SFB 823

Residual-based inference on moment hypotheses, with an application to testing for constant correlation

Matei Demetrescu, Dominik Wied

Nr. 37/2016

Discussion Paper

Residual-based inference on moment hypotheses, with an application to testing for constant correlation

Matei Demetrescu*
Christian-Albrechts-University of Kiel

Dominik Wied†
University of Cologne

July 21, 2016

Abstract

Often, inference on moment properties of unobserved processes are conducted on the basis of estimated counterparts obtained in a preliminary step. In some situations, the use of residuals instead of the true quantities affects inference even in the limit, while in others there is no asymptotic residual effect. For the case of statistics based on partial sums of nonlinear functions of the residuals, we give here a characterization of the conditions under which the residual effect does not vanish as the sample size goes to infinity (generic regularity conditions provided). An o verview of methods to account for the residual effect is also provided. The analysis extends to models with change points in parameters at estimated time, in spite of the discontinuous manner in which the break time enters the model of interest. To illustrate the usefulness of the results, we propose a test for constant correlations allowing for breaks at unknown time in the marginal means and variances. We find, in Monte Carlo simulations and in an application to US and German stock returns, that not accounting for changes in the marginal moments has severe consequences.

Key words: Two-step procedure; Estimation error; Cumulated sums; Bootstrap; Structural break;

JEL classification: C12 (Hypothesis Testing)

1 Introduction

In many situations, estimated quantities are used for inferring on the properties of a latent data generating process. For example, in the linear regression model, researchers might investigate the third and fourth moments of residuals in order to test the normality of error terms; see Jarque and Bera (1980). Another example are tests for no structural breaks: Brown et al. (1975) use recursive residuals for testing the constancy of parameters in the linear model, while Ploberger and Krämer (1992) do the same with OLS residuals.[1] (Co)Variance stability tests have been proposed by Aue et al. (2009). More recently, Borowski et al. (2014) and Dette et al. (2015) consider a setting, where a time-varying signal function is added to a stochastic error term and residuals are used to test for constancy of the variance of the error term. Dette et al. (2015) also

*Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany, email: mdeme@stat-econ.uni-kiel.de.

†Institute for Econometrics and Statistics, University of Cologne, Germany, email: dwied@uni-koeln.de. Financial support by DFG (SFB 823, project A1) is gratefully acknowledged.

[1] Such stability tests for slope parameters can be conducted in more general frameworks, one well-known example being the work of Andrews (1993); see also Andrews and Ploberger (1994) and Hansen (2000).

consider testing for auto-correlation constancy in the case of time-varying variances which, among others, improves aspects of previous work of Wied, Krämer, and Dehling (2012), who test for cross-correlation constancy under the assumption of constant, yet unknown, variances.

The purpose of this paper is to provide a general analysis about the relationship between the limiting distribution of test statistics based on residuals and of the test statistics based on the unobservable counterparts.[2] To keep the problem tractable, we shall focus on statistics based on sums or partial sums of some transformation of the residuals of interest. Using residuals instead of the true series may have an effect[3] on the statistics under scrutiny, but this not need be the case in general. For instance, in the case of the OLS CUSUM test, the limit distribution is the supremum of the absolute value of a Brownian bridge, while it would base on the Brownian motion if one used the unobservable disturbances (Ploberger and Krämer, 1992). On the other hand, the distribution of the Jarque-Bera test for normality is claimed to remain unchanged in such situations, see Jarque and Bera (1980, p. 257) (although, as a byproduct of our analysis, we actually show the claim to be unsubstantiated), while Chen and Fan (2006) and Chan et al. (2009) show that the asymptotic distributions of estimators in copula models are not influenced by taking residuals from fitting marginal models.

Regularity conditions assumed, the effect of using residuals depends on both the filter which maps the unobservable terms of interest into observations and on the statistic of interest. To conduct the analysis, two types of filters are considered here, one which is continuous in unknown parameters and one which exhibits discontinuities in some of the variables allowing us e.g. to deal with change points. The unknown parameters are estimated with a full-sample estimator or with a recursive estimator.

The main contribution of this paper is to characterize the specific circumstances under which the residual effect appears or not. Moreover, we discuss selected aspects of asymptotic and bootstrap corrections for the cases where the residual effect is not asymptotically negligible. For instance, it turns out that the residual effect does not emerge in the scenario of Borowski et al. (2014) (which is based on the variance constancy test in Wied, Arnold, Bissantz, and Ziggel, 2012) if the signal function is piecewise constant and the break point fractions can be consistently estimated. Borowski et al. (2014) provided simulation evidence for this conjecture, but did not give a formal proof. Furthermore, the theoretical result in our paper complements the applicability of the variance constancy test in Dette et al. (2015), who only consider a smooth signal function and do not deal with the question if there might be situations in which the limit distribution remains the same.

We illustrate the details of our characterization on the basis of a test for constant correlations under breaks in marginal means or variances, with an application to the correlation of US and German stock market returns. In this regard, we improve the literature in several ways. While Dette et al. (2015) focus on auto-correlations, we propose a residual-based test for constant cross-correlations in the case of time-varying variances and show that taking residuals changes the limit distribution. In particular, we directly improve the work of Wied et al. (2012) by relaxing the assumption of constant variances and find e.g. that the breaks in marginal variances significantly changes the dating of correlation breaks.

The remainder of the paper is structured as follows. Section 2 introduces the setting in a formal way. Section 3 provides the asymptotic arguments for the smooth case and discusses the conditions under which the use of residuals instead of the true series does (does not) have an asymptotic effect, together with some asymptotic and bootstrap corrections. Section 4 addresses the case of structural changes and shows that plugging in an estimated break time is asymptotically equivalent to employing the true break time in what concerns the

[2] Note that our approach is somewhat related to two other branches in the literature. The first one is the topic of generated regressors, see Mammen et al. (2012), where people analyze the effect of estimating regressors on subsequent estimation problems. The second one is the topic of two-stage parameter estimation, see Newey and McFadden (1994), where the effect of the first on the second estimation step is analyzed.

[3] In conjunction with tests for distribution, this is often called the Durbin effect (Durbin, 1973).

residual effect. Section 5 gives concrete examples, introduces the new correlation constancy test, and gives Monte Carlo illustrations for the proposed test. Section 6 provides the application to the correlation of US and German stock markets. The proofs have been gathered in the appendix.

2 Setup

Suppose one is interested in inference about the moment properties of some data generating process [DGP] on the basis of a sample $\boldsymbol{Z}_t \in \mathbb{R}^K$, $t = 1, \ldots, n$, for which the partial sums $\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \boldsymbol{g}(\boldsymbol{Z}_t)$ are relevant. The particular shape of the function $\boldsymbol{g} \mapsto \mathbb{R}^L$ depends on the question of interest; e.g. $\boldsymbol{g}(z) = \left(z^3, z^4\right)$ for the test of Jarque and Bera (1980) and $g(\boldsymbol{z}) = z_1 z_2$ for pairwise covariances (or correlations, if $Z_{t,1}$ and $Z_{t,2}$ are standardized).

We however assume that one only observes n values, say $\boldsymbol{X}_t$, $t = 1, \ldots, n$, of some (nonlinear) filter of the variables of interest $\boldsymbol{Z}_t$; quite often, $\boldsymbol{Z}_t$ are disturbances in a (regression) model or $\boldsymbol{Z}_t$ are standardized versions of $\boldsymbol{X}_t$. In time series, one may well have a linear finite-order filter where $\boldsymbol{Z}_t$ are the innovations of a moving average process say, $\boldsymbol{X}_t = \sum_{j=0}^{q} B_j \boldsymbol{Z}_{t-j}$. To nest all these possible scenarios, we assume a parametric relation between the two, in the most general case

$$\boldsymbol{X}_t = \boldsymbol{f}\left(\boldsymbol{Z}_t, \boldsymbol{Z}_{t-1}, \ldots, {}^{t}\!/\!{}_{n}; \boldsymbol{\theta}\right).$$

Let the length M of the parameter vector $\boldsymbol{\theta}$ be finite.

In practice, the true values $\boldsymbol{\theta}_0$ of the parameters are not known so the filter $\boldsymbol{f}$ cannot be inverted to give the necessary $\boldsymbol{Z}_t$. Rather, one is forced to resort to estimates thereof, i.e. residuals $\hat{\boldsymbol{Z}}_t$ based on some estimators $\hat{\boldsymbol{\theta}}$ of the unknown parameters. The relation between the limit distributions based on $\hat{\boldsymbol{Z}}_t$ and those based on $\boldsymbol{Z}_t$ depends on both $\boldsymbol{g}$ and $\boldsymbol{f}$, as well as on the properties of the estimators $\hat{\boldsymbol{\theta}}$, which we assume to belong to the family of generalized method-of-moments [GMM] estimators (Hansen, 1982), which includes e.g. M estimators as a particular case.

The above formulation is fairly general. For instance, the dependence of $\boldsymbol{f}$ on the index t allows one to model e.g. trends, say in an additive model such as $\boldsymbol{X}_t = {}^{t}\!/\!{}_{n}\, \boldsymbol{\theta} + \boldsymbol{Z}_t$. Additivity is not critical for the analysis, while the smoothness properties of $\boldsymbol{f}$ are.

Regarding smoothness, we shall consider two situations. In the first, $\boldsymbol{f}$ is smooth in the parameters $\boldsymbol{\theta}$. In the second, we model discontinuities explicitly in form of change points (structural breaks).[4] In a simple case, say for the mean, we may encounter $\mathrm{E}\left(\boldsymbol{X}_t\right) = \boldsymbol{\mu}_1$, $1 \leq t < N$ and $\mathrm{E}\left(\boldsymbol{X}_t\right) = \boldsymbol{\mu}_2$, $N \leq t < n$, so, considering $N = [\lambda n]$ for some $\lambda \in (0,1)$, one may work with the model $\boldsymbol{X}_t = \boldsymbol{Z}_t + \boldsymbol{\mu}_1 \mathbb{I}\left({}^{t}\!/\!{}_{n} < \lambda\right) + \boldsymbol{\mu}_2 \mathbb{I}\left({}^{t}\!/\!{}_{n} \geq \lambda\right)$ with $\mathrm{E}\left(\boldsymbol{Z}_t\right) = \boldsymbol{0}$ and $\mathbb{I}$ the indicator function.[5] Here, $\boldsymbol{f}\left(\boldsymbol{z}, {}^{t}\!/\!{}_{n}, (\boldsymbol{\mu}, \lambda)\right) = \boldsymbol{z} + \boldsymbol{\mu}_1 \mathbb{I}\left({}^{t}\!/\!{}_{n} < \lambda\right) + \boldsymbol{\mu}_2 \mathbb{I}\left({}^{t}\!/\!{}_{n} \geq \lambda\right)$ is discontinuous in the parameter λ, but smooth in $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. This will be captured more generally via the model

$$\boldsymbol{X}_t = \boldsymbol{f}\left(\boldsymbol{Z}_t, \boldsymbol{Z}_{t-1}, \ldots, {}^{t}\!/\!{}_{n}; \boldsymbol{\theta}_1\right) \mathbb{I}\left({}^{t}\!/\!{}_{n} < \lambda\right) + \boldsymbol{f}\left(\boldsymbol{Z}_t, \boldsymbol{Z}_{t-1}, \ldots, {}^{t}\!/\!{}_{n}; \boldsymbol{\theta}_2\right) \mathbb{I}\left({}^{t}\!/\!{}_{n} \geq \lambda\right),$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are taken to be estimated for each subsample using the same method as in the smooth case. In the most general case one may allow for a finite number of such discontinuity points. Although this is a particular case of a time-dependent filter, we treat it separately due to its practical relevance and because of the discontinuity in λ. We deal with this situation in more detail in Section 4 and focus for now on the case without breaks.

[4] The arguments regarding breaks could likely be extended to discuss threshold models; we do not pursue the topic here, though.

[5] Although one may add an extra n in the notation to acknowledge the triangular array structure of such DGPs, we omit this to ease notation.

We shall assume the (causal) filter generating $\boldsymbol{X}_t$ to be invertible in the sense that there exists a (causal) filter $\boldsymbol{h}$ such that the series $\boldsymbol{Z}_t$ is uniquely given by

$$\boldsymbol{Z}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}\right),$$

i.e. $\boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}\right) = \boldsymbol{Z}_t \ \forall t$ iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ with $\boldsymbol{\theta}_0$ the true parameter value. The corresponding representation for breaks, when needed, is given by

$$\boldsymbol{Z}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}_1\right) \mathbb{I}\left(t/n < \lambda\right) + \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}_2\right) \mathbb{I}\left(t/n \geq \lambda\right) \tag{1}$$

and is assumed to hold uniquely as well.

In the case of time-series models, except for finite-order (nonlinear) autoregressive models, the initial conditions play a role since the full relevant past of $\boldsymbol{X}_t$ is not available in finite samples. In such situations, one may have to resort to truncated versions of the involved filters, $\boldsymbol{Z}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_1, t/n; \boldsymbol{\theta}\right)$, and impose technical conditions such as $\sup_{s\in[0,1]} \frac{1}{\sqrt{n}} \left\| \sum_{t=1}^{[sn]} \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_1, t/n; \boldsymbol{\theta}\right) - \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}\right) \right\| \xrightarrow{p} 0$ that ensure the difference between the truncated and the unfeasible filter to be asymptotically negligible. We do not further pursue this topic.

Given a sample $\{\boldsymbol{X}_t\}$, $t = 1, \ldots, n$, and an estimator for the unknown true parameter values $\boldsymbol{\theta}_0$, we may thus estimate the variables of interest $\boldsymbol{Z}_t$. We consider two possible estimation scenarios, first a full-sample approach delivering the estimator $\hat{\boldsymbol{\theta}}$, and, second, an adaptive, or recursive, approach (i.e. based on the sample $1, \ldots, t$) delivering the sequence of estimators $\hat{\boldsymbol{\theta}}_t$. Note that $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_n$, but also that time variation in $\boldsymbol{\theta}$ is only allowed if modelling it explicitly (like the break case introduced above). Recursive estimation is involved e.g. in the case of inference on correlations (Wied et al., 2012), where the sample variances in the denominator of the relevant correlation coefficient are computed up to time t, but has a much longer history; see Kianifard and Swallow (1996) for an earlier review. Assuming a that GMM-type estimator with $N \geq M$ moment restrictions is available for estimating $\boldsymbol{\theta}$, we may represent it as

$$\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0 = \left(\sum_{j=1}^{t} B'_{j,n} W_n \sum_{j=1}^{t} B_{j,n} \right)^{-1} \sum_{j=1}^{t} B'_{j,n} W_n \sum_{j=1}^{t} A_{j,n} + R_{t,n}$$

with suitable limiting behavior of these generic components $B_{j,n}$ $(N \times M)$, $A_{j,n}$ $(N \times 1)$ and $R_{t,n}$ $(M \times 1)$; see Assumption 1 below. For simplicity, the $N \times N$ GMM weighting matrix W_n is not computed recursively. The components $A_{j,n}$, $B_{j,n}$ and $R_{t,n}$ depend explicitly on $\boldsymbol{X}_t$, and implicitly (via the DGP) on $\boldsymbol{\theta}_0$.

The residuals are given as

$$\hat{\boldsymbol{Z}}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_1, t/n; \hat{\boldsymbol{\theta}}\right) \quad \text{or} \quad \tilde{\boldsymbol{Z}}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_1, t/n; \hat{\boldsymbol{\theta}}_t\right),$$

and inference on $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)$ is based on the partial sums of the transformed residuals,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right) \quad \text{or} \quad \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right), \quad s \in [0,1].$$

We now outline high-level assumptions on the DGP and the estimators that allow for a discussion of the residual effect in a generic framework.

Assumption 1 *With "$\Rightarrow$" denoting weak convergence in a space of cadlag functions on* $[0,1]$ *endowed with a suitable metric, it holds that:*

1. $\sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{t=1}^{[sn]}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)-\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \\ \frac{1}{n}\sum_{t=1}^{[sn]}A_{t,n}\end{pmatrix}\Rightarrow\boldsymbol{\Psi}\left(s\right)$, *where* $\Psi\left(s\right)$ *is an* $L+N$*-dimensional Gaussian process with* $\boldsymbol{\Psi}\left(0\right)=0$ *a.s. and* $\mathrm{Cov}\left(\boldsymbol{\Psi}\left(1\right)\right)=\Xi$;
2. $\frac{1}{n}\sum_{t=1}^{[sn]}B_{t,n}\Rightarrow\Pi\left(s\right)$ *where* $\Pi\left(s\right)$ *is a deterministic* $N\times M$ *matrix of Lipschitz functions, of rank* M *at all* $s\in(0,1]$, $\Pi\left(0\right)=0$; *furthermore,* $\sqrt{n}\sup_{s\in[\epsilon,1]}\left|R_{[sn],n}\right|\xrightarrow{p}0$, $\epsilon\in(0,1)$, *and* $W_n\xrightarrow{p}W$ *with* W *a positive definite matrix;*
3. $\frac{1}{n}\sum_{t=1}^{[ns]}\left.\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\Rightarrow\boldsymbol{\tau}\left(s\right)$ *where* $\boldsymbol{\tau}\left(s\right)$ *is a deterministic matrix of differentiable functions;*[6]
4. *For some neighbourhood* $\Phi_n=\left\{\boldsymbol{\theta}^*:\|\boldsymbol{\theta}^*-\boldsymbol{\theta}_0\|<Cn^{-1/2+\epsilon},\ 0<\epsilon<1/2,\ C>0\right\}$ *of* $\boldsymbol{\theta}_0$,

$$\sup_{\boldsymbol{\theta}_t^*\in\Phi_n;t=1,\dots,n}\left\|\left.\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^*}-\left.\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\|\xrightarrow{p}0$$

   where $\boldsymbol{Z}_t^*=\boldsymbol{h}\left(\boldsymbol{X}_t,\dots,;\boldsymbol{\theta}_t^*\right)$.

The assumption first specifies the joint behavior of the sample moment conditions for estimation and the relevant sample moments. Under weak stationarity and short memory of the involved quantities, the limit process $\boldsymbol{\Psi}\left(s\right)$ is a Brownian motion. But a more general Gaussian process is allowed for; e.g. slowly varying variances can be encompassed and $\boldsymbol{\Psi}$ has independent Gaussian, but not stationary increments. This may be the case under local stationarity of the DGP; see e.g. Hansen (2000) and, more recently, Zhou (2013), for specific parameter stability tests under local stationarity.

The first two conditions together also allow us to describe the asymptotic behavior of the estimators of $\boldsymbol{\theta}$. Note that the recursive estimators $\hat{\boldsymbol{\theta}}_t$ do not have proper asymptotics for $t=O\left(1\right)$. Still, for any $0<\epsilon<1$, we have as a consequence of Assumption 1 the weak convergence

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{[sn]}-\boldsymbol{\theta}_0\right)\Rightarrow\left(\Pi'(s)\,W\,\Pi(s)\right)^{-1}\Pi'(s)\,W\,\boldsymbol{\Psi}_{(L+1):N}\left(s\right)\quad\text{for } s\in[\epsilon,1],$$

for any $0<\epsilon<1$. The convergence does not extend to $[0,1]$. To deal with this situation one typically adds a step showing that $\hat{\boldsymbol{\theta}}_t$ for $t\in\{1,\dots,[\epsilon n]\}$ do not have an asymptotic effect on the statistic of interest as $\epsilon\to0$. See e.g. Wied et al. (2012). This may require additional assumptions on the behavior of $R_{t,n}$ for "small" t. Since they would depend on the particular statistic to be analyzed, we do not attempt to give a set of conditions here and recommend a case-by-case discussion. Obviously, this is not relevant when using full-sample estimation.

Condition 3 introduces the essential quantity involved in the residual effect: we show in Section 3 that the residual effect vanishes in the limit if $\boldsymbol{\tau}$ is zero. But there are other interesting special cases where the residual effect vanishes when $\boldsymbol{\tau}$ has specific forms; see Section 3.2 for the precise details.

Condition 4 imposes a form of uniform smoothness of the relevant model components. Essentially, the approximation error due to linearization of the estimation noise $\hat{\boldsymbol{Z}}_t-\boldsymbol{Z}_t$ is assumed to be controlled for in a neighbourhood of $\boldsymbol{\theta}_0$ that is "small enough" to avoind imposing unrealistic assumptions but still "large enough" to contain the estimators $\hat{\boldsymbol{\theta}}$ $(\hat{\boldsymbol{\theta}}_t)$ with probability approaching unity. This could e.g. be achieved by bounding the elements of the Hessians of $\boldsymbol{g}$ and $\boldsymbol{h}$ or suitable bounds for the parameter space, but the properties of $\boldsymbol{Z}_t$ also play a role, so imposing moment properties on $\boldsymbol{Z}_t$ may relax the requirements on $\boldsymbol{g}$ or $\boldsymbol{h}$. This too has to be discussed on a case-by-case basis.

As a general remark, it comes natural to assume some form of short memory, say mixing properties, for $\boldsymbol{Z}_t$ and require that the assumed model $\boldsymbol{f}$ be restricted in such a way that the resulting random elements

[6]This is the line vector version of the gradient and the conformable version of the Jacobian.

($\boldsymbol{Z}_t$, $\boldsymbol{X}_t$, $A_{t,n}$ and $B_{t,n}$) be mixing themselves, which can then be used to establish the required weak convergence results. See e.g. Davidson (1994, Chapter 29) for sets of suitable technical conditions. Bootstrap implementations may require additional smoothness conditions themselves. Note however that e.g. unit root or cointegrated DGPs are largely excluded since, in such nonstandard cases, $\hat{\boldsymbol{\theta}}_{[ns]} - \boldsymbol{\theta}_0$ would typically be non-Gaussian in the limit, and the convergence rate would not be $\sqrt{n}$; while accounting for this is not conceptually difficult, the notational effort is not trivial and we do not further consider this topic here.

The second assumption is of relevance for the kind of statistics we look at, which require estimation of scaling matrices. For feasibility of test statistics based on (sample) moments, the following assumption regarding normalization is useful.

Assumption 2 *There exists an estimator $\hat{\Xi}$ such that $\hat{\Xi} \xrightarrow{p} \Xi$.*

Often, HAC estimators (Newey and West, 1987; Andrews, 1991) would be employed for estimation of Ξ based on residuals and sample moment conditions, although they are not the only choice (see e.g. Phillips et al., 2006). Note that HAC estimators are often consistent even for data generating processes that are only locally stationary; see e.g. Cavaliere (2004) for the case of time-varying variances.

Assumption 1 implies weak convergence of the centered partial sums of $\boldsymbol{g}$ and of the moment conditions $A_{j,n}$. It will be convenient to standardize the limit processes such that, with

$$\Xi = \begin{pmatrix} \Omega & \Lambda' \\ \Lambda & \Sigma \end{pmatrix},$$

we may write

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \Rightarrow \Omega^{1/2} \boldsymbol{\Gamma}\left(s\right)$$

where $\boldsymbol{\Gamma}(s) = \Omega^{-1/2} \Psi_{1:L}(s)$ is a Gaussian process with $\boldsymbol{\Gamma}(1) \sim \mathcal{N}(0, I_L)$, and

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{[ns]} - \boldsymbol{\theta}_0\right) \Rightarrow \left(\Pi'(s)\, W\, \Pi(s)\right)^{-1} \Pi'(s)\, W\, \Sigma^{1/2} \boldsymbol{\Theta}\left(s\right)$$

on $[\epsilon, 1]$, where $\boldsymbol{\Theta}(s) = \Sigma^{-1/2} \boldsymbol{\Psi}_{(L+1):(L+N)}(s)$ is a Gaussian process with $\boldsymbol{\Theta}(1) \sim \mathcal{N}(0, I_N)$.

If one can base the tests directly on $\boldsymbol{Z}_t$, then only $\boldsymbol{\Gamma}(s)$ and Ω will be relevant for inference. Otherwise, Σ, Λ, Π, $\boldsymbol{\Theta}$ and $\boldsymbol{\tau}$ would play a role. We discuss this role in the following section.

3 The residual effect

The effect depends on what kind of statistics one is interested in. For estimating $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)$ via sample averages of $\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right)$, it is quite plausible that there is no asymptotic effect and we do not discuss this formally. But for centered, normalized partial sums, the picture is different as has been studied in numerous particular cases (see e.g. Bai and Ng, 2005, Theorem 1, for a formulation for higher-order moments of $\boldsymbol{Z}_t$ in a linear regression setup).

3.1 Residual-based partial sums

We formulate the first result in the following

Proposition 1 *Under Assumption 1, it holds as $T \to \infty$ that*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \Rightarrow \Omega^{1/2}\boldsymbol{\Gamma}\left(s\right) + \boldsymbol{\tau}\left(s\right)\left(\Pi'(1)\, W\, \Pi(1)\right)^{-1} \Pi'(1)\, W\, \Sigma^{1/2}\boldsymbol{\Theta}\left(1\right)$$

and, on $[\epsilon, 1]$ for any $0 < \epsilon < 1$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \Rightarrow \Omega^{1/2}\boldsymbol{\Gamma}\left(s\right) + \left(\int_0^s \boldsymbol{\Theta}'\left(r\right) \left(\Sigma^{1/2}\right)' W'\Pi(r) \left(\Pi'(r)\, W\, \Pi(r)\right)^{-1} \mathrm{d}\boldsymbol{\tau}'\left(r\right)\right)'.$$

Proof: *See the Appendix.*

Remark 1 *Although $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ are in general distinct, they are allowed to have common components; in fact, it is not excluded that they are identical in particular situations. The latter happens e.g. in the simple case of demeaning where $\hat{\boldsymbol{\theta}} = \bar{\boldsymbol{X}}$ so $\hat{\boldsymbol{Z}}_t = \boldsymbol{X}_t - \bar{\boldsymbol{X}}$, where $\boldsymbol{\Gamma} \equiv \boldsymbol{\Theta}$ and the proposition reduces, in the full-sample estimation scenario, to the well-known result of a Brownian bridge.*

Remark 2 *The proposition requires the inverse filter $\boldsymbol{h}$ to be differentiable in $\boldsymbol{\theta}$. This does not exclude structural breaks in the parameters, as long as the break time is known. We examine this situation more closely in Section 4, where we also prove that an unknown break time λ can be dealt with as well, in spite of entering the model in a discontinuous setup, provided that the estimate is precise enough; see Proposition 2 for details.*

The main implication of the proposition is that the residual effect appears for partial sums whenever $\boldsymbol{\tau}$ is not zero. Tests based on partial sums would not be affected if $\boldsymbol{\tau}\left(s\right) = \mathbf{0}$ for all $s \in [0,1]$,[7] but there are some additional situations where specific tests are not affected even if $\boldsymbol{\tau} \neq \mathbf{0}$.

3.2 Implications for selected tests

We first discuss testing simple hypotheses on the expectation of $\boldsymbol{g}\left(\boldsymbol{Z}_t\right)$. The null is of the form $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right) = \boldsymbol{\mu}^{(0)}$, and the Wald-type test statistic against alternatives of the form $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right) \neq \boldsymbol{\mu}^{(0)}$ is

$$\mathcal{T} = n\left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}^{(0)}\right)' \Omega^{-1} \left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}^{(0)}\right)$$

where $\bar{\boldsymbol{g}}$ is the sample average of $\boldsymbol{g}\left(\boldsymbol{Z}_t\right)$. The scale matrix Ω is typically unknown and is replaced by an estimate $\hat{\Omega}$; this would typically be the corresponding block of $\hat{\Xi}$, so a consistent estimator is available under Assumption 2.

The naive feasible versions of the test statistic are

$$\hat{\mathcal{T}} = n\left(\hat{\bar{\boldsymbol{g}}} - \boldsymbol{\mu}^{(0)}\right)' \hat{\Omega}^{-1} \left(\hat{\bar{\boldsymbol{g}}} - \boldsymbol{\mu}^{(0)}\right)$$

and

$$\tilde{\mathcal{T}} = n\left(\tilde{\bar{\boldsymbol{g}}} - \boldsymbol{\mu}^{(0)}\right)' \hat{\Omega}^{-1} \left(\tilde{\bar{\boldsymbol{g}}} - \boldsymbol{\mu}^{(0)}\right)$$

where $\hat{\bar{\boldsymbol{g}}}$ is the sample average of $\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right)$ and $\tilde{\bar{\boldsymbol{g}}}$ the sample average of $\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right)$.

[7] Newey and McFadden (1994) derive a similar condition under which the first-stage estimation has no effect on the limiting distribution of the second-stage estimators.

It follows from Proposition 1 together with Assumption 2 that, under the null $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right) = \boldsymbol{\mu}_0$

$$\hat{\mathcal{T}} \xrightarrow{d} \hat{\mathbf{\Gamma}}'\left(1\right) \hat{\mathbf{\Gamma}}\left(1\right)$$

where

$$\hat{\mathbf{\Gamma}}\left(s\right) = \mathbf{\Gamma}\left(s\right) + \Omega^{-1/2}\boldsymbol{\tau}\left(s\right)\left(\Pi'(1)\,W\,\Pi(1)\right)^{-1}\Pi'(1)\,W\,\Sigma^{1/2}\mathbf{\Theta}\left(1\right)$$

and

$$\tilde{\mathcal{T}} \xrightarrow{d} \tilde{\mathbf{\Gamma}}'\left(1\right) \tilde{\mathbf{\Gamma}}\left(1\right)$$

where

$$\tilde{\mathbf{\Gamma}}\left(s\right) = \mathbf{\Gamma}\left(s\right) + \Omega^{-1/2}\left(\int_0^s \mathbf{\Theta}'\left(r\right)\left(\Sigma^{1/2}\right)' W'\Pi(r)\left(\Pi'(r)\,W\,\Pi(r)\right)^{-1} \mathrm{d}\boldsymbol{\tau}'\left(r\right)\right)'.$$

Without residuals, $\mathcal{T} \xrightarrow{d} \mathbf{\Gamma}\left(1\right)' \mathbf{\Gamma}\left(1\right)$ under the null and follows as such a χ^2_L limiting null distribution according to Assumption 1, so the naive feasible versions are not pivotal in general, except for the obvious situation where $\boldsymbol{\tau} = \mathbf{0}$ for all $s \in [0,1]$; the other exception is when $\boldsymbol{\tau}(1) = \mathbf{0}$, at least for full-sample estimation, as pointed out by the following

Corollary 1 *Under Assumptions 1 – 2, the statistics $\mathcal{T}$, $\hat{\mathcal{T}}$ and $\tilde{\mathcal{T}}$ are asymptotically equivalent under the null if $\boldsymbol{\tau}\left(s\right) = \mathbf{0}$ for all $s \in [0,1]$. Furthermore, the statistics $\mathcal{T}$ and $\hat{\mathcal{T}}$ are asymptotically equivalent if $\boldsymbol{\tau}(1) = \mathbf{0}$.*

It is not straightforward (but also not inconceivable) to imagine a situation where $\boldsymbol{\tau}(1) = \mathbf{0}$ but $\boldsymbol{\tau}$ is not zero. Still, $\boldsymbol{\tau}(s) = \mathbf{0}$ for all $s \in [0,1]$ is the more plausible mechanism of making the residual effect negligible in this case. We give some examples in Section 5, while the following subsection considers correction strategies.

Moving on to testing hypotheses of constancy, $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_1\right)\right) = \ldots = \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_n\right)\right)$ the classical multivariate CUSUM statistic is given by

$$Q_n = \max_{1\leq j\leq n} \frac{j}{\sqrt{n}} \sqrt{\left(\boldsymbol{S}_j - \boldsymbol{S}_n\right)' \Omega^{-1}\left(\boldsymbol{S}_j - \boldsymbol{S}_n\right)} \qquad \text{with} \qquad \boldsymbol{S}_j = \frac{1}{j}\sum_{t=1}^{j} \boldsymbol{g}\left(\boldsymbol{Z}_t\right),$$

while the naive feasible versions are

$$\hat{Q}_n = \max_{1\leq j\leq n} \frac{j}{\sqrt{n}} \sqrt{\left(\hat{\boldsymbol{S}}_j - \hat{\boldsymbol{S}}_n\right)' \hat{\Omega}^{-1}\left(\hat{\boldsymbol{S}}_j - \hat{\boldsymbol{S}}_n\right)} \qquad \text{with} \qquad \hat{\boldsymbol{S}}_j = \frac{1}{j}\sum_{t=1}^{j} \boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right) \tag{2}$$

and

$$\tilde{Q}_n = \max_{1\leq j\leq n} \frac{j}{\sqrt{n}} \sqrt{\left(\tilde{\boldsymbol{S}}_j - \tilde{\boldsymbol{S}}_n\right)' \hat{\Omega}^{-1}\left(\tilde{\boldsymbol{S}}_j - \tilde{\boldsymbol{S}}_n\right)} \qquad \text{with} \qquad \tilde{\boldsymbol{S}}_j = \sum_{t=1}^{j} \boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right).$$

As a consequence of Proposition 1 and Assumption 2, we have

$$\hat{Q}_n \Rightarrow \sup_{s\in[0,1]} \sqrt{\left(\hat{\mathbf{\Gamma}}\left(s\right) - s\hat{\mathbf{\Gamma}}\left(1\right)\right)'\left(\hat{\mathbf{\Gamma}}\left(s\right) - s\hat{\mathbf{\Gamma}}\left(1\right)\right)},$$

and

$$\tilde{Q}_n \Rightarrow \sup_{s\in[0,1]} \sqrt{\left(\tilde{\mathbf{\Gamma}}\left(s\right) - s\tilde{\mathbf{\Gamma}}\left(1\right)\right)'\left(\tilde{\mathbf{\Gamma}}\left(s\right) - s\tilde{\mathbf{\Gamma}}\left(1\right)\right)}.$$

Had one computed the statistic using the unobserved $\boldsymbol{Z}_t$, the following well-known (pivotal) distribution would have been obtained,

$$Q_n \Rightarrow \sup_{s\in[0,1]} \sqrt{\left(\mathbf{\Gamma}\left(s\right) - s\mathbf{\Gamma}\left(1\right)\right)'\left(\mathbf{\Gamma}\left(s\right) - s\mathbf{\Gamma}\left(1\right)\right)};$$

so it is interesting to ask, when is the distribution not affected by the residual effect.

Again, $\hat{Q}_n$ and $\tilde{Q}_n$ are asymptotically equivalent with Q_n when $\boldsymbol{\tau}(s) = \mathbf{0}$; but, in addition, there is another interesting case where equivalence of CUSUM statistics is given, at least for $\hat{Q}_n$:

Corollary 2 *Under Assumptions 1 – 2, the statistics Q_n, $\hat{Q}_n$ and $\tilde{Q}_n$ are asymptotically equivalent if $\boldsymbol{\tau}\left(s\right) = \mathbf{0}$ for all $s \in [0,1]$. Moreover, the statistics Q_n and $\hat{Q}_n$ are asymptotically equivalent if $\boldsymbol{\tau}\left(s\right) = s\boldsymbol{\tau}$ for some constant $L \times M$ matrix $\boldsymbol{\tau}$.*

The condition under which the corollary holds is likely to be fulfilled in strictly stationary data generating processes, and unlikely to be fulfilled in data generating processes with structural breaks; see Section 5 for examples. Essentially, it requires first-order stationarity of $\left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, but note that this actually is compatible with breaks when $\boldsymbol{\tau} = \mathbf{0}$.

Finally, note that one may resort to a Cramér-von Mises type functional instead of the sup functional; this does not affect the validity of Corollary 2.

3.3 Asymptotic and bootstrap corrections

For the cases where there is a residual effect, corrections are required. We first discuss the more straightforward case of simple hypotheses, $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right) = \boldsymbol{\mu}^{(0)}$.

If basing the test on residuals with full-sample parameter estimation, we note that, under the null,

$$\sqrt{n}\left(\bar{\boldsymbol{g}} - \boldsymbol{\mu}^{(0)}\right) \Rightarrow \Omega^{1/2}\mathbf{\Gamma}\left(1\right) + \boldsymbol{\tau}\left(1\right)\left(\Pi'(1)\, W\, \Pi(1)\right)^{-1}\Pi'(1)\, W\, \Sigma^{1/2}\mathbf{\Theta}\left(1\right)$$

which is actually multivariate normally distributed, so making the distribution of this quadratic form pivotal is just a matter of using the right covariance matrix estimator: $\hat{\Omega}$ is only correct when $\boldsymbol{\tau}$ is zero; see the corollaries above. Otherwise, one should have used

$$\left(I_L; W'\Pi(1)\left(\Pi'(1)\, W\, \Pi(1)\right)^{-1}\boldsymbol{\tau}'\left(1\right)\right)\hat{\Xi}\begin{pmatrix} I_L \\ \boldsymbol{\tau}\left(1\right)\left(\Pi'(1)\, W\, \Pi(1)\right)^{-1}\Pi'(1)\, W \end{pmatrix} \tag{3}$$

instead of $\hat{\Omega}$. This situation is quite often encountered in the literature; see e.g. Bai and Ng (2005).

This correction is not available for recursive estimation of the parameters. The difference is that $\mathrm{Cov}\left(\tilde{\mathbf{\Gamma}}\left(1\right)\right)$ depends on the entire path of $\boldsymbol{\tau}$ which makes a correct estimation of the required covariance matrix difficult. While this is feasible, it would perhaps be easier to resort to a bootstrap scheme, as is not uncommon in the literature; see e.g. Zhou (2013) and Hansen (2000).

This too is not without disadvantages, though; see the discussion below.

Note also the following. If $\boldsymbol{g}\left(\boldsymbol{Z}_t\right)$ is weakly stationary then $\Omega^{1/2}\mathbf{\Gamma}$ is a Brownian motion, fully specified by Ω, and $\mathbf{\Gamma}$ is just a vector of independent Wiener processes. Under time-varying 2nd moments of $\boldsymbol{g}\left(\boldsymbol{Z}_t\right)$, however, the process $\Omega^{1/2}\mathbf{\Gamma}$ would have nonlinear (at best piecewise linear) quadratic covariation. In this case $\mathbf{\Gamma}$ cannot be a vector of independent Wiener processes, and setting the covariance matrix at $s = 1$ to be unity only norms Ω as quadratic covariance matrix of the limit process of centered partial sums of $\boldsymbol{g}\left(\boldsymbol{Z}_t\right)$;

the same holds for $\mathbf{\Theta}$, and the test statistic is not asymptotically pivotal under the null. E.g. Zhou (2013) suggests the use of the block wild bootstrap to accommodate locally stationary DGPs.

Moving on to the case of moment constancy tests, it is worth asking the question whether $\hat{Q}_n$ or $\tilde{Q}_n$ could be corrected using the right covariance matrix estimator, like in the case of simple hypotheses. This is more difficult to achieve since the test statistic depends on the entire path of Ψ and not only on the properties of $\mathbf{\Gamma}$ and $\mathbf{\Theta}$ at $s = 1$. For such a correction to work, one needs linear combinations of $\mathbf{\Gamma}$ and $\mathbf{\Theta}$ to have the same properties as $\mathbf{\Gamma}$ only. This, as can be easily checked, is the case when $\mathbf{\Gamma}$ and $\mathbf{\Theta}$ are Gaussian processes with covariance profile of the form $\eta(s)\Upsilon$ with $\eta(s)$ a suitable scalar function and Υ a constant positive definite matrix, but not in general. Should the correction be applicable, this works immediately for $\hat{Q}_n$, but becomes decisively more complex for $\tilde{Q}_n$ where the integral of $\mathbf{\Theta}$ over $[0, s]$ is a Gaussian process, but no Brownian motion.

Finally, since the analytical corrections may stop short of being straightforward, and sometimes nonlinear quadratic covariances need to be accounted for, bootstrap implementations of the tests suggest themselves to obtain asymptotically correctly-sized inference.

Since the effect depends also on the properties of estimator $\hat{\boldsymbol{\theta}}$ (in particular on $A_{t,n}$ or $B_{t,n}$), on which it is difficult to get more precise without becoming too model-specific, a thorough analysis of bootstrap validity is out of reach. Instead, we would rather like to emphasize some pitfalls associated to standard (block) i.i.d. and wild bootstrap schemes.

Denote by $\boldsymbol{X}^*_{t,b}$ the bootstrapped sample (which may be obtained either by bootstrapping $\boldsymbol{X}_t$, or by bootstrapping $\hat{\boldsymbol{Z}}_t$ or $\tilde{\boldsymbol{Z}}_t$ and filtering through an estimated version of $\boldsymbol{f}$). For testing, we shall assume that the null is suitably imposed when bootstrapping.[8]

Then, with "$\overset{p}{\Rightarrow}$" denoting weak convergence in probability and E^* expectation in the bootstrap population, the critical step would be to ensure that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\left(\boldsymbol{g}\left(\hat{\boldsymbol{Z}}^*_{t,b}\right)-\mathrm{E}^*\left(\boldsymbol{g}\left(\boldsymbol{Z}^*_{t,b}\right)\right)\right)\overset{p}{\Rightarrow}\Omega^{1/2}\mathbf{\Gamma}(s)+\boldsymbol{\tau}(s)\left(\Pi'(1)\,W\,\Pi(1)\right)^{-1}\Pi'(1)\,W\,\Sigma^{1/2}\mathbf{\Theta}(1)$$

for the full sample estimation, and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\left(\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}^*_{t,b}\right)-\mathrm{E}^*\left(\boldsymbol{g}\left(\boldsymbol{Z}^*_{t,b}\right)\right)\right)\overset{p}{\Rightarrow}\Omega^{1/2}\mathbf{\Gamma}(s)+\left(\int_0^s\mathbf{\Theta}'(r)\left(\Sigma^{1/2}\right)'W'\Pi(r)\left(\Pi'(r)\,W\,\Pi(r)\right)^{-1}\mathrm{d}\boldsymbol{\tau}'(r)\right)'$$

for recursive estimation. In other words, the bootstrapped partial sums should converge in a suitable mode (weakly in probability) to the same limit process as in Proposition 1, such that the residual effect is correctly replicated by the bootstrap procedure.

This, however, is not guaranteed with any bootstrap scheme. Consider e.g. the well-understood case of the i.i.d. bootstrap performed on $\boldsymbol{X}_t$. Then, the bootstrap samples do not replicate serial correlation or nonstationarities of the DGP. One could of course use the block bootstrap to side-step the first issue, and resort to the residual i.i.d. bootstrap, if the source of the nonstationarity lies in the filter or in the structure of the estimator.

If on the other hand the quantities $\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right)$ or $A_{t,n}$ are not stationary, but only piecewise locally stationary, one may have resort to wild or block wild bootstraps as suggested by Hansen (2000) or Zhou (2013) in related contexts. A seminal reference for this bootstrap is Wu (1986). This too is not always going to lead to valid results. To see why, take a basic example where $A_{t,n} = \boldsymbol{a}(\boldsymbol{X}_t)$. Then, wild bootstrapping $\boldsymbol{X}_t$ or

[8] This may not be difficult if constancy is of interest, but one may have to go at some lengths to impose say zero skewness in the bootstrap population.

$\hat{\boldsymbol{Z}}_t$ ($\tilde{\boldsymbol{Z}}_t$), even in block versions, does not produce the desired result in general: in an extreme case, $\boldsymbol{g}$ or $\boldsymbol{a}$ may be even functions such as the square, and using e.g. Rademacher random variables $R_{t,b}$ to generate bootstrap samples $\boldsymbol{X}^*_{t,b} = \boldsymbol{X}_t R_{t,b}$ would not give bootstrap sampling variability at all. But the issue is more subtle, because even if we don't use the Rademacher distribution, the covariance of $\boldsymbol{g}\left(\boldsymbol{X}^*_{t,b}\right)$ and $\boldsymbol{a}\left(\boldsymbol{X}^*_{t,b}\right)$ need not equal the covariance of $\boldsymbol{g}(\boldsymbol{X}_t)$ and $\boldsymbol{a}(\boldsymbol{X}_t)$.[9] (A related case of wild bootstrap failure is given in Brüggemann et al., 2016.) The solution here would be to block wild bootstrap $\boldsymbol{g}(\boldsymbol{Z}_t)$ and $A_{t,n}$ *jointly*, e.g. $(\boldsymbol{g}(\boldsymbol{X}_t), \boldsymbol{a}(\boldsymbol{X}_t))^* = (\boldsymbol{g}(\boldsymbol{X}_t), \boldsymbol{a}(\boldsymbol{X}_t)) R_{t,b}$. The bottom line is that bootstrapping without understanding the asymptotics of the residual effect is likely to fail.

4 Structural changes

Let $D_{t,\lambda} = \mathbb{I}(t/n > \lambda)$ for some nontrivial break time $\lambda \in (0,1)$ and write the model with breaks as outlined in Subsection 3.1,

$$\boldsymbol{h}_\lambda(\boldsymbol{\vartheta}) = \boldsymbol{h}(\boldsymbol{\theta}_1)(1 - D_{t,\lambda}) + \boldsymbol{h}(\boldsymbol{\theta}_2) D_{t,\lambda}$$

where $\boldsymbol{\vartheta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$. We only model one break at a common time for all parameters to avoid notational overhead.

In this model having formally $2M$ parameters, observations for $t < \lambda n$ are noninformative about $\boldsymbol{\theta}_2$ (and the other way round), so we make the convention that

$$\hat{\boldsymbol{\theta}}_{t,1} - \boldsymbol{\theta}_{1,0} = \begin{cases} \left(\sum_{j=1}^t B_{j,n}' W_n \sum_{j=1}^t B_{j,n}\right)^{-1} \sum_{j=1}^t B_{j,n}' W_n \sum_{j=1}^t A_{j,n} + R_{t,n} & t < \lambda n \\ \left(\sum_{j=1}^{\lambda n} B_{j,n}' W_n \sum_{j=1}^{\lambda n} B_{j,n}\right)^{-1} \sum_{j=1}^{\lambda n} B_{j,n}' W_n \sum_{j=1}^{\lambda n} A_{j,n} + R_{\lambda n,n} & t \geq \lambda n \end{cases}$$

and

$$\hat{\boldsymbol{\theta}}_{t,2} - \boldsymbol{\theta}_{2,0} = \begin{cases} 0 & t < \lambda n \\ \left(\sum_{j=\lambda n+1}^t B_{j,n}' W_n \sum_{j=\lambda n+1}^t B_{j,n}\right)^{-1} \sum_{j=\lambda n+1}^t B_{j,n}' W_n \sum_{j=\lambda n+1}^t A_{j,n} + R_{t,n} & t \geq \lambda n \end{cases}$$

where the components essentially obey Assumption 1 for the two subsamples, $1 \leq t < \lambda_0 n$ and $\lambda_0 n < t \leq n$. Since, in this formulation, the model has as parameter vector $\boldsymbol{\vartheta}$, this leads to a specific structure of the quantities of relevance; say $\boldsymbol{\Psi}_\lambda$, the analog of $\boldsymbol{\Psi}$ for the case with breaks, is given by

$$\boldsymbol{\Psi}_\lambda(s) = \begin{pmatrix} \boldsymbol{\Gamma}(s) \\ \boldsymbol{\Theta}(s)\,\mathbb{I}(s < \lambda) + \boldsymbol{\Theta}(\lambda)\,\mathbb{I}(s \geq \lambda) \\ (\boldsymbol{\Theta}(s) - \boldsymbol{\Theta}(\lambda))\,\mathbb{I}(s \geq \lambda) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}(s) \\ \boldsymbol{\Theta}_\lambda(s) \end{pmatrix},$$

while

$$\Pi_\lambda(s) = \begin{pmatrix} \Pi(s)\,\mathbb{I}(s < \lambda) + \Pi(\lambda)\,\mathbb{I}(s \geq \lambda) \\ (\Pi(s) - \Pi(\lambda))\,\mathbb{I}(s \geq \lambda) \end{pmatrix}$$

and the GMM weighting matrix $W_{n\lambda}$ has a block-diagonal structure,

$$W_{n\lambda} = \begin{pmatrix} W_n & 0 \\ 0 & W_n \end{pmatrix};$$

also,

$$\boldsymbol{\tau}_\lambda(s) = \begin{pmatrix} \boldsymbol{\tau}_{\boldsymbol{\theta}_1}(s)\,\mathbb{I}(s < \lambda) + \boldsymbol{\tau}_{\boldsymbol{\theta}_1}(\lambda)\,\mathbb{I}(s \geq \lambda) & (\boldsymbol{\tau}_{\boldsymbol{\theta}_2}(s) - \boldsymbol{\tau}_{\boldsymbol{\theta}_2}(\lambda))\,\mathbb{I}(s \geq \lambda) \end{pmatrix}$$

[9] Consider e.g. g the identity function and a the square function; then, unless $\mathrm{E}\left(R^3_{t,b}\right) = 1$, the wild bootstrap fails.

with obvious notation $\boldsymbol{\tau}_{\boldsymbol{\theta}_{1,2}}(s)$.

If the true break date λ_0 is known, Assumption 3 is not needed and Proposition 1 leads immediately to

Corollary 3 *Under the assumptions of Proposition 1, it holds as $T \to \infty$ that*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_{t,\lambda_0}\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \Rightarrow \Omega^{1/2}\boldsymbol{\Gamma}(s) + \boldsymbol{\tau}_{\lambda_0}(s) \left(\Pi'_{\lambda_0}(s)\, W_\lambda\, \Pi_{\lambda_0}(s)\right)^{-1} \Pi'_{\lambda_0}(s)\, W_\lambda\, \Sigma^{1/2}_{\lambda_0} \boldsymbol{\Theta}_{\lambda_0}(1)$$

and, on $[\epsilon, \lambda_0] \cup [\lambda_0 + \epsilon, 1]$ for any $0 < \epsilon < \min\{\lambda_0, 1 - \lambda_0\}$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_{t,\lambda_0}\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \Rightarrow \Omega^{1/2}\boldsymbol{\Gamma}(s) + \left(\int_0^s \boldsymbol{\Theta}'_{\lambda_0}(r) \left(\Sigma^{1/2}_{\lambda_0}\right)' W'_\lambda \Pi_{\lambda_0}(r) \left(\Pi'_{\lambda_0}(r)\, W_\lambda\, \Pi_{\lambda_0}(r)\right)^{-1} \mathrm{d}\boldsymbol{\tau}'_{\lambda_0}(r) \right)'.$$

When it comes to unknown break times, we may not treat an estimated λ the same way as an estimated $\boldsymbol{\theta}$ due to the discontinuity of the indicator function. It turns out, however, that plugging in an estimated λ, should its convergence rate be high enough (see e.g. Bai, 1997) is asymptotically equivalent to plugging in the true λ.

To establish this equivalence, we shall however need an additional assumption, since, in the cases cases where one has no knowledge on the true break date, one ends up using data from one regime to estimate the parameters of the other. E.g., the moment conditions $A_{j,n}$ need not have zero expectation anymore in the wrong regime, and $\boldsymbol{h}(\boldsymbol{X}_t, \ldots; \boldsymbol{\theta}) \neq \boldsymbol{Z}_t$ if $\boldsymbol{X}_t$ comes from the wrong regime, but we require minimal regularity conditions would side-step this problem if an estimated break time is close enough to the true one.

Assumption 3 *It holds that*

1. *$A_{j,n}$ is uniformly (in j, n) $L_{2+\alpha}$-bounded and $B_{j,n}$ is uniformly (in j, n) $L_{1+\alpha}$-bounded for some $\alpha > 0$;*
2. *$\sqrt{n} \sup_{s \in [\epsilon, \lambda_0] \cup [\lambda_0 + \epsilon, 1]} \left| R_{[sn],n} \right| \xrightarrow{p} 0$, $0 < \epsilon < \min\{\lambda_0, 1 - \lambda_0\}$; $\sqrt{n} \sup_{s \in [\lambda_0, \lambda_0 + \epsilon]} \left| R_{[sn],n} - R_{[\lambda_0 n],n} \right| \xrightarrow{p} 0$;*
3. *For $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_{1,2}$, $\max_{t=1,\ldots,n} \left\| \boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \ldots; \bar{\boldsymbol{\theta}}\right)\right) \right\| = o_p\left(\sqrt{n}\right)$ and $\max_{t=1,\ldots,n} \left\| \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z} = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots; \bar{\boldsymbol{\theta}}\right)} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}} \right\| = o_p(n)$;*
4. *For $\bar{\Phi}_n = \left\{\boldsymbol{\theta}^* : \left\|\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}\right\| < C n^{-1/2+\epsilon},\ 0 < \epsilon < 1/2,\ C > 0\right\}$, $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}_{1,2}$,*

$$\sup_{\boldsymbol{\theta}^*_t \in \bar{\Phi}_n; t=1,\ldots,n} \left\| \left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z} = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots; \boldsymbol{\theta}^*_t\right)} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*_t} - \left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z} = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots; \bar{\boldsymbol{\theta}}\right)} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}} \right\| \xrightarrow{p} 0.$$

We also introduce some extra notation. Namely, $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ depend on the assumed break time, so we make this dependence explicit by writing $\hat{\boldsymbol{\theta}}_1(\lambda)$ etc. for $\lambda = \lambda_0$ or $\lambda = \hat{\lambda}$. They lead to residuals $\hat{\boldsymbol{Z}}_t(\lambda)$ and $\tilde{\boldsymbol{Z}}_t(\lambda)$. We examine the difference between the partial sums of $\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_{t,\lambda_0}\right)$ and $\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_{t,\hat{\lambda}}\right)$ in the following

Proposition 2 *Let $\hat{\lambda} = \lambda_0 + O_p\left(n^{-1}\right)$ and $0 < \underline{\lambda} \leq \hat{\lambda} \leq \overline{\lambda} < 1$ a.s. Then, under Assumptions 1 and 3, it holds as $T \to \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_{t,\hat{\lambda}}\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \Rightarrow \Omega^{1/2}\boldsymbol{\Gamma}(s) + \boldsymbol{\tau}_{\lambda_0}(s) \left(\Pi'_{\lambda_0}(s)\, W\, \Pi_{\lambda_0}(s)\right)^{-1} \Pi'_{\lambda_0}(s)\, W\, \Sigma^{1/2}_{\lambda_0} \boldsymbol{\Theta}_{\lambda_0}(1)$$

and, on $[\epsilon, \lambda_0] \cup [\lambda_0 + \epsilon, 1]$ *for any* $0 < \epsilon < \min\{\lambda_0, 1 - \lambda_0\}$*,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_{t,\hat{\lambda}}\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right) \right) \Rightarrow \Omega^{1/2}\boldsymbol{\Gamma}(s) + \left(\int_0^s \boldsymbol{\Theta}'_{\lambda_0}(r) \left(\Sigma_{\lambda_0}^{1/2}\right)' W' \Pi_{\lambda_0}(r) \left(\Pi'_{\lambda_0}(r) W \Pi_{\lambda_0}(r)\right)^{-1} \mathrm{d}\boldsymbol{\tau}'_{\lambda_0}(r) \right)'.$$

Proof: *See the Appendix.*

These are the same limits as in Corollary 3 so the effect of plugging in an estimated break time is indeed asymptotically negligible.

Remark 3 *Also Dette and Wied (2016) make use of the fact that certain limit distributions do not change if one replaces the true breakpoint fraction* t *with an estimator which converges faster than* $\sqrt{n}$ *to* t*. Dette and Wied (2016) propose tests for relevant changes in time series models based on a CUSUM-approach. Their tests are based on the integral of certain differences between estimated moments. The variance of the integral depends on* t *and the convergence rate of the integral is* $\sqrt{n}$ *so that* $\hat{t} - t$ *must be* $o_p(\sqrt{n})$*.*

Remark 4 *Should there be no break, the break time estimator can typically be shown to converge in distribution, and the weak limit in Corollary 3 changes; note that a different limiting distribution of statistics of interest would arise (one taking the behavior of* $\hat{\lambda}$ *into account). Since we explicitly model a break, we don't pursue this topic here.*

Remark 5 *Sofar,* $\boldsymbol{g}$ *has been assumed to be smooth. Since we focus on capturing structural breaks, this is a natural assumption to make. We may however speculate as to what happens if* $\boldsymbol{g}$ *is only piecewise smooth. Assuming e.g. continuity of* $\boldsymbol{g}$ *with jump discontinuity its the derivatives, it should suffice to assume continuous density of* $\boldsymbol{Z}_t$ *to ensure that the results still hold. If letting* $\boldsymbol{g}$ *itself exhibit a jump discontinuity, one may formally apply the result from Proposition 1 to conclude that the density of* $\boldsymbol{Z}_t$ *at the discontinuity plays a role in quantifying the effect;*[10] *we leave this topic for further research.*

5 Some examples

5.1 Examples without breaks

Let us first consider testing hypotheses about the higher-order moments of a (univariate latent) i.i.d. series Z_t in a location-scale model,

$$X_t = \mu + \sigma Z_t \quad \text{with} \quad Z_t \sim iid\,(0, 1)\,.$$

Given that we work under iid sampling, the assumptions in Section 3 can easily be shown to hold, provided that enough moments of Z_t are finite and the parameter space is compact, so we do not spell out the details here to save space.

Letting

$$\hat{Z}_t = \frac{X_t - \hat{\mu}}{\hat{\sigma}} \quad \text{with} \quad \hat{\sigma}^2 = \frac{1}{n}\sum (X_t - \hat{\mu})^2 \quad \text{and} \quad \hat{\mu} = \bar{X},$$

we may test hypotheses about the skewness μ_3 of Z_t (or equivalently the standardized skewness of X_t) building on the statistic

$$\mathcal{T} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\hat{Z}_t^3 - \mu_3^{(0)} \right).$$

[10] This is less of a trick than one might think; see Phillips (1991).

The relevant quantities are

$$g(z) = z^3, \qquad \boldsymbol{\theta} = \left(\mu, \sigma^2\right)' \qquad \text{and} \qquad h(x) = \frac{x - \theta_1}{\sqrt{\theta_2}},$$

such that

$$\frac{\partial g}{\partial z} = 3z^2 \qquad \text{and} \qquad \frac{\partial h}{\partial \boldsymbol{\theta}} = \left(-\frac{1}{\sqrt{\theta_2}}, -\frac{1}{2}\frac{x - \theta_1}{\theta_2^{3/2}}\right),$$

leading to

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{[ns]} \left.\frac{\partial g}{\partial z}\right|_{z=Z_t} \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= \frac{1}{n}\sum_{t=1}^{[ns]} 3Z_t^2 \left(-\frac{1}{\sigma_0}, -\frac{1}{2}\frac{Z_t}{\sigma_0^2}\right) \\ &\Rightarrow -3s\left(\frac{1}{\sigma_0}, \frac{\mu_{3,0}}{2\sigma_0^2}\right) \equiv \boldsymbol{\tau}(s). \end{aligned}$$

Hence

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]} \left(\hat{Z}_t^3 - \mu_3^{(0)}\right) \Rightarrow \Omega^{1/2}\Gamma(s) - 3s\left(\frac{1}{\sigma_0}, \frac{\mu_{3,0}}{2\sigma_0^2}\right) \Sigma^{1/2}\boldsymbol{\Theta}(1)$$

where

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]} \begin{pmatrix} Z_t^3 - \mu_3 \\ \sigma Z_t \\ \sigma^2 Z_t^2 - \sigma^2 \end{pmatrix} \Rightarrow \boldsymbol{\Psi}(s) \equiv \begin{pmatrix} \Omega^{1/2}\Gamma(s) \\ \Sigma^{1/2}\boldsymbol{\Theta}(s) \end{pmatrix}$$

with $\boldsymbol{\Psi}$ a Brownian motion with quadratic covariation process

$$[\boldsymbol{\Psi}](s) = s \begin{pmatrix} \mu_{6,0} - \mu_{3,0}^2 & \sigma_0\mu_{4,0} & \sigma_0^2(\mu_{5,0} - \mu_{3,0}) \\ \sigma_0\mu_{4,0} & \sigma_0^2 & \sigma_0^3\mu_{3,0} \\ \sigma_0^2(\mu_{5,0} - \mu_{3,0}) & \sigma_0^3\mu_{3,0} & \sigma_0^4(\mu_{4,0} - 1) \end{pmatrix},$$

hence $\Omega = \mu_{6,0} - \mu_{3,0}^2$, $\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_0^3\mu_{3,0} \\ \sigma_0^3\mu_{3,0} & \sigma_0^4(\mu_{4,0} - 1) \end{pmatrix}$ and $\Lambda = \begin{pmatrix} \sigma\mu_4 \\ \sigma_0^2(\mu_{5,0} - \mu_{3,0}) \end{pmatrix}$. Also, $\Pi(s) = sI_2$ is this case, as we deal with estimators that are essentially sample averages. (This is the case for the following examples as well.)

We note that demeaning always has an effect on the partial sums, but whether estimating the variance has an effect or not depends explicitly on the true skewness $\mu_{3,0}$ of the considered DGP. If one is interested in testing the constancy of the skewness, both effects cancel out in the statistic according to Corollary 2.

Note also that Jarque and Bera (1980) claim that there is no effect when testing the null of normality in the Pearson family of distributions. Jarque and Bera (1980, p. 257) indicate $m_3^2/6m_2^3$ as unfeasible statistic, with $m_k = n^{-1}\sum_{t=1}^{n} Z_t^k$, and the analog $\hat{m}_3^2/6\hat{m}_2^3$, with $\hat{m}_k = n^{-1}\sum_{t=1}^{n} \hat{Z}_t^k$, as residual-based one. So, as it is known that the residual-based statistic works, their conclusion seems correct. However, since the 6th centered moment of the normal distribution is $15\sigma^6$, it is immediately seen that the statistic $m_3^2/6m_2^3$ is not χ_1^2 in the limit (and the correct unfeasible statistic would have been $m_3^2/15m_2^3$), so the residual effect is actually present, as discussed above.

Now, for testing the kurtosis of Z_t, h is the same but

$$g(z) = z^4 \qquad \text{and} \qquad \frac{\partial g}{\partial z} = 4z^3,$$

such that

$$\frac{1}{n}\sum_{t=1}^{[ns]}\left.\frac{\partial g}{\partial z}\right|_{z=Z_t}\left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n}\sum_{t=1}^{[ns]}4Z_t^3\left(-\frac{1}{\sigma_0},-\frac{1}{2}\frac{Z_t}{\sigma_0^2}\right)$$
$$\Rightarrow -4s\left(\frac{\mu_3}{\sigma_0},\frac{\mu_4}{2\sigma_0^2}\right)\equiv\boldsymbol{\tau}(s).$$

The process $\boldsymbol{\Psi}(s)$ (in particular the component $\Gamma(s)$) is different,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]}\left(\begin{array}{c} Z_t^4-\mu_{4,0} \\ \sigma_0 Z_t \\ \sigma_0^2 Z_t^2-\sigma_0^2\end{array}\right)\Rightarrow\boldsymbol{\Psi}(s),$$

having a different quadratic covariation,

$$[\boldsymbol{\Psi}](s)=s\left(\begin{array}{ccc}\mu_{8,0}-\mu_{4,0}^2 & \sigma_0\mu_{5,0} & \sigma_0^2(\mu_{6,0}-\mu_{4,0}) \\ \sigma_0\mu_{5,0} & \sigma_0^2 & \sigma_0^3\mu_{3,0} \\ \sigma_0^2(\mu_{6,0}-\mu_{4,0}) & \sigma_0^3\mu_{3,0} & \sigma_0^4(\mu_{4,0}-1)\end{array}\right).$$

Contrary to the case of the skewness, estimating the variance has an effect on the partial sums irrespective of the skewness, but the actual skewness $\mu_{3,0}$ controls now whether demeaning has an effect. Again, if interested in the constancy of the kurtosis, both effects cancel out and the asymptotics is not affected by the residual effect.

These are more or less familiar cases that have been extensively discussed in the literature (see e.g. Bai and Ng, 2005). Let us now put some bivariate cases into perspective, say the covariance of some bivariate $\boldsymbol{X}_t$ which has unknown mean but only the covariance (matrix) is subject to inference. Then,

$$g(\boldsymbol{z})=z_1z_2,\qquad \hat{\boldsymbol{Z}}_t=\boldsymbol{X}_t-\bar{\boldsymbol{X}}_t\qquad\text{and}\qquad \boldsymbol{h}(x)=\left(\begin{array}{c}x_1-\theta_1\\ x_2-\theta_2\end{array}\right)$$

with $\hat{\theta}_1=\hat{\mu}_1$ and $\hat{\theta}_2=\hat{\mu}_2$. Hence

$$\frac{\partial g}{\partial z_1}=z_2\quad\frac{\partial g}{\partial z_2}=z_1,\qquad\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}=\left(\begin{array}{cc}-1 & 0\\ 0 & -1\end{array}\right),$$

leading to

$$\frac{1}{n}\sum_{t=1}^{[ns]}\left.\frac{\partial g}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{n}\sum_{t=1}^{[ns]}(-Z_{t2},-Z_{t1})$$
$$\Rightarrow \mathbf{0}$$

such that here the distribution is not asymptotically affected compared to the test based on $Z_{t,1}Z_{t,2}$ directly.

Then again, if looking at the correlation ρ rather than the covariance of Z_{t1} and Z_{t2}, the residual effect is present. We have like before

$$g(\boldsymbol{z})=z_1z_2,$$

but, for $i=1,2$, we have that

$$\hat{Z}_{ti}=\frac{X_{ti}-\hat{\mu}_i}{\hat{\sigma}_i}$$

with $\hat{\mu}_i = \bar{X}_i$ and

$$\hat{\sigma}_i^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_{ti} - \bar{X}_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\left(Z_{ti} - \bar{Z}_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 Z_{ti}^2 + O_p\left(n^{-1}\right),$$

such that, with $\theta_3 = \sigma_1^2$ and $\theta_4 = \sigma_2^2$, we write

$$\boldsymbol{h}\left(\boldsymbol{x}\right) = \begin{pmatrix} \frac{x_1 - \theta_1}{\sqrt{\theta_3}} \\ \frac{x_2 - \theta_2}{\sqrt{\theta_4}} \end{pmatrix}.$$

While $\frac{\partial g}{\partial \boldsymbol{z}}$ is the same as in the case of the covariance,

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -\frac{1}{\sigma_1} & 0 & -\frac{1}{2}\frac{x_1-\mu_1}{\sigma_1^3} & 0 \\ 0 & -\frac{1}{\sigma_2} & 0 & -\frac{1}{2}\frac{x_2-\mu_2}{\sigma_2^3} \end{pmatrix}$$

such that

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{[ns]} \left.\frac{\partial g}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= \frac{1}{n}\sum_{t=1}^{[ns]} \left(Z_{t2}, Z_{t1}\right) \begin{pmatrix} -\frac{1}{\sigma_{1,0}} & 0 & -\frac{1}{2}\frac{Z_{t1}}{\sigma_{1,0}^2} & 0 \\ 0 & -\frac{1}{\sigma_{2,0}} & 0 & -\frac{1}{2}\frac{Z_{t2}}{\sigma_{2,0}^2} \end{pmatrix} \\ &\Rightarrow -\rho_0\, s \begin{pmatrix} 0 & 0 & \frac{1}{2\sigma_{1,0}^2} & \frac{1}{2\sigma_{2,0}^2} \end{pmatrix} \equiv \boldsymbol{\tau}\left(s\right) \end{aligned}$$

and the variance estimation matters whenever the correlation is nonzero, but not the demeaning. Kicking out the irrelevant zero elements, $\boldsymbol{\tau}\left(s\right) = -\rho_0\, s \begin{pmatrix} \frac{1}{2\sigma_{1,0}^2} & \frac{1}{2\sigma_{2,0}^2} \end{pmatrix}$ and the relevant Brownian motion is

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]} \begin{pmatrix} Z_{t1}Z_{t2} - \rho_0 \\ \sigma_{1,0}^2 Z_{t1}^2 - \sigma_{1,0}^2 \\ \sigma_{2,0}^2 Z_{t2}^2 - \sigma_{2,0}^2 \end{pmatrix} \Rightarrow \boldsymbol{\Psi}\left(s\right)$$

with quadratic covariation

$$\left[\boldsymbol{\Psi}\right]\left(s\right) = s\begin{pmatrix} \mathrm{E}\left(Z_{t1}^2 Z_{t2}^2\right) - \rho_0^2 & \sigma_{1,0}^2\left(\mathrm{E}\left(Z_{t1}^3 Z_{t2}\right) - \rho_0\right) & \sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1} Z_{t2}^3\right) - \rho_0\right) \\ \sigma_{1,0}^2\left(\mathrm{E}\left(Z_{t1}^3 Z_{t2}\right) - \rho_0\right) & \sigma_{1,0}^4\left(\mu_{4,1,0} - 1\right) & \sigma_{1,0}^2\sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1}^2 Z_{t2}^2\right) - 1\right) \\ \sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1} Z_{t2}^3\right) - \rho_0\right) & \sigma_{1,0}^2\sigma_{2,0}^2\left(\mathrm{E}\left(Z_{t1}^2 Z_{t2}^2\right) - 1\right) & \sigma_{2,0}^4\left(\mu_{4,2,0} - 1\right) \end{pmatrix}.$$

If interested in tests on constant correlation, $\boldsymbol{\tau}$ is linear in s so the estimation effect cancels out.

5.2 Examples with breaks

Let us now consider situations where there is a break in the mean. Concretely, let

$$X_t = \mu_{0,1}\mathbb{I}\left(t < \lambda_0 n\right) + \mu_{0,2}\mathbb{I}\left(t \geq \lambda_0 n\right) + Z_t, \qquad Z_t \sim iid\left(0, \sigma^2\right),$$

and test simple hypotheses on $\mathrm{E}\left(Z_t^2\right)$. Therefore, $g\left(z\right) = z^2$ and

$$\hat{Z}_t = Z_t - \left(\hat{\mu}_1 - \mu_{1,0}\right)\left(1 - D_{t,\lambda_0}\right) + \left(\hat{\mu}_2 - \mu_{2,0}\right) D_{t,\lambda_0}.$$

Then, $\frac{\partial g}{\partial z} = 2z$, $\frac{\partial h}{\partial \boldsymbol{\theta}} = \begin{pmatrix} -\left(1 - D_{t,\lambda_0}\right) & -D_{t,\lambda_0} \end{pmatrix}$ such that

$$\frac{1}{n}\sum_{t=1}^{[sn]} \left.\frac{\partial g}{\partial z}\right|_{z=Z_t} \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{2}{n}\sum_{t=1}^{[sn]} Z_t \begin{pmatrix} -\left(1 - D_{t,\lambda_0}\right) & -D_{t,\lambda_0} \end{pmatrix} \Rightarrow \begin{pmatrix} 0 & 0 \end{pmatrix}$$

and there is no residual effect from piecewise demeaning on the variance test.

This extends to the case of tests on the correlation if the breaks accounted for are only in the mean but not in the variance as follows. Let

$$\boldsymbol{X}_t = \boldsymbol{\mu}_{1,0}\left(1 - D_{t,\lambda_0}\right) + \boldsymbol{\mu}_{2,0} D_{t,\lambda_0} + \left(\begin{array}{cc} \sigma_{1,0} & 0 \\ 0 & \sigma_{2,0} \end{array}\right) \boldsymbol{Z}_t$$

with λ_0 known. We still have

$$g\left(\boldsymbol{z}\right) = z_1 z_2,$$

but

$$\hat{Z}_{ti} = \frac{X_{ti} - \hat{\mu}_{1,i}\left(1 - D_{t,\lambda}\right) - \hat{\mu}_{2,i} D_{t,\lambda}}{\hat{\sigma}_i} \tag{4}$$

such that, with $\theta_1 = \mu_1$, $\theta_2 = \mu_2$, $\theta_3 = \sigma_1^2$ and $\theta_4 = \sigma_2^2$, and defining for brevity $\bar{D}_{t,\lambda} = 1 - D_{t,\lambda}$, we obtain

$$\boldsymbol{h}_\lambda\left(\boldsymbol{x}\right) = \left(\begin{array}{c} \frac{x_1 - \theta_1 \bar{D}_{t,\lambda} - \theta_2 D_{t,\lambda}}{\sqrt{\theta_5}} \\ \frac{x_2 - \theta_3 \bar{D}_{t,\lambda} - \theta_4 D_{t,\lambda}}{\sqrt{\theta_6}} \end{array}\right).$$

While $\frac{\partial g}{\partial \boldsymbol{z}} = (z_2, z_1)$, we now have

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}} = -\left(\begin{array}{cccccc} \frac{1}{\sigma_1}\bar{D}_{t,\lambda} & \frac{1}{\sigma_1} D_{t,\lambda} & 0 & 0 & \frac{1}{2}\frac{x_1 - \mu_{1,1}\bar{D}_{t,\lambda} - \mu_{2,1} D_{t,\lambda}}{\sigma_1^3} & 0 \\ 0 & 0 & \frac{1}{\sigma_2}\bar{D}_{t,\lambda} & \frac{1}{\sigma_2} D_{t,\lambda} & 0 & \frac{1}{2}\frac{x_2 - \mu_{2,1} - \mu_{2,2} D_{t,\lambda}}{\sigma_2^3} \end{array}\right),$$

hence

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{[ns]} \left.\frac{\partial g}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= -\frac{1}{n}\sum_{t=1}^{[ns]} (Z_{t2}, Z_{t1}) \left(\begin{array}{cccccc} \frac{1}{\sigma_{1,0}}\bar{D}_{t,\lambda_0} & \frac{1}{\sigma_{1,0}} D_{t,\lambda_0} & 0 & 0 & \frac{1}{2}\frac{Z_{t1}}{\sigma_{1,0}^2} & 0 \\ 0 & 0 & \frac{1}{\sigma_{2,0}}\bar{D}_{t,\lambda_0} & \frac{1}{\sigma_{2,0}} D_{t,\lambda_0} & 0 & \frac{1}{2}\frac{Z_{t2}}{\sigma_{2,0}^2} \end{array}\right) \\ &\Rightarrow -\rho_0\, s \left(\begin{array}{cccccc} 0 & 0 & 0 & 0 & \frac{1}{2\sigma_{1,0}^2} & \frac{1}{2\sigma_{2,0}^2} \end{array}\right) \equiv \boldsymbol{\tau}_{\lambda_0}\left(s\right) \end{aligned}$$

and only the variance estimation has an effect on the limiting behavior of the partial sums, which would cancel out if testing the constancy of the correlation. The relevant Brownian motion is the same as for demeaning only, and breaks in the mean (accounted for) do not matter for testing the correlation either.

Finally, if allowing for a break in the variance, say a model

$$\boldsymbol{X}_t = \left(\begin{array}{cc} \sqrt{\sigma_{1,1}^2\left(1 - D_{t,\lambda}\right) + \sigma_{1,2}^2 D_{t,\lambda}} & 0 \\ 0 & \sqrt{\sigma_{2,1}^2\left(1 - D_{t,\lambda}\right) + \sigma_{2,2}^2 D_{t,\lambda}} \end{array}\right) \boldsymbol{Z}_t$$

(for simplicity with known zero mean since demeaning does not have an asymptotic effect in this setup), we obtain

$$\hat{Z}_{ti} = \frac{X_{ti}}{\sqrt{\hat{\sigma}_{i,1}^2\left(1 - D_{t,\lambda}\right) + \hat{\sigma}_{i,2}^2 D_{t,\lambda}}} \qquad \text{and} \quad \boldsymbol{h}\left(\boldsymbol{x}\right) = \left(\begin{array}{c} \frac{x_1}{\sqrt{\theta_1 \bar{D}_{t,\lambda} + \theta_2 D_{t,\lambda}}} \\ \frac{x_2}{\sqrt{\theta_3 \bar{D}_{t,\lambda} + \theta_4 D_{t,\lambda}}} \end{array}\right)$$

and consequently

$$\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}} = -\frac{1}{2}\left(\begin{array}{cccc} \frac{x_1 \bar{D}_{t,\lambda}}{\left(\sigma_{1,1}^2 \bar{D}_{t,\lambda} + \sigma_{1,2}^2 D_{t,\lambda}\right)^{3/2}} & \frac{x_1 D_{t,\lambda}}{\left(\sigma_{1,1}^2 \bar{D}_{t,\lambda} + \sigma_{1,2}^2 D_{t,\lambda}\right)^{3/2}} & 0 & 0 \\ 0 & 0 & \frac{x_1 \bar{D}_{t,\lambda}}{\left(\sigma_{2,1}^2 \bar{D}_{t,\lambda} + \sigma_{2,2}^2 D_{t,\lambda}\right)^{3/2}} & \frac{x_1 D_{t,\lambda}}{\left(\sigma_{2,1}^2 \bar{D}_{t,\lambda} + \sigma_{2,2}^2 D_{t,\lambda}\right)^{3/2}} \end{array}\right).$$

Then,

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{[ns]} \left.\frac{\partial g}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial h}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\
&= -\frac{1}{2n}\sum_{t=1}^{[ns]} \left(Z_{t2}, Z_{t1}\right) \begin{pmatrix} \frac{Z_{t1}\bar{D}_{t,\lambda_0}}{\sigma^2_{1,1,0}\bar{D}_{t,\lambda_0}+\sigma^2_{1,2,0}D_{t,\lambda_0}} & \frac{Z_{t1}D_{t,\lambda_0}}{\sigma^2_{1,1,0}\bar{D}_{t,\lambda_0}+\sigma^2_{1,2,0}D_{t,\lambda_0}} & 0 & 0 \\ 0 & 0 & \frac{Z_{t2}\bar{D}_{t,\lambda_0}}{\sigma^2_{2,1,0}\bar{D}_{t,\lambda_0}+\sigma^2_{2,2,0}D_{t,\lambda_0}} & \frac{Z_{t2}D_{t,\lambda_0}}{\sigma^2_{2,1,0}\bar{D}_{t,\lambda_0}+\sigma^2_{2,2,0}D_{t,\lambda_0}} \end{pmatrix} \\
&\Rightarrow -\frac{\rho_0}{2}\begin{pmatrix} \frac{\mathbb{I}(s<\lambda_0)}{\sigma^2_{1,1,0}}s & \frac{\mathbb{I}(s\geq\lambda_0)}{\sigma^2_{1,2,0}}(s-\lambda_0) & \frac{\mathbb{I}(s<\lambda_0)}{\sigma^2_{2,1,0}}s & \frac{\mathbb{I}(s\geq\lambda_0)}{\sigma^2_{2,2,0}}(s-\lambda_0) \end{pmatrix} \equiv \boldsymbol{\tau}_{\lambda_0}(s)
\end{aligned}
$$

which is piecewise linear for $s \in [0,1]$. Hence the effect of accounting for breaks in the variance is not negligible when concerned about the correlation, not even when testing the constancy, unless $\rho_0 = 0$. The corresponding process is also not a Brownian motion,

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{[sn]} \begin{pmatrix} Z_{t1}Z_{t2}-\rho_0 \\ \sigma^2_{1,1,0}\left(Z^2_{t1}-1\right)\left(1-D_{t,\lambda_0}\right) \\ \sigma^2_{1,2,0}\left(Z^2_{t1}-1\right)D_{t,\lambda_0} \\ \sigma^2_{2,1,0}\left(Z^2_{t2}-1\right)\left(1-D_{t,\lambda_0}\right) \\ \sigma^2_{2,0,2}\left(Z^2_{t2}-1\right)D_{t,\lambda_0} \end{pmatrix} \Rightarrow \boldsymbol{\Psi}_{\lambda_0}(s) \equiv \begin{pmatrix} \Omega^{1/2}\Gamma(s) \\ \Sigma^{1/2}\boldsymbol{\Theta}_{\lambda_0}(s) \end{pmatrix}.
$$

This motivates us in defining a new test for constant correlations, since the above one accounts for breaks in variances and means in a perhaps more convenient way. Concretely, compute (2) with

$$
g\left(\hat{Z}_{t1}, \hat{Z}_{t2}\right) = \hat{Z}_{t1}\hat{Z}_{t2} \qquad \text{and} \qquad \hat{Z}_{ti} = \frac{X_{ti} - \hat{\mu}_{1,i}\left(1-D_{t,\lambda}\right) - \hat{\mu}_{2,i}D_{t,\lambda}}{\sqrt{\hat{\sigma}^2_{i,1}\left(1-D_{t,\lambda}\right) + \hat{\sigma}^2_{i,2}D_{t,\lambda}}}. \tag{5}
$$

Since the residual effect is difficult to account for analytically in this case, we resort to a bootstrap procedure and for this reason we may use as standardizing matrix $\hat{\Omega}$ the sample variance of $\hat{Z}_{t1}\hat{Z}_{t2}$. The change point is either known, $\lambda = \lambda_0$, or can be estimated superconsistently, $\lambda = \hat{\lambda}$. We analyze the behavior of the new, robust test in the following subsection and use it with real data in Section 6.

5.3 Experimental evidence on the robustified constant correlation test

5.3.1 Robustness with respect to non-constant variances

In this subsection, we analyze the finite-sample behavior of the test for constant correlation if the marginal variances are time-varying. A simulation study illustrates the robustness with respect to non-constant variances of our new test in contrast to the non-robust Wied et al. (2012)-test. Moreover, we will see that the new test has considerable power in finite samples. The new robustified test is based on (2) in combination with (5) but without demeaning in the numerator as we generate the series with zero mean.

First, for analyzing the size properties, we generate independent data from a bivariate normal distribution with constant correlation 0.4. The marginal variances are 1 in the first half of the sample and take the values $\{0.1, 0.2, \ldots, 1.9, 2\}$ in the second part of the sample. The sample size is 500 and we use 10000 Monte Carlo replications. The critical values of our new test are obtained by an i.i.d. bootstrap based on drawing with replacement from the joint empirical distribution of the demeaned X_{t1} and X_{t2}; for this, we use 199 bootstrap repetitions to keep the computational effort to a minimum. After being drawn, the bootstrap samples are transformed as follows: the univariate series are split into two parts based on the estimated variance change points in the original sample and both parts are variance-standardized such that they have the same empirical variance as the original series. We consider both the case of estimated and of true variance change point locations. In both cases, we take the validity of the block bootstrap to be granted.

The plot of the empirical sizes is given in Figure 1. One sees that our test generally keeps its size, in particular also if the variance change point locations are estimated. Practically, there are no differences between the test with true and the test with estimated locations, although the size is marginally lower in the latter case if the true variances do not change. The size of the nonrobust Wied et al. (2012)-test is smaller than α in the case of decreasing variances and larger than α in the case of increasing variances. The intuition to this comes from the structure of the non-robust test in which successively estimated correlations are compared. In the extreme case that the variances are zero in the second part, the recursive correlations do not change any more after the middle. So, the supremum of the correlation differences is attained only in the first half of the sample, which leads to a smaller test statistic. On the other hand, if the variances are extremely large in the second half, there is an extreme, sudden shift towards ± 1 in the successively estimated correlations slightly after the middle. The mechanism leading to this behavior is ultimately the sensitivity of the empirical correlation coefficient with respect to outliers. This peak leads to a high test statistic and thus to higher rejection rates.

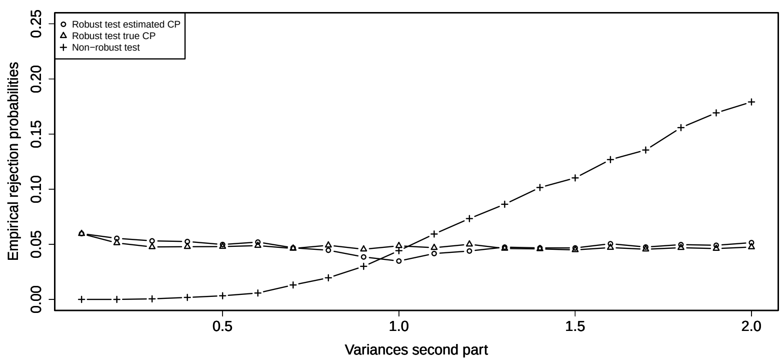

Figure 1: Empirical rejection probabilities of the non-robust and the robust test in a setting with constant cross-correlations and non-constant marginal variances

Figure 2 shows the empirical power of both tests in a setting under which the Wied et al. (2012)-test works, i.e., we generate i.i.d. data from a zero-mean bivariate normal distribution with constant unity marginal variances. The cross-correlation is 0.4 in the first half of the sample and takes the values $\{-0.4, -0.3, \ldots, 0.7, 0.8\}$ in the second part of the sample. One sees that the power of both tests is rather similar, although, not surprisingly, robustifying has a minor cost in terms of power for changes to higher values of the correlation coefficient. Again, there are practically no consequences of plugging in an estimated break time.

Figure 3 shows the empirical power of both tests in a setting under which the Wied et al. (2012)-test does not work, i.e. we generate independent data from a bivariate normal distribution with zero mean and constant marginal variances 1 in the first half and 2 in the second half of the sample. The cross-correlation is 0.4 in the first half of the sample and takes the values $\{-0.4, -0.3, \ldots, 0.7, 0.8\}$ in the second part of the sample. One sees that our new test has high power in the case of a large jump. The non-robust test has higher rejection frequencies than the new test but, of course, it must be taken into account that it is quite oversized.

5.3.2 Robustness with respect to non-constant expectations

This subsection repeats the analysis from the last subsection, but with a focus on non-constant expectations and not on non-constant variances. This means that the residuals of our new robust test are obtained

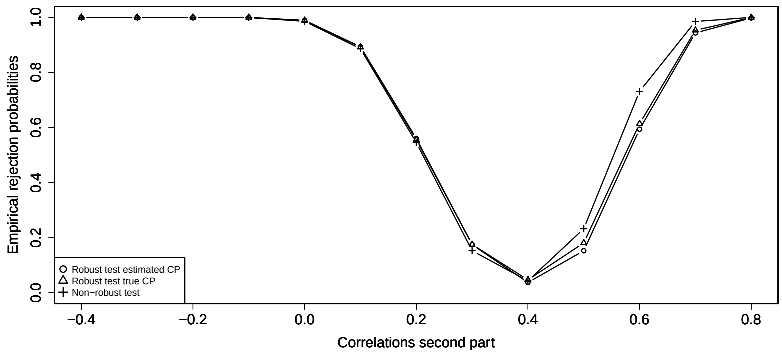

Figure 2: Empirical rejection probabilities of the non-robust and the robust test in a setting with changing cross-correlations and constant marginal variances

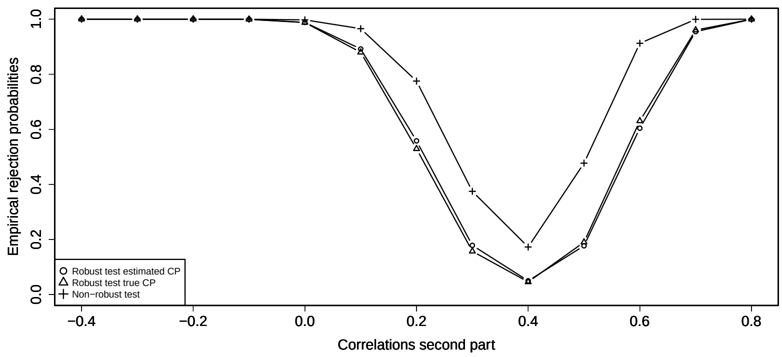

Figure 3: Empirical rejection probabilities of the non-robust and the robust test in a setting with changing cross-correlations and changing marginal variances

by filtering out change points in the first moment, i.e. use residuals from (4). Since this does not induce a residual effect (see Subsection 5.2), we do not have to use a bootstrap approximation. Instead, the asymptotic distribution of our test statistic is $\sup_{s\in[0,1]} |B(s)|$, where $B(\cdot)$ is a Brownian bridge. For significance level $\alpha = 0.05$, the critical value is 1.358.

At first, we analyze the size in a setting in which the variances are constant, equal to 1, and the expectations take the value 0 in the first half and $\{-1, -0.9, \ldots, 0.9, 1\}$ in the second half of the sample.

The results are plotted in Figure 4. More so than in the variance case, Figure 1, estimating the change point makes no difference in the robust test's behavior; the test is slightly conservative in both cases. The Wied et al. (2012)-case is oversized if the expectations decrease or increase.

Figure 5 compares (in a way similar to Figure 2) the robust and nonrobust tests in a setting with constant expectation zero. As in Figure 4, estimating change point locations does not make any difference compared to using the true change point locations. However, the Wied et al. (2012)-test performs relatively better when the change in the correlation is upwards (cf. Figure 2).

Finally, Figure 6 (in a similar way as Figure 3) shows the empirical power of both tests in a setting under which the Wied et al. (2012)-test does not work, i.e., the expectations are zero in the first half and unity in the second half of the sample. The result is, at first sight, quite interesting: While our new robust test has considerable power, which increases with the difference of the correlation in the second half of the sample (cf. Figure 5), the power curve of the Wied et al. (2012)-test has a minimum at 0.1. One must of course consider that the Wied et al. (2012)-test rejects almost every time under the null for the unity jump in the

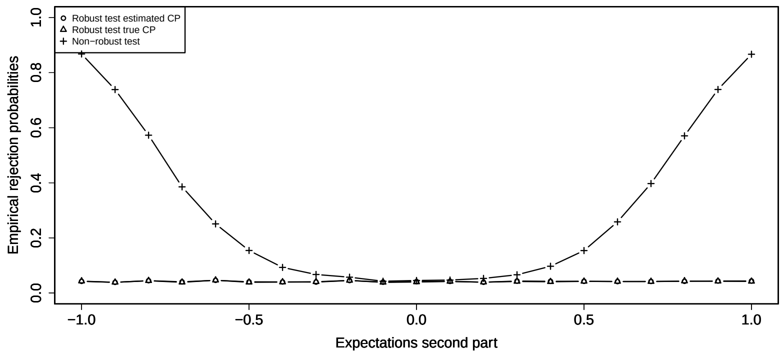

Figure 4: Empirical rejection probabilities of the non-robust and the robust test in a setting with constant cross-correlations and non-constant marginal expectations

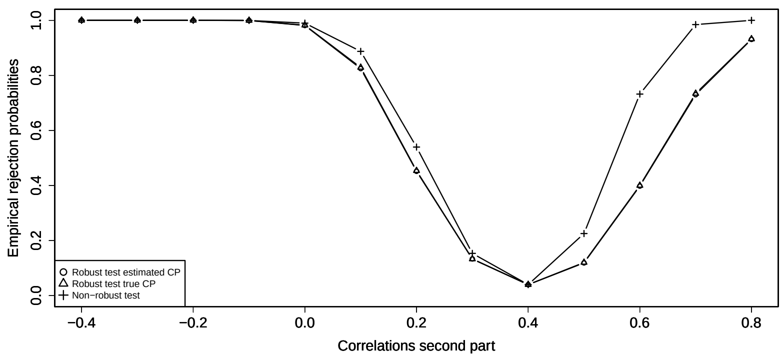

Figure 5: Empirical rejection probabilities of the non-robust and the robust test in a setting with changing cross-correlations and constant marginal expectations

mean, so it is actually not surprising that the non-robust test, in addition to not controlling size, is also severely biased.

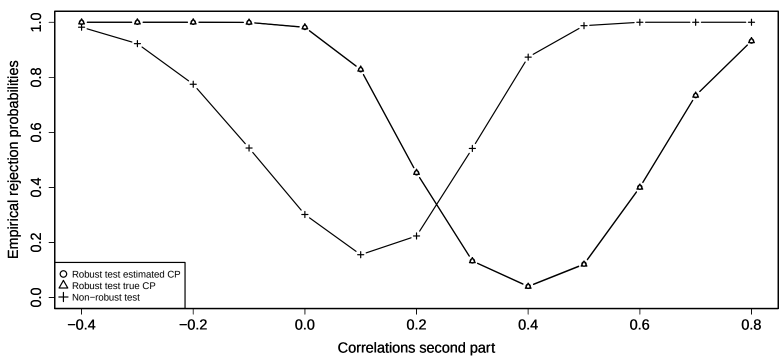

Figure 6: Empirical rejection probabilities of the non-robust and the robust test in a setting with changing cross-correlations and changing marginal expectations

6 Correlation of stock returns

In this section, we provide an empirical illustration of our methods, whereas we focus on the cross-correlation constancy case and revisit the analysis in Wied et al. (2012) using the robustified test. We thus reexamine the correlation of DAX and S&P 500 returns around the insolvency of Lehman Brothers in September 2008, which is often considered as the climax of the global financial crisis 2007-2008. Concretely, we use data from the beginning of 2005 until the end of 2009, which yields $T = 1244$ daily continuous returns, i.e., the first difference of the log-prices.

A picture of empirical correlations calculated in a rolling window of 50 days (Figure 7 a) gives some evidence for increasing correlations around the climax in the spirit of the "diversification meltdown"-hypothesis. This evidence is supported by the outcome of the test proposed in Wied et al. (2012), with a statistic given by

$$\max_{2\leq j\leq n} P(j) \quad \text{with} \quad P(j) = \left|\hat{D}\frac{j}{\sqrt{n}}\left(\hat{\rho}_j - \hat{\rho}_n\right)\right|,$$

where $\hat{\rho}_j$ are recursively estimated correlations and $\hat{D}$ is a kernel-based estimator for the asymptotic variance of $\hat{\rho}_n$ (for the exact implementation details see Wied et al., 2012).

Figure 7 b) shows the graph of the function $P(j)$ and it is clearly seen that the maximum is larger than the critical value (at the significance level 0.05) of 1.358. The (argmax) estimator for the break date is February 20th, 2008.

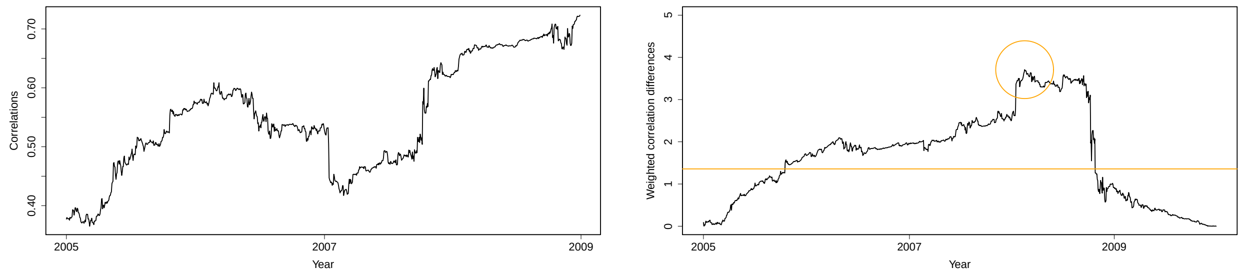

Figure 7: (a) Rolling correlations / (b) weighted differences of successively calculated correlations

A potential problem arises due to the fact that this test does not accommodate an (asymptotically non-vanishing) shift in the marginal variances. Instead, the power of the test close to 0 in the case of a sudden decrease and close to 1 in the case of a sudden increase; see Figure 1. Figures 8 a) and b) show the empirical variances calculated in rolling windows of 50 days of the two returns, respectively. There is evidence for a model with two variance regimes, where the variance in the second regime is higher than in the first one. This is confirmed by an application of the variance constancy test from Wied, Arnold, Bissantz, and Ziggel (2012) in combination with a binary segmentation algorithm applied in a similar way as in Galeano and Wied (2014). Applied on the two time series, the test yields a variance change point at the 14th of January 2008 for the DAX series and at the 3rd of September 2008 for the S&P500 series. After this, the data is split into the interval before the change point (including the point) and after in order to test in both segments again. To account for multiple testing, the smallest of the two p-values is compared with the significance level $\alpha = 1 - 0.95^{1/2}$. If smaller, a new change point is detected at the argmax of the corresponding series, the time series is split at this point again. The procedure is repeated with decreasing significance levels until no further change points can be found or until the distance between further change points is smaller than $0.05 \cdot T$. a refinement step is applied in order to improve the precision of the estimators. Here, the test is applied on each interval, which contains exactly one change point, and only statistically significant change points are kept. After this refinement step, no other change points except of the ones from the first step remain. We consider them as fixed in the following and no further variance change point estimations are performed, neither in the tests themselves nor in the bootstrap replications.

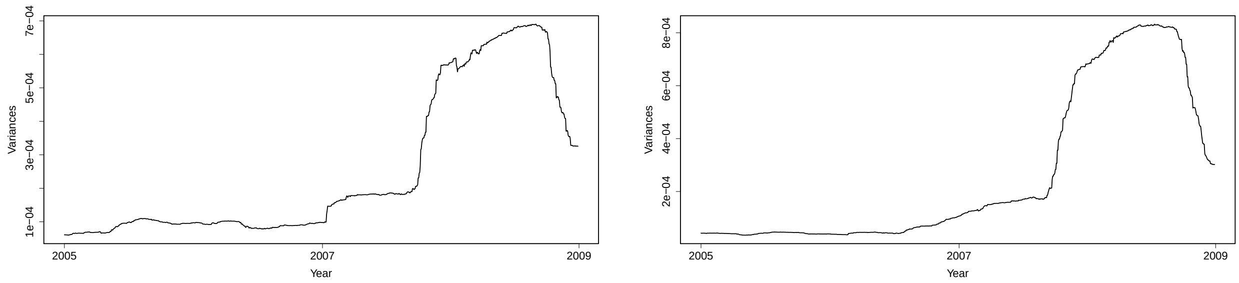

Figure 8: Rolling variances of (a) the DAX and (b) the S&P500 returns

We apply the test from (2) in combination with (5) which explicitly allows for a two-regime-model in the variances. The mean of daily returns is taken to be negligible, so we do not demean the series. Due to the complexity of the limit distribution, we rely on a bootstrap approximation following Subsection 5.3.1, with one modification: we resort to a block bootstrap, as the ACF of the product of the residuals $\hat{Z}_{t,1}\hat{Z}_{t,2}$ from (5) reveals autocorrelation (Figure 9) (once we eliminate variance breaks, stationarity of the series is plausible under the null of no changing correlations and we see no need to account for further possible nonstationarities). Consequently, we draw non-overlapping blocks of length $T^{1/3}$ and use $B = 9999$ bootstrap replications. Figure 10 shows a similar graph as Figure 7 b) for (2). The hypothesis of constant cross-correlation is rejected under these milder assumptions as well, but the date of the change point (estimated by the arg max statistic) is located half an year earlier, at the 9th of July 2007. Although small, the date can be tied to the 2007 liquidity crisis marking the beginning of the global financial crisis; the timing of the correlation break by the nonrobust test in February 2008 can be seen as a confusion with the variance break in January 2008 of the DAX returns series.

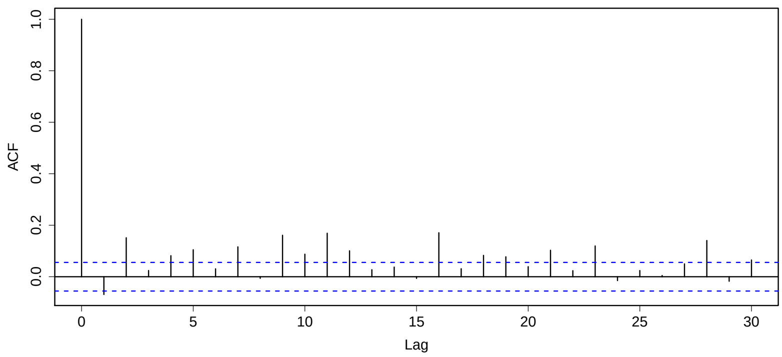

Figure 9: ACF of the residuals (4)

Moreover, Figure 10 raises doubt at the one-break-assumption. In particular, there is some evidence for at least one other change point after the 9th of July 2007. For clarification, we apply a binary segmentation algorithm in a similar way as in Galeano and Wied (2014) as described above. Before the iteration step, we get the additional dates 2nd of April 2009 in step 2 and 26th of September 2008 in step 3. In the iteration step, all three change points remain statistically significant, but the location of the point 2nd of April 2009 changes to the 2nd of December 2008. In the iteration step, the p-value of all tests is smaller than 0.001.

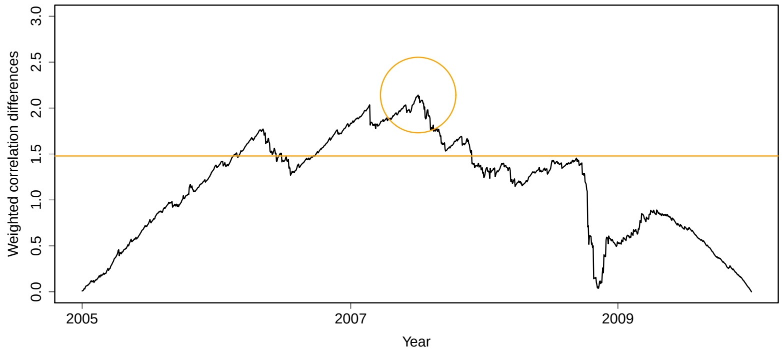

Figure 10: Weighted differences of successively calculated correlations (without the assumption of constant variances)

Regime	Correlation
Jan 4th 2005 - Jul 9th 2007	0.478
Jul 10th 2007 - Sep 25th 2008	0.505
Sep 26th 2008 - Dec 1st 2008	0.711
Dec 2nd 2008 - Dec 30th 2009	0.672

Table 1: Estimated regimes and corresponding empirical correlations

Table 6 gives an overview of the estimated regimes and the corresponding correlations. To sum up, we find that the correlation severely increases at the end of September 2008, corresponding quite closely to the Lehman bancruptcy, and drops somewhat in 2009 as the crysis appears to be getting under control.

Appendix

Before providing the main proofs, we state and prove an auxiliary result.

Lemma 1 *It holds under Assumptions 1 and 3 that*

$$\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_1\left(\lambda_0\right) = o_p\left(n^{-1/2}\right) = \hat{\boldsymbol{\theta}}_2\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_2\left(\lambda_0\right).$$

as $n \to \infty$*, provided that* $\boldsymbol{\theta}_{1,0} \neq \boldsymbol{\theta}_{2,0}$.

Proof of Lemma 1

Let us first discuss the behavior of

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_1\left(\lambda_0\right) &= \left(\sum_{j=1}^{\hat{\lambda}n} B_{j,n}' W_n \sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right)^{-1} \sum_{j=1}^{\hat{\lambda}n} B_{j,n}' W_n \sum_{j=1}^{\hat{\lambda}n} A_{j,n} + R_{\hat{\lambda}n,n} \\
&\quad - \left(\sum_{j=1}^{\lambda_0 n} B_{j,n}' W_n \sum_{j=1}^{\lambda_0 n} B_{j,n}\right)^{-1} \sum_{j=1}^{\lambda_0 n} B_{j,n}' W_n \sum_{j=1}^{\lambda_0 n} A_{j,n} - R_{\lambda_0 n,n} \\
&= P_n^{-1}\left(\hat{\lambda}\right) Q_n\left(\hat{\lambda}\right) - P_n^{-1}\left(\lambda_0\right) Q_n\left(\lambda_0\right) + R_{\hat{\lambda}n,n} - R_{\lambda_0 n,n},
\end{aligned}$$

where $P_n(\lambda) = \sum_{j=1}^{\lambda n} B_{j,n}' W_n \sum_{j=1}^{\lambda n} B_{j,n}$ and $Q_n(\lambda) = \sum_{j=1}^{\lambda n} B_{j,n}' W_n \sum_{j=1}^{\lambda n} A_{j,n}$, such that

$$P_n^{-1}(\lambda_0) = O_p\left(n^{-2}\right) \qquad \text{and} \qquad Q_n(\lambda_0) = O_p\left(n^{3/2}\right)$$

given the behavior of the individual components from Assumption 1 and 3. Since both λ_0 and $\hat{\lambda}$ (w.p. 1) are interior points of $[0,1]$, we also have from Assumption 3 that

$$\left|R_{\hat{\lambda}n,n} - R_{\lambda_0 n,n}\right| = o_p\left(n^{-1/2}\right)$$

for either $\hat{\lambda} \leq \lambda_0$ or $\hat{\lambda} > \lambda_0$. Furthermore,

$$\begin{aligned} &\left\|P_n^{-1}\left(\hat{\lambda}\right) Q_n\left(\hat{\lambda}\right) - P_n^{-1}(\lambda_0) Q_n(\lambda_0)\right\| \\ &\quad \leq \left\|P_n^{-1}\left(\hat{\lambda}\right) - P_n^{-1}(\lambda_0)\right\| \left\|Q_n\left(\hat{\lambda}\right)\right\| + \left\|P_n^{-1}(\lambda_0)\right\| \left\|Q_n\left(\hat{\lambda}\right) - Q_n(\lambda_0)\right\|. \end{aligned}$$

To assess $\left\|Q_n\left(\hat{\lambda}\right) - Q_n(\lambda_0)\right\|$, write

$$\left\|Q_n\left(\hat{\lambda}\right) - Q_n(\lambda_0)\right\| \leq \left\|\sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right\| \|W_n\| \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} A_{j,n}\right\| + \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| \|W_n\| \left\|\sum_{j=1}^{\lambda_0 n} A_{j,n}\right\|$$

where we make the convention that $\sum_{j=\hat{\lambda}n}^{\lambda_0 n} = -\sum_{j=\lambda_0 n}^{\hat{\lambda}n}$ if $\hat{\lambda} > \lambda$, such that

$$\left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} A_{j,n}\right\| \leq n\left|\lambda_0 - \hat{\lambda}\right| \sup_{1 \leq j \leq n} \|A_{j,n}\| = O_p\left(n^{1/(2+\alpha)}\right).$$

(The uniform $L_{2+\alpha}$ boundedness of $A_{j,n}$ has been used to derive the magnitude of the maximum.) We the have analogously that

$$\left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| = O_p\left(n^{1/(1+\alpha)}\right),$$

such that

$$\left\|\sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right\| = O_p\left(n^{1/(1+\alpha)}\right) \leq \left\|\sum_{j=1}^{\lambda_0 n} B_{j,n}\right\| = O_p\left(n^{1/(1+\alpha)}\right) + \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| = O_p\left(n^{1/(1+\alpha)}\right) = O_p(n)$$

and, summing up, that

$$\left\|Q_n\left(\hat{\lambda}\right) - Q_n(\lambda_0)\right\| = O_p\left(\max\left\{n^{1+1/(2+\alpha)}, n^{1/2+1/(1+\alpha)}\right\}\right) = o_p\left(n^{3/2}\right).$$

Furthermore, this implies that

$$\left\|Q_n\left(\hat{\lambda}\right)\right\| \leq \|Q_n(\lambda_0)\| + \left\|Q_n\left(\hat{\lambda}\right) - Q_n(\lambda_0)\right\| = O_p\left(n^{3/2}\right).$$

Now, Lütkepohl (1996, Section 8.4.1, Eq. (11c)) implies that

$$\left\|n^2 P_n^{-1}\left(\hat{\lambda}\right) - n^2 P_n^{-1}(\lambda_0)\right\| \leq \left\|n^2 P_n^{-1}(\lambda_0)\right\| \frac{\left\|n^2 P_n^{-1}(\lambda_0)\right\| \left\|\frac{1}{n^2} P_n\left(\hat{\lambda}\right) - \frac{1}{n^2} P_n(\lambda_0)\right\|}{1 - \left\|n^2 P_n^{-1}(\lambda_0)\right\| \left\|\frac{1}{n^2} P_n\left(\hat{\lambda}\right) - \frac{1}{n^2} P_n(\lambda_0)\right\|}$$

if $\left\|n^2 P_n^{-1}(\lambda_0)\right\| \left\|\frac{1}{n^2} P_n\left(\hat{\lambda}\right) - \frac{1}{n^2} P_n(\lambda_0)\right\| < 1$ and $\left\|n^2 P_n(\lambda_0)\left(\frac{1}{n^2} P_n\left(\hat{\lambda}\right) - \frac{1}{n^2} P_n(\lambda_0)\right)\right\| < 1$, where

$$\left\|P_n\left(\hat{\lambda}\right) - P_n(\lambda_0)\right\| \leq 2\left\|\sum_{j=1}^{\hat{\lambda}n} B_{j,n}\right\| \|W_n\| \left\|\sum_{j=\hat{\lambda}n}^{\lambda_0 n} B_{j,n}\right\| = O_p\left(n^{1+1/(1+\alpha)}\right) = o_p\left(n^2\right).$$

Consequently, $\left(\frac{1}{n^2} P_n\left(\hat{\lambda}\right) - \frac{1}{n^2} P_n(\lambda_0)\right) \xrightarrow{p} 0$ so the two conditions are fulfilled and we have that

$$\left\|P_n^{-1}\left(\hat{\lambda}\right) - P_n^{-1}(\lambda_0)\right\| = o_p\left(n^{-2}\right).$$

Summing up,

$$\left\|P_n^{-1}\left(\hat{\lambda}\right) Q_n\left(\hat{\lambda}\right) - P_n^{-1}(\lambda_0) Q_n(\lambda_0)\right\| = o_p\left(n^{-1/2}\right) = \hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_1(\lambda_0).$$

The result for $\hat{\boldsymbol{\theta}}_2\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_2(\lambda_0)$ is derived analogously and we omit the details.

Proof of Proposition 1

Use the mean value theorem to expand the vector function $\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right)$ elementwise about $\boldsymbol{\theta}_0$ to obtain with $\boldsymbol{Z}_t^* = \boldsymbol{h}(\boldsymbol{X}_t, \ldots; \boldsymbol{\theta}^*)$

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} g_l\left(\hat{\boldsymbol{Z}}_t\right) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} g_l(\boldsymbol{Z}_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} - \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\end{aligned}$$

where $\boldsymbol{\theta}^*$ is a convex combination of $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$. Since $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p\left(n^{-1/2}\right)$, $\boldsymbol{\theta}^*$ belongs to a $\sqrt{n}$-neighbourhood of $\boldsymbol{\theta}_0$ and thus to Φ_n; we pick $\boldsymbol{\theta}_t^* = \boldsymbol{\theta}^*$ $1 \leq t \leq n$, and Assumption 1 ensures uniform negligibility of the third term on the r.h.s. for $l = 1, \ldots, L$,

$$\begin{aligned}&\left\|\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} - \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right\| \\ &\qquad \leq \left\|\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right\| \sup_{\boldsymbol{\theta}^*, t} \left\|\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} - \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\| \\ &\qquad \xrightarrow{p} 0.\end{aligned}$$

The first result follows with Assumption 1 and the CMT.

Let us now examine the case of the recursive estimation scheme. Since $g_l\left(\tilde{\boldsymbol{Z}}_t\right)$ is a function of $\hat{\boldsymbol{\theta}}_t$, we have n convex combinations $\boldsymbol{\theta}_t^*$ of $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_t$ in the mean-value expansion about $\boldsymbol{\theta}_0$, leading to

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} g_l\left(\tilde{\boldsymbol{Z}}_t\right) &= \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} g_l(\boldsymbol{Z}_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0\right) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} \left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^*} - \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right) \left(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0\right).\end{aligned}$$

Since $\sup_{s\in[\epsilon,1]} \left\|\hat{\boldsymbol{\theta}}_{[sn]} - \boldsymbol{\theta}_0\right\| = O_p\left(n^{-1/2}\right)$ when $\boldsymbol{\Psi}$ is bounded in probability, the third term on the r.h.s. is immediately seen to vanish like before, such that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\boldsymbol{g}\left(\boldsymbol{Z}_t\right)+\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\left.\frac{\partial\boldsymbol{g}}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\left(\hat{\boldsymbol{\theta}}_t-\boldsymbol{\theta}_0\right)+o_p\left(1\right)$$

where the o_p term is uniform on $[\epsilon, 1]$, and the result is completed with Assumption 1 and the CMT.

Proof of Proposition 2

The desired asymptotic equivalence follows for the case of full-sample estimation from the condition

$$\sup_{s\in[0,1]}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\left(\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)\right)-\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\lambda_0\right)\right)\right)\right|=o_p\left(1\right).$$

Examining $\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)$, we have (writing explicitly only the dependence on $\boldsymbol{X}_t$ to simplify notation)

$$\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)\right)=\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)\left(1-D_{t,\hat{\lambda}}\right)+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\hat{\lambda}\right)\right)\right)D_{t,\hat{\lambda}}$$

and analogously

$$\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\lambda_0\right)\right)=\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\left(1-D_{t,\lambda_0}\right)+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\lambda_0\right)\right)\right)D_{t,\lambda_0}$$

such that

$$\begin{aligned}\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\hat{\lambda}\right)\right)-\boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\left(\lambda_0\right)\right) &= \boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)\left(1-D_{t,\hat{\lambda}}\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\left(1-D_{t,\lambda_0}\right)\\ &\quad+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\hat{\lambda}\right)\right)\right)D_{t,\hat{\lambda}}-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_2\left(\lambda_0\right)\right)\right)D_{t,\lambda_0}.\\ &= M_t+N_t\end{aligned}$$

Then,

$$\begin{aligned}M_t &= \left(\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\right)\left(1-D_{t,\hat{\lambda}}\right)+\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)\left(D_{t,\lambda_0}-D_{t,\hat{\lambda}}\right)\\ &= M_{1t}+M_{2t}.\end{aligned}$$

Now, $D_{t,\hat{\lambda}}$ is either zero or unity, so we may focus on $\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)-\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right)$ in discussing cumulated sums of M_{1t}, for which we resort to the mean value theorem elementwise and obtain like in the proof of Proposition 1 that, for each l, and $t\leq\lambda_0 n$,

$$\begin{aligned}g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right) &= g_l\left(\boldsymbol{Z}_t\right)+\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)-\boldsymbol{\theta}_{1,0}\right)\\ &\quad+\left(\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^*}-\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\right)\left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)-\boldsymbol{\theta}_{1,0}\right)\end{aligned}$$

and

$$\begin{aligned}g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t,\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)\right)\right) &= g_l\left(\boldsymbol{Z}_t\right)+\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\left(\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)-\boldsymbol{\theta}_{1,0}\right)\\ &\quad+\left(\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t^{0*}}-\left.\frac{\partial g_l}{\partial\boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t}\left.\frac{\partial\boldsymbol{h}}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}}\right)\left(\hat{\boldsymbol{\theta}}_1\left(\lambda_0\right)-\boldsymbol{\theta}_{1,0}\right)\end{aligned}$$

for suitable $\boldsymbol{\theta}_t^*$ ($\boldsymbol{\theta}_t^{0*}$) between $\boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)$ (between $\boldsymbol{\theta}_{1,0}$ and $\hat{\boldsymbol{\theta}}_1(\lambda_0)$), such that, for all $1 \le t \le \lambda_0 n$,

$$\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right) - \boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right) = \left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}} \left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_1(\lambda_0)\right) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

where the $o_p\left(\frac{1}{\sqrt{n}}\right)$ term is uniform in t following Assumption 3. For $t > \lambda_0 n$, we expand $g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right)$ and $g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right)$ about the same $\boldsymbol{\theta}_{1,0}$, but note that $\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{\theta}_{1,0}) \neq \boldsymbol{Z}_t$ for t in the second regime. We obtain however similarly

$$\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right)\right)\right) - \boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right) = \left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{\theta}_{1,0})} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}} \left(\hat{\boldsymbol{\theta}}_1\left(\hat{\lambda}\right) - \hat{\boldsymbol{\theta}}_1(\lambda_0)\right) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

thanks to Assumption 3. Using now Lemma 1, we obtain immediately

$$\sup_{s\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[sn]} M_{1t} \right| = o_p(1).$$

For M_{2t} we note that $\sum \left| D_{t,\hat{\lambda}} - D_{t,\lambda_0} \right| = O_p(1)$ since $\hat{\lambda} - \lambda_0 = O_p\left(n^{-1}\right)$. Then, for each $t < \lambda_0 n$ and l, write again

$$\begin{aligned} g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right) &= g_l(\boldsymbol{Z}_t) + \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}} \left(\hat{\boldsymbol{\theta}}_1(\lambda_0) - \boldsymbol{\theta}_{1,0}\right) \\ &\quad + \left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t^*} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} - \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}} \right) \left(\hat{\boldsymbol{\theta}}_1(\lambda_0) - \boldsymbol{\theta}_{1,0}\right) \end{aligned}$$

where $\sup_{t=1,\dots,n} |g_l(\boldsymbol{Z}_t)| = o_p(\sqrt{n})$ and $\sup_{t=1,\dots,n} \left\| \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}} \right\| = o_p(n)$ thanks to Assumption 3, and the third summand on the r.h.s. can be dealt with using Assumption 3 such that

$$\sup_{t=1,\dots,\lambda_0 n} \left\| \boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right) \right\| = o_p(\sqrt{n}).$$

For each $t \ge \lambda_0 n$ and l, we have like before

$$\begin{aligned} g_l\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right) &= g_l(\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{\theta}_{1,0})) + \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{\theta}_{1,0})} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}} \left(\hat{\boldsymbol{\theta}}_1(\lambda_0) - \boldsymbol{\theta}_{1,0}\right) \\ &\quad + \left(\left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{\theta}^*)} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} - \left.\frac{\partial g_l}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{\theta}_{1,0})} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_{1,0}} \right) \left(\hat{\boldsymbol{\theta}}_1(\lambda_0) - \boldsymbol{\theta}_{1,0}\right) \end{aligned}$$

and Assumption 3 leads analogously to

$$\max_{\lambda_0 n \le t \le n} \left\| \boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right) \right\| = o_p(\sqrt{n})$$

such that, summing up,

$$\sup_{s\in[0,1]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{[sn]} M_{2t} \right| \le \sup_{t=1,\dots,n} \left\| \boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \hat{\boldsymbol{\theta}}_1(\lambda_0)\right)\right) \right\| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| D_{t,\hat{\lambda}} - D_{t,\lambda_0} \right| = o_p(1).$$

The partial sums of N_t are evaluated in the same manner and the first result follows.

The case of recursive estimation follows along the same lines (but taking into account the fact that, at the beginning of the sample and after the break, the recursive estimator does not have proper asymptotics) and we omit the details.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica 62*(6), 1383–1414.

Aue, A., S. Hörmann, L. Horváth, M. Reimherr, et al. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics 37*(6B), 4046–4087.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory 13*(3), 315–352.

Bai, J. and S. Ng (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics 23*(1), 49–60.

Borowski, M., N. Rudak, B. Hussong, D. Wied, S. Kuhnt, and W. Tillmann (2014). On- and offline detection of structural breaks in thermal spraying processes. *Journal of Applied Statistics 41*(5), 1073–1090.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B 37*(2), 149–192.

Brüggemann, R., C. Jentsch, and C. Trenkler (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics 191*(1), 69–85.

Cavaliere, G. (2004). Unit root tests under time-varying variances. *Econometric Reviews 23*(3), 259–292.

Chan, H.-N., J. Chen, X. Chen, Y. Fan, and L. Peng (2009). Statistical inference for multivariate residual copula of GARCH models. *Statistica Sinica 19*, 53–70.

Chen, X. and Y. Fan (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics 135*, 125–154.

Davidson, J. (1994). *Stochastic Limit Theory.* Oxford University Press.

Dette, H. and D. Wied (2016). Detecting relevant changes in time series models. *Journal of the Royal Statistical Society Series B 78*, 371–396.

Dette, H., W. Wu, and Z. Zhou (2015). Change point analysis of second order characteristics in non-stationary time series. Mimeo.

Durbin, J. (1973). *Distribution theory for tests based on the sample distribution function*, Volume 9. Siam.

Galeano, P. and D. Wied (2014). Multiple break detection in the correlation structure of random variables. *Computational Statistics and Data Analysis 76*, 262–282.

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics 97*(1), 93–115.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–1054.

Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters 6*(3), 255–259.

Kianifard, F. and W. H. Swallow (1996). A review of the development and application of recursive residuals in linear models. *Journal of the American Statistical Association 91*(433), 391–400.

Lütkepohl, H. (1996). *Handbook of matrices.* John Wiley and Sons.

Mammen, E., C. Rothe, and M. Schienle (2012). Nonparametric regression with nonparametrically generated covariates. *Annals of Statistics 40*, 1132–1170.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society 55*(3), 703–708.

Phillips, P. C. B. (1991). A shortcut to LAD estimator asymptotics. *Econometric Theory 7*(4), 450–463.

Phillips, P. C. B., Y. Sun, and S. Jin (2006). Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation. *International Economic Review 47*(3), 837–894.

Ploberger, W. and W. Krämer (1992). The CUSUM test with OLS residuals. *Econometrica 60*(2), 271–285.

Wied, D., M. Arnold, N. Bissantz, and D. Ziggel (2012). A new fluctuation test for constant variances with applications to finance. *Metrika 75*(8), 1111–1127.

Wied, D., W. Krämer, and H. Dehling (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory 28*(3), 570–589.

Wu, C. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics 14*, 1261–1350.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*(502), 726–740.