{R_\tau}$ as the respective ML estimators. Further, we maximize (B.9) numerically, as discussed in Appendix A.2 in the main document.

This yields the following bootstrapped test statistic:

$$T^{\flat}_{\mathcal{I},\tau} = -\frac{\mathcal{U}_{\mathcal{I}}}{2}\log(2\pi) - \frac{\mathcal{U}_{\mathcal{I}}}{2} - \frac{\mathcal{U}_{L_\tau}}{2}\log\hat{\sigma}^{*2}_{L_\tau} - \frac{\mathcal{U}_{R_\tau}}{2}\log\hat{\sigma}^{*2}_{R_\tau} - f^{\flat}_{CHa}(\hat{\sigma}^{\flat *2}_{CHa}). \tag{B.10}$$

Multiplicative bias correction. Considering the multiplicative bias correction for σ^2, we have

$$\hat{\mu} = \frac{\mathcal{U}_{L_\tau}\hat{\sigma}^2_{R_\tau}\bar{x}^{*}_{L_\tau} + \mathcal{U}_{R_\tau}\hat{\sigma}^2_{L_\tau}\bar{\tilde{x}}^{*}_{R_\tau}}{\mathcal{U}_{L_\tau}\hat{\sigma}^2_{R_\tau} + \mathcal{U}_{R_\tau}\hat{\sigma}^2_{L_\tau}}, \tag{B.11}$$

such that the bootstrapped log-likelihood function over both intervals equals

$$\begin{aligned} f^{\flat}_{CHm}(\sigma^2) =& \mathcal{L}^{\flat}_{L_\tau}\left(\hat{\mu}, \sigma^2 \mid X_{t\in L_\tau}\right) \\ &+ \mathcal{L}^{\flat}_{R_\tau}\left(\hat{\mu} + \hat{\mu}_a, \sigma^2\hat{\sigma}^2_m \mid X_{t\in R_\tau}\right) \\ =& -\frac{\mathcal{U}_{\mathcal{I}}}{2}\log(2\pi) - \frac{\mathcal{U}_{L_\tau}}{2}\log\sigma^2 - \frac{\mathcal{U}_{L_\tau}}{2}\frac{\hat{\sigma}^{*2}_{L_\tau}}{\sigma^2} \\ &- \frac{\mathcal{U}_{R_\tau}}{2}\log\left(\sigma^2\hat{\sigma}^2_m\right) - \frac{\mathcal{U}_{R_\tau}}{2}\frac{\hat{\tilde{\sigma}}^{*2}_{R_\tau}}{\sigma^2\hat{\sigma}^2_m} \\ &- \frac{\mathcal{U}_{L_\tau}\mathcal{U}^2_{R_\tau}}{2}\frac{\sigma^2\left(\bar{x}^{*}_{L_\tau} - \bar{\tilde{x}}^{*}_{R_\tau}\right)^2}{\left[\mathcal{U}_{L_\tau}\left(\sigma^2\hat{\sigma}^2_m\right) + \mathcal{U}_{R_\tau}\sigma^2\right]^2} \\ &- \frac{\mathcal{U}_{R_\tau}\mathcal{U}^2_{L_\tau}}{2}\frac{\left(\sigma^2\hat{\sigma}^2_m\right)\left(\bar{\tilde{x}}^{*}_{R_\tau} - \bar{x}^{*}_{L_\tau}\right)^2}{\left[\mathcal{U}_{L_\tau}\left(\sigma^2\hat{\sigma}^2_m\right) + \mathcal{U}_{R_\tau}\sigma^2\right]^2}. \end{aligned} \tag{B.12}$$

As in (2.28) and (2.29), we find the closed-form unique maximizer of (B.12). It takes the form

$$\hat{\sigma}^{\flat *2}_{CHm} = \frac{1}{\mathcal{U}_{\mathcal{I}}}\left(\mathcal{U}_{L_\tau}\hat{\sigma}^{**2}_{L_\tau} + \mathcal{U}_{R_\tau}\hat{\tilde{\sigma}}^{**2}_{R_\tau}\frac{\hat{\sigma}^2_{L_\tau}}{\hat{\sigma}^2_{R_\tau}}\right). \tag{B.13}$$

Thus, the bootstrapped test statistic is given by

$$T^{\flat}_{\mathcal{I},\tau} = -\frac{\mathcal{U}_{\mathcal{I}}}{2}\log(2\pi) - \frac{\mathcal{U}_{\mathcal{I}}}{2} - \frac{\mathcal{U}_{L_\tau}}{2}\log\hat{\sigma}^{*2}_{L_\tau} - \frac{\mathcal{U}_{R_\tau}}{2}\log\hat{\sigma}^{*2}_{R_\tau} - f^{\flat}_{CHm}\left(\hat{\sigma}^{\flat *2}_{CHm}\right). \tag{B.14}$$

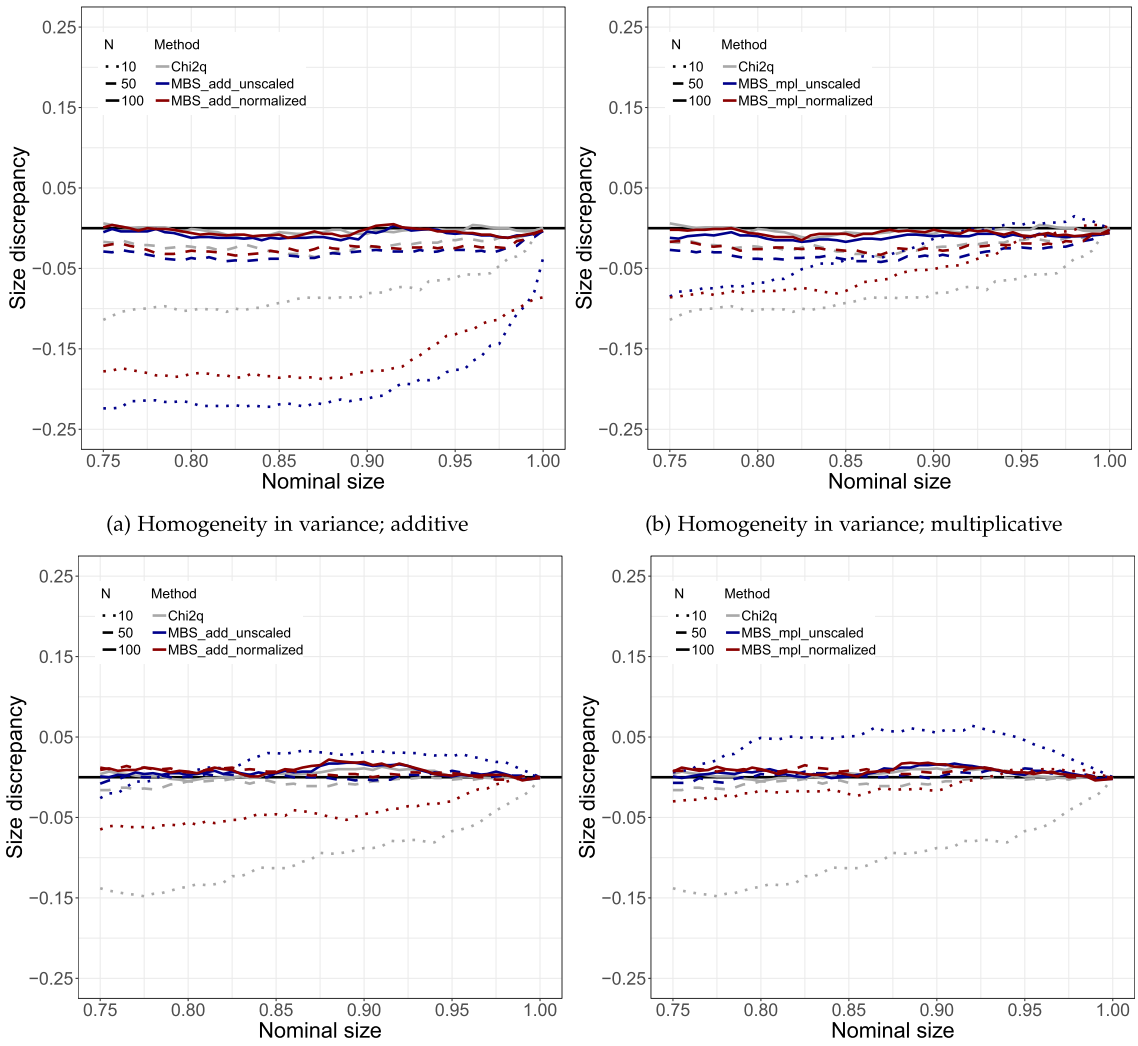
B.4 Simulation Evidence Based on Unscaled Weights

To compare the size of the proposed homogeneity tests relying on the multiplier bootstrap, with either (*i*) normalized weights (as in the main document) or (*ii*) unscaled weights (as outlined in this supplement), we follow the design of the respective simulation study motivated in Section 3.1.

In summary, we consider: weights $\{u_t\}$ from a Poisson distribution with $\lambda = 1$, $B = 1000$, $M = 1000$, $n_{L_\tau} = n_{R_\tau} = \{5, 25, 50\}$ with $n_{\mathcal{I}} = \{10, 50, 100\}$ as well as $\alpha = \{0.025, 0.05, 0.1\}$. Again, we consider two DGPs; we sample from a normal distribution $\mathcal{N}(0,1)$ and from a standardized $t(5)$-distribution.

Figures 12 and 13 illustrate the size discrepancies for the respective DGPs. A comparison between normalized weights and unscaled weights reveals that the normalization tends to reduce size distortions, especially for moderately large sample sizes, such as $n_{\mathcal{I}} = \{50, 100\}$. This effect is appears to be pronounced when testing for homogeneity in variance under correct model specification, regardless of the bias correction applied. Under model misspecification, normalizing the weights improves the size of both tests across all considered bias correction setups, except for small sample sizes, such as $n_{\mathcal{I}} = 10$ with $n_{L_\tau} = n_{R_\tau} = 5$.

B.5 Size Discrepancy: Unscaled Weights in the MBS



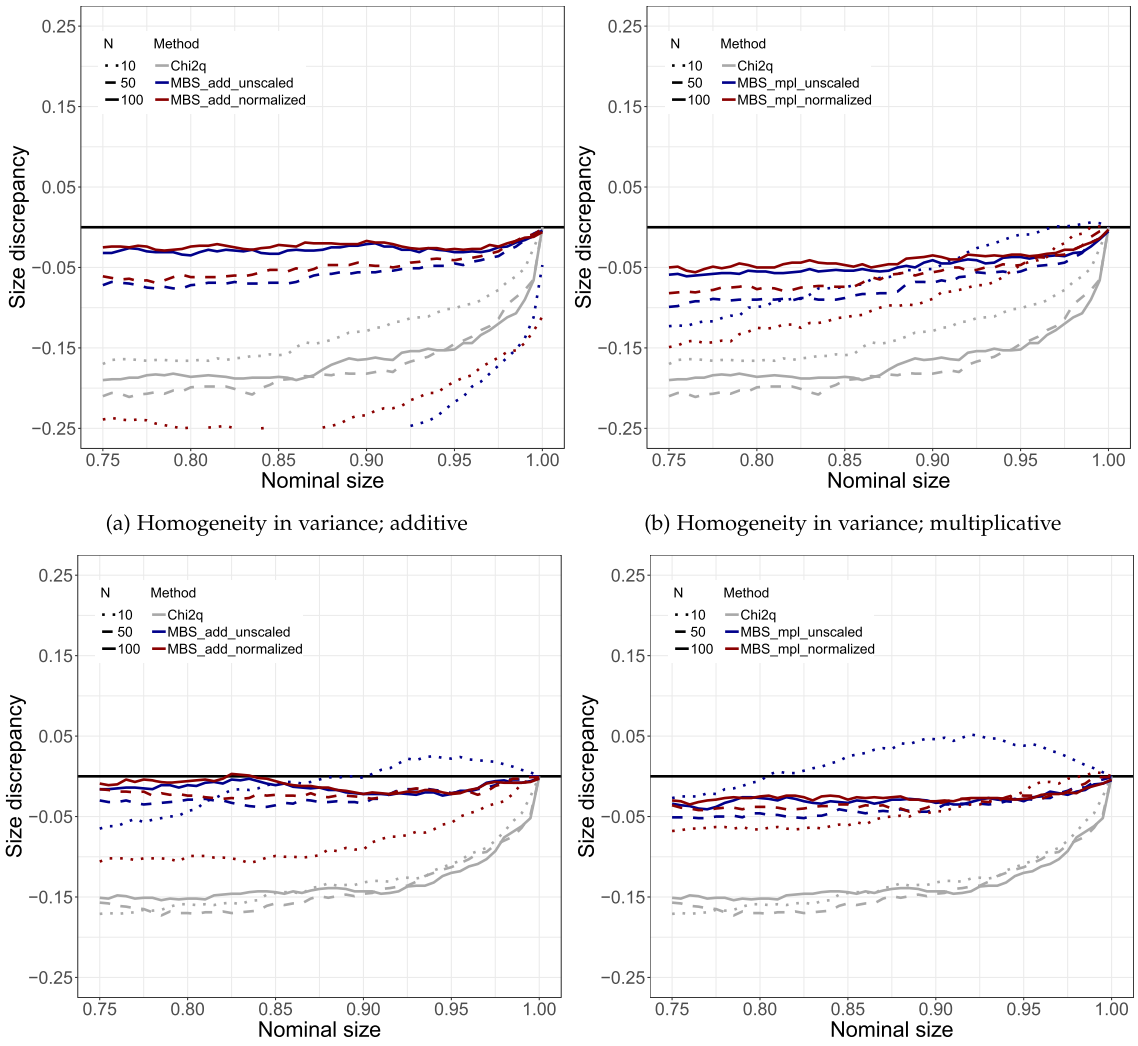
(a) Homogeneity in variance; additive

(b) Homogeneity in variance; multiplicative

(c) Complete homogeneity; additive

(d) Complete homogeneity; multiplicative

Figure 12: Size discrepancies (under H_0) for both tests in the case of the *i.i.d.* DGP, where we have $X \sim \mathcal{N}$. Red lines refer to tests utilizing normalized MBS weights, while blue ones refer to tests incorporating unscaled realizations. Results from the χ^2-test are illustrated by grey lines. Sample sizes are depicted in dotted, dashed, or solid lines for $n_{\mathcal{I}} = \{10, 50, 100\}$, respectively. Configuration of LCP: $M = 1000$, $B = 1000$ and $u \sim Poi(1)$.



(a) Homogeneity in variance; additive

(b) Homogeneity in variance; multiplicative

(c) Complete homogeneity; additive

(d) Complete homogeneity; multiplicative

Figure 13: Size discrepancies (under H_0) for both tests in the case of the *i.i.d.* DGP, where we have $X \sim t_5$. Red lines refer to tests utilizing normalized MBS weights, while blue ones refer to tests incorporating unscaled realizations. Results from the χ^2-test are illustrated by grey lines. Sample sizes are depicted in dotted, dashed, or solid lines for $n_{\mathcal{I}} = \{10, 50, 100\}$, respectively. Configuration of LCP: $M = 1000$, $B = 1000$ and $u \sim Poi(1)$.

Accepted 18 July 2025

REFERENCES

[1] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31(3)** 307–327. https://doi.org/10.1016/0304-4076(86)90063-1. MR0853051

[2] Chen, Y., Härdle, W. K. and Pigorsch, U. (2010). Localized Realized Volatility Modeling. *Journal of the American Statistical Association* **105(492)** 1376–1393. https://doi.org/10.1198/jasa.2010.ap09039. MR2796557

[3] Čížek, P., Härdle, W. and Spokoiny, V. (2009). Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models. *The Econometrics Journal* **12**(2) 248–271. https://doi.org/10.1111/j.1368-423X.2009.00292.x. MR2562386

[4] Davies, L., Höhenrieder, C. and Krämer, W. (2012). Recursive computation of piecewise constant volatilities. *Computational Statistics and Data Analysis* **56** 3623–3631. https://doi.org/10.1016/j.csda.2010.06.027. MR2943916

[5] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50(4)** 987–1007. https://doi.org/10.2307/1912773. MR0666121

[6] Francq, C. and Zakoïan, J. -M. (2019) *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons. https://doi.org/10.1002/9780470670057. MR3186556

[7] Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2006). A Haar-Fisz Technique for Locally Stationary Volatility Estimation. *Biometrika* **93(3)** 687–704. https://doi.org/10.1093/biomet/93.3.687. MR2261451

[8] Giacomini, E., Härdle, W. K. and Spokoiny, V. (2009). Inhomogeneous Dependence Modeling with Time-Varying Copulae. *Journal of Business & Economic Statistics* **27(2)** 224–234. https://doi.org/10.1198/jbes.2009.0016. MR2516440

[9] Härdle, W. K., Simar, L. and Fengler, M. R. (2024) *Applied Multivariate Statistical Analysis*, 6th ed. Springer. https://doi.org/10.1007/978-3-031-63833-6. MR4807482

[10] Härdle, W. K., Hautsch, N. and Mihoci, A. (2015). Local Adaptive Multiplicative Error Models for High-Frequency Forecasts. *Journal of Applied Econometrics* **30(4)** 529–550. https://doi.org/10.1002/jae.2376. MR3358635

[11] Härdle, W. K., Herwartz, H. and Spokoiny, V. (2003). Time inhomogeneous multiple volatility modelling. *Journal of Financial Econometrics* **1(1)** 55–95.

[12] Khowaja, K., Saef, D., Sizov, S. and Härdle, W. K. (2020). Data Analytics Driven Controlling: bridging statistical modeling and managerial intuition. *IRTG 1792 Discussion Paper* 2020–026.

[13] Klochkov, Y., Härdle, W. K. and Xu, X. (2019). Localizing Multivariate CAViaR. *IRTG 1792 Discussion Paper* 2019–007.

[14] Lee, A., Sandberg, R. and Sucarrat, G. (2025). Robust Estimation and Inference for Time-varying Unconditional Volatility. Unpublished manuscript.

[15] Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *The Annals of Statistics* **32(2)** 577–602. https://doi.org/10.1214/009053604000000102. MR2060170

[16] Mikosch, T. and Stărică, C. (2004). Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *The Review of Economics and Statistics* **86(1)** 378–390.

[17] Polzehl, J. and Spokoiny, V. (2006). Propagation-separation approach for local likelihood estimation. *Probability Theory and Related Fields* **135** 335–362. https://doi.org/10.1007/s00440-005-0464-1. MR2240690

[18] Preuss, P., Puchstein, R. and Dette, H. (2015). Detection of Multiple Structural Breaks in Multivariate Time Series. *Journal of the American Statistical Association* **110(510)** 654–668. https://doi.org/10.1080/01621459.2014.920613. MR3367255

[19] Spokoiny, V. (1998). Estimation of a Function with Discontinuities via Local Polynomial Fit with an Adaptive Window Choice. *The Annals of Statistics* **26(4)** 1356–1378. https://doi.org/10.1214/aos/1024691246. MR1647669

[20] Spokoiny, V. (2009). Multiscale local change point detection with applications to value-at-risk. *The Annals of Statistics* **37(3)** 1405–1436. https://doi.org/10.1214/08-AOS612. MR2509078

[21] Spokoiny, V. and Zhilova, M. (2015). Bootstrap confidence sets under model misspecification. *The Annals of Statistics* **43(6)** 2653–2675. https://doi.org/10.1214/15-AOS1355. MR3405607

[22] Stock, J. H. and Watson, M. W. (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual* **17** 159–218.

[23] Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics & Statistics* **87(3)** 503–522.

[24] Suvorikova, A. and Spokoiny, V. (2017). Multiscale change point detection. *Theory of Probability & Its Applications* **61(4)** 665–691. https://doi.org/10.1137/S0040585X97T988411. MR3632534

[25] Taylor, S. J. (1986). Modelling Financial Time Series. *John Wiley and Sons* **Chichester**.

[26] Taylor, S. J. (1987). Forecasting the volatility of currency exchange rates. *International Journal of Forecasting* **3**(1) 159–170.

[27] Van Bellegem, S. and von Sachs, R. (2004). Forecasting economic time series with unconditional time-varying variance. *International Journal of Forecasting* **20**(4) 611–627.

[28] Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics* **9(1)** 60–62.

[29] Xu, K. -L. and Phillip, P. C. B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics* **142** 65–90. https://doi.org/10.1016/j.jeconom.2007.06.001. MR2408736

[30] Zbonakova, L., Li, X. and Härdle, W. K. (2018). Penalized Adaptive Forecasting with Large Information Sets and Structural Changes. *IRTG 1792 Discussion Paper* 2018–039.

Matthias R. Fengler. School of Economics and Political Science, University of St.Gallen & Swiss Finance Institute, Switzerland.
E-mail address: matthias.fengler@unisg.ch

Bruno Jäger. Institute of Strategy and Marketing, Eastern Switzerland University of Applied Sciences, Switzerland.
E-mail address: bruno.jaeger@ost.ch

Ostap Okhrin. Faculty of Transportation, TUD Dresden University of Technology, Germany.
E-mail address: ostap.okhrin@tu-dresden.de