Econometrics Journal (2018), volume **01**, pp. 1–22.
Article No. ectj??????

Testing for constant correlation of filtered series under structural change[1]

Matei Demetrescu[†] and Dominik Wied[‡]

[†] *Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Germany.*
E-mail: mdeme@stat-econ.uni-kiel.de

[‡] *Institute for Econometrics and Statistics, University of Cologne, Germany.*
E-mail: dwied@uni-koeln.de

Received:

Summary

The paper proposes a test for constant correlations allowing for breaks at unknown times in the marginal means and variances. Theoretically and in an application to US and German stock returns, we find that not accounting for changes in the marginal moments has severe consequences. This is because incorrect standardization of the series transfers to the sample correlations onto which the tests are built. Correcting for variance breaks at unknown time will have an asymptotic effect. To discuss adjustments, we tackle the issue more generally by considering partial-sums based inference on moment properties of unobserved processes which is conducted on the basis of estimated counterparts obtained in a preliminary step. The paper gives a characterization of the conditions under which the effect of filtering does not vanish asymptotically. The analysis extends to models with breaks in parameters at estimated time.

Key words: Bootstrap; Estimation Error; Partial Sums; Structural Break; Two-Step Procedure
JEL classification: C12 (Hypothesis Testing)

1. INTRODUCTION

Testing for time-varying moments and dependencies is of considerable interest in statistics and econometrics, in particular financial econometrics. This is motivated, among others, by the fact that correlations of asset returns increase in times of crises, just like their volatility.

(Co)Variance stability tests have e.g. 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 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 "almost" constant, yet unknown, variances. Such tools turned out to be useful, e.g., for forecasting risk measures like value at risk and expected shortfall, see Berens et al. (2015).

The drawback of such constant correlation tests is that (up to slight changes) the marginal variances are assumed to be constant under the null hypothesis of constant correlation. Yet changes in the marginal variances of the series of interest may easily create the impression of

[1]The authors would like to thank Jörg Breitung, as well as two anonymous referees and the editor (Dennis Kristensen) for suggestions significantly improving the paper.

a change in the correlation. Were the true time-varying variances known, one could simply use existing tests. However, if the marginal variances have to be estimated, the limit distribution derived under the assumption of constant means and variances may be affected.

More generally, estimated quantities are routinely 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 classical 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]

The paper develops procedures to test the null hypothesis of no changes in moments (say the pairwise correlation) of series that have possibly been filtered by means of GMM parameter estimators Hansen (1982). For instance, unless the variance of the series is known, one would have to at least standardize the series using estimated means and standard deviations, such that some filtering is virtually always conducted in practice. Importantly, we also allow the parameters of the filter to change at unknown times. The obvious example is a break in the marginal variances of the examined series; such breaks are easily mistaken for breaks in the correlation such that they have to be accounted for.

This extends the literature on structural breaks in the GMM framework. We go for instance beyond the setup of Andrews (1993), who considers the problem of testing constancy of a subset of model parameters by proposing sup-LM, sup-LM and sup-Wald tests, but requires nuisance parameters to be constant. Regarding CUSUM-type tests, Wied (2013) proposes for instance a test for constant parameters in a spatial autoregressive model for stock returns based on GMM-estimators (see also Lee et al., 2003 for CUSUM-type tests based on ML-estimators) and Zeileis (2005) provides a unified approach for structural change testing with score functions. Yet, like Andrews (1993), all these papers assume stationarity under the null hypothesis. Departing from stationarity under the null, Gagliardinia et al. (2005) propose robust GMM-based tests for parameter breaks, but focus on local deviations from stationarity having outliers in mind. Similarly, Wied et al. (2012)-test assume "almost" constant variances and show that the asymptotic distribution of their bivariate correlation test statistic remains the same if the variances change slightly (with the change vanishing asymptotically) and the ratio between the two variances remains constant. Such robustness to local departures from constancy of (nuisance) parameters is given more generally for GMM based inference, see Li and Müller (2009). But, in this paper, we consider unrestricted global changes, i.e. cases, where e.g. the variance permanently jumps from one value to another (arbitrary) one at unknown time. Such phenomena are not uncommon in financial data; see e.g. Rapach and Strauss (2008).

Concretely, we provide a generic discussion on the relation between the limiting distribution of test statistics based on partial sums of filtered series and of the test statistics based on the unobservable counterparts.[2] Using the filtered instead of the true series may have an effecton the statistics under scrutiny, but this need not 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)

[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).

[2] Note that our analysis 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 the effect of estimating regressors on subsequent estimation problems is analyzed. 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.

(although, as a byproduct of our analysis, we show the claim to be unsubstantiated), while Chen and Fan (2006), Chan et al. (2009) and Bücher et al. (2015) show asymptotic distributions of estimators in copula models to be unaffected by taking residuals from marginal models.

Our main results concern statistics based on partial sums of some transformation of the filtered series of interest. This covers the main case of interest, namely correlations, but allows the application of the main results to other situations of interest in applied work, say testing higher-order moments of latent variables. This extends the discussion of general specification tests provided by Newey (1985) and Tauchen (1985), who focus on sample sums rather than partial sums. The limiting behavior of normalized partial sums is essential for analyzing the parameter stability tests mentioned above. Clearly, the effect of using filtered series depends on both the filter which maps the unobservable terms of interest into observations and on the statistic of interest. The unknown parameters are estimated with a full-sample estimator or with a recursive estimator, and two types of filters are considered here, one which is continuous in unknown parameters and one which exhibits discontinuities, allowing us e.g. to deal with abrupt changes.[3] The analysis of the case with breaks at unknown time appears to be new in the literature.

The remainder of the paper is structured as follows. We give the formal setting in Section 2.

In Section 3, the paper firstly provides the asymptotic arguments for the smooth case together with a discussion of the conditions under which the use of the filtered instead of the true series does (does not) have an asymptotic effect, secondly addresses the case of structural changes and shows that plugging in an estimated break time is asymptotically equivalent to employing the true break time, and, thirdly, touches on the issue of asymptotic and bootstrap corrections. Here, it turns out that the filtering 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. That estimating the time of breaks does not affect the limiting behavior parallels the findings of Qu and Perron (2007) in Gaussian Quasi-ML estimation of regression models.

Section 4 introduces the new correlation constancy test, and gives Monte Carlo illustrations for the proposed test. We then provide an application to the correlation of US and German stock markets. 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; 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. Finally, we 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 proofs and additional material have been gathered in an online supplement. To be precise, Section A contains the proofs, Section B and C analytical derivations about the residual effects in particular models (B about covariance and correlation testing, Section C about the Jarque-Bera test) and Section D details about bootstrap approximations (which we refer to later on).

2. THE 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 $\frac{1}{\sqrt{n}} \sum_{t=1}^{[sn]} \boldsymbol{g}(\boldsymbol{Z}_t)$ are relevant. The leading case will be $g(\boldsymbol{z}) = z_1 z_2$ for pairwise covariances or, when $Z_{t,1}$ and $Z_{t,2}$ are standardized, correlations; some of the results are of more general

[3] Such discontinuities, and especially uncertainty about their timing, make an analysis based on tools such as the theory of contiguous distribution families developed in Le Cam (1960) less tractable.

applicability (e.g. the popular test for normality of Jarque and Bera, 1980 is recovered for $\boldsymbol{g}(z) = \left(z^3, z^4\right)$) so we consider the extra notation to be worth the effort.

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 $\boldsymbol{X}_t$, and $\boldsymbol{Z}_t$ and $\boldsymbol{X}_t$ have the same dimension. 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 take

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

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, resulting in filtered series $\hat{\boldsymbol{Z}}_t$ based on some estimators $\hat{\boldsymbol{\theta}}$ of the unknown parameters. One may equivalently regard $\hat{\boldsymbol{Z}}_t$ as model residuals and we use the terms residuals and filtered series interchangeably. We assume the estimators $\hat{\boldsymbol{\theta}}$ to belong to the family of generalized method-of-moments [GMM] estimators Hansen (1982), which includes e.g. M estimators as a particular case.

This formulation is fairly general. E.g., the dependence of $\boldsymbol{f}$ on the index t allows one to model trends, say $\boldsymbol{X}_t = t/n\, \boldsymbol{\theta} + \boldsymbol{Z}_t$. Additivity is not critical, but 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). 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.[4] 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. The leading case for this scenario is the model

$$\boldsymbol{X}_t = \boldsymbol{\mu}_{1,0}\left(1 - D_{t,\lambda}\right) + \boldsymbol{\mu}_{2,0} D_{t,\lambda} + \begin{pmatrix} \sqrt{\sigma_{1,1}^2\left(1 - D_{t,\lambda}\right) + \sigma_{1,2}^2 D_{t,\lambda}} & 0 \\ 0 & \sqrt{\sigma_{2,1}^2\left(1 - D_{t,\lambda}\right) + \sigma_{2,2}^2 D_{t,\lambda}} \end{pmatrix} \boldsymbol{Z}_t, \quad (2.1)$$

where $D_{t,\lambda} = \mathbb{I}\left(t/n \geq \lambda\right)$ and we are interested in the constancy of the correlation of the two components of $\boldsymbol{Z}_t$. Importantly, we will allow the break location λ to be unknown.

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 3.2 and focus for now on the case without breaks.

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 assumed to hold uniquely as well,

$$\boldsymbol{Z}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}_1\right) \mathbb{I}\left(t/n < \lambda\right) + \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}_2\right) \mathbb{I}\left(t/n \geq \lambda\right). \quad (2.2)$$

[4]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.

For time-series models, except for finite-order (nonlinear) autoregressions, the initial conditions matter, since the relevant past of $\boldsymbol{X}_t$ is not available in finite samples. One then often resorts 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 require e.g. $\sup_{s\in[0,1]} n^{-1/2} \left\| \sum_{t=1}^{[sn]} \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_1, {}^{t}\!/\!{}_{n}; \boldsymbol{\theta}\right) - \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, {}^{t}\!/\!{}_{n}; \boldsymbol{\theta}\right) \right\| \xrightarrow{p} 0$, ensuring asymptotic equivalence of the truncated and the unfeasible filters; we won't elaborate on the 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). Recursive estimation is involved e.g. in the case of inference on correlations Wied et al. (2012), but has a much longer history; see Kianifard and Swallow (1996) for an earlier review. The GMM-type estimator of $\boldsymbol{\theta}$ with $N \geq M$ moment restrictions are represented 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 2.1 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 $\boldsymbol{\theta}_0$. In the case of estimating the expectation with the arithmetic mean, one has $B_{j,n} = W_n = 1$, $A_{j,n}$ would be the observations and $R_{t,n} = 0$.

The residuals are given as $\hat{\boldsymbol{Z}}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_1, {}^{t}\!/\!{}_{n}; \hat{\boldsymbol{\theta}}\right)$ or $\tilde{\boldsymbol{Z}}_t = \boldsymbol{h}\left(\boldsymbol{X}_t, \ldots, \boldsymbol{X}_1, {}^{t}\!/\!{}_{n}; \hat{\boldsymbol{\theta}}_t\right)$, and inference on $\mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)$ is based on the partial sums of the transformed residuals,

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

We now give high-level assumptions that allow for a general discussion of the filtration effect.

Assumption 2.1. *With "$\Rightarrow$" denoting weak convergence in a space of cadlag functions on $[0,1]$ endowed with a suitable metric, it holds that:*

1 $\sqrt{n} \begin{pmatrix} \frac{1}{n} \sum_{t=1}^{[sn]} \left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right) - \mathrm{E}\left(\boldsymbol{g}\left(\boldsymbol{Z}_t\right)\right)\right) \\ \frac{1}{n} \sum_{t=1}^{[sn]} A_{t,n} \end{pmatrix} \Rightarrow \boldsymbol{\Psi}(s)$*, where* $\Psi(s)$ *is an* $L+N$*-dimensional Gaussian process with* $\boldsymbol{\Psi}(0) = 0$ *a.s. and* $\mathrm{Cov}\left(\boldsymbol{\Psi}(1)\right) = \Xi$*;*

2 $\frac{1}{n} \sum_{t=1}^{[sn]} B_{t,n} \Rightarrow \Pi(s)$ *where* $\Pi(s)$ *is a deterministic* $N \times M$ *matrix of Lipschitz functions, of rank* M *at all* $s \in (0,1]$, $\Pi(0) = 0$*; 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 $\frac{1}{n} \sum_{t=1}^{[sn]} \left.\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\right|_{\boldsymbol{z}=\boldsymbol{Z}_t} \left.\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \Rightarrow \boldsymbol{\tau}(s)$ *where* $\boldsymbol{\tau}(s)$ *is a deterministic matrix of differentiable functions (the gradient is a line vector and the Jacobian is built in a conformable manner);*

4 $\exists\ 0 < \epsilon < {}^{1}\!/\!{}_{2}$ *s.t., for a neighbourhood* $\Phi_n = \left\{\boldsymbol{\theta}^* : \|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\| < C n^{-1/2+\epsilon},\ C > 0\right\}$ *of* $\boldsymbol{\theta}_0$*,*

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

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

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

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 $t = O(1)$. Still, for any $0 < \epsilon < 1$, we have as a consequence of Assumption 2.1 the weak convergence

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

for any $0 < \epsilon < 1$. The convergence does not extend to $[0, 1]$ in general.[5] To deal with this situation one typically adds a step showing that $\hat{\boldsymbol{\theta}}_t$ for $t \in \{1, \ldots, [\epsilon n]\}$ do not have an asymptotic effect on the statistic of interest as $\epsilon \to 0$. 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 filtration effect. It is known (following e.g. Tauchen, 1985) that the residual effect vanishes in the limit of the full-sample sums if $\boldsymbol{\tau}$ defined in Assumption 2.1 is zero. But there are other interesting special cases for $\boldsymbol{\tau}$ where the residual effect vanishes; see Section 3.1 for the 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 avoid 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 $\boldsymbol{h}$, 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 strong 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 ($\boldsymbol{Z}_t$, $\boldsymbol{X}_t$, $A_{t,n}$ and $B_{t,n}$) be strong 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. Moreover, bootstrap implementations (see Section D in the supplement) may require additional smoothness conditions themselves. Note however that e.g. unit root or cointegrated DGPs are largely excluded since, in such nonstandard cases, $\hat{\boldsymbol{\theta}}_{[sn]} - \boldsymbol{\theta}_0$ would typically be non-Gaussian in the limit, and the convergence rate would not be $\sqrt{n}$; while accounting for this is not difficult in principle, the notational effort is not trivial and we do not further consider this topic here.

To construct test statistics based on the partial sums of $g(\hat{\boldsymbol{Z}}_t)$ (or $g(\boldsymbol{Z}_t)$), knowledge on Ξ is needed in general. Since this is typically not the case in practice, (consistent) estimation thereof is required. HAC estimators Newey and West (1987); Andrews (1991) are often used to estimate Ξ based on $\hat{\boldsymbol{Z}}_t$ and sample moment conditions $A_{t,n}$, although they are not the only choice (see e.g. Phillips et al., 2006). Note that HAC estimators are often consistent even under locally stationary; see e.g. Cavaliere (2004) for the case of time-varying variances. The focus of the paper being on the residual effect, we assume directly that a consistent estimator exists.

[5] Recursive trend adjustment is an exception; see Born and Demetrescu (2015).

ASSUMPTION 2.2. *There exists an estimator* $\hat{\Xi}$ *such that* $\hat{\Xi} \xrightarrow{p} \Xi$.

Assumption 2.1 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

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

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

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

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. MAIN RESULTS

While the residual effect is well understood for full-sample sums and smoothness conditions (see, among many others, Bai and Ng, 2005, Theorem 1, for a formulation for higher-order moments of $\boldsymbol{Z}_t$ in linear regressions), not much work has been done on the behavior of normalized partial sums based on filtered series with breaks at unknown time. To keep the paper self-contained we shall begin with a presentation of the smooth filter case and introduce breaks at unknown times afterwards. The relative advantages and disadvantages of various methods of accounting for the filtering effect are then briefly addressed in Section D in the supplement.

3.1. Residual-based partial sums

PROPOSITION 3.1. *Under Assumption 2.1, it holds as* $n \to \infty$ *that, on* $[0, 1]$,

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

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

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

REMARK 3.1. *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 known Brownian bridge.*

REMARK 3.2. *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 3.2, 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 3.2 for details.*

The obvious implication of the proposition is that the filtering effect appears for partial sums whenever $\boldsymbol{\tau}$ is not zero. Tests based on partial sums would not be affected if $\boldsymbol{\tau}(s) = \mathbf{0}$ for all $s \in [0,1]$,[6] but there are additional situations where specific tests are not affected even if $\boldsymbol{\tau} \neq \mathbf{0}$.

We first test simple hypotheses on the expectation of $\boldsymbol{g}(\boldsymbol{Z}_t)$. The null is of the form $\mathrm{E}(\boldsymbol{g}(\boldsymbol{Z}_t)) = \boldsymbol{\mu}^{(0)}$, and the Wald-type test statistic against alternatives of the form $\mathrm{E}(\boldsymbol{g}(\boldsymbol{Z}_t)) \neq \boldsymbol{\mu}^{(0)}$ 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 $\boldsymbol{g}(\boldsymbol{Z}_t)$. The scale matrix Ω is typically estimated, $\hat{\Omega}$; this would be the corresponding block of $\hat{\Xi}$, so a consistent estimator is available under Assumption 2.2.

The naive feasible versions of the test statistic are

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

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 $\boldsymbol{g}\left(\tilde{\boldsymbol{Z}}_t\right)$.

It follows from Proposition 3.1 and Assumption 2.2 that, under the null $\mathrm{E}(\boldsymbol{g}(\boldsymbol{Z}_t)) = \boldsymbol{\mu}_0$

$$\hat{\mathcal{T}} \xrightarrow{d} \hat{\boldsymbol{\Gamma}}'(1)\,\hat{\boldsymbol{\Gamma}}(1) \qquad \text{and} \qquad \tilde{\mathcal{T}} \xrightarrow{d} \tilde{\boldsymbol{\Gamma}}'(1)\,\tilde{\boldsymbol{\Gamma}}(1)$$

where

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

Without residuals, $\mathcal{T} \xrightarrow{d} \boldsymbol{\Gamma}(1)'\boldsymbol{\Gamma}(1)$ under the null, following as such a χ^2_L limiting null distribution (cf. Assumption 2.1), so the naive feasible versions are not pivotal in general, except for the obvious $\boldsymbol{\tau} = \mathbf{0}$ 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, which we include for completeness.

COROLLARY 3.1. *Under Assumptions 2.1 – 2.2, $\mathcal{T}$, $\hat{\mathcal{T}}$ and $\tilde{\mathcal{T}}$ are asymptotically equivalent under the null if $\boldsymbol{\tau}(s) = \mathbf{0}$ for all $s \in [0,1]$. Furthermore, the same holds for $\mathcal{T}$ and $\hat{\mathcal{T}}$ 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 discuss the test for constant correlation in Section 4 and provide additional examples in the supplement.

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

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

while the naive feasible versions are

$$\hat{Q}_n = \max_{1 \le j \le n} \frac{j}{\sqrt{n}} \sqrt{\left(\hat{\boldsymbol{S}}_j - \hat{\boldsymbol{S}}_n\right)' \hat{\Omega}^{-1} \left(\hat{\boldsymbol{S}}_j - \hat{\boldsymbol{S}}_n\right)} \qquad \text{with} \qquad \hat{\boldsymbol{S}}_j = \frac{1}{j}\sum_{t=1}^{j} \boldsymbol{g}\left(\hat{\boldsymbol{Z}}_t\right) \tag{3.3}$$

[6]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.

and

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

As a consequence of Proposition 3.1 and Assumption 2.2, we have

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

(assuming for the sake of the exposition that $\tilde{\boldsymbol{\Gamma}}(s)$ is defined for $s \in [0,1]$, keeping in mind that the second result in Proposition 3.1 only holds for $[\epsilon, 1]$).

Working with the unobserved $\boldsymbol{Z}_t$, the following well-known (pivotal) distribution

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

would have been obtained, so we ask, when is the distribution not affected by the filtering effect. Again, $\hat{Q}_n$ and $\tilde{Q}_n$ are asymptotically equivalent with Q_n when $\boldsymbol{\tau}(s) = \boldsymbol{0}$; but, in addition, there is another interesting case where equivalence of CUSUM statistics is given, at least for $\hat{Q}_n$:

COROLLARY 3.2. *Under Assumptions 2.1 – 2.2, the statistics Q_n, $\hat{Q}_n$ and $\tilde{Q}_n$ are asymptotically equivalent if $\boldsymbol{\tau}(s) = \boldsymbol{0}$ for all $s \in [0,1]$. Moreover, the statistics Q_n and $\hat{Q}_n$ are asymptotically equivalent if $\boldsymbol{\tau}(s) = s\boldsymbol{\tau}$ 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 B in the supplement for the concrete case of testing constancy of correlations under breaks in the marginal variances. Essentially, it requires first-order stationarity of $\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{z}}\Big|_{\boldsymbol{z}=\boldsymbol{Z}_t} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, but note that this actually is compatible with breaks when $\boldsymbol{\tau} = \boldsymbol{0}$.

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 3.2.

3.2. The residual effect under structural changes

In this subsection, we refer to the leading example of Section 2, equation (2.1), in which we are interested in testing for constant correlations under potentially nonconstant means and variances.

Let $D_{t,\lambda} = \mathbb{I}\left(t/n \ge \lambda\right)$ for some generic nontrivial break time $\lambda \in (0,1)$ and write the model with breaks as outlined in Section 2,

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

where $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$. Moreover, $\boldsymbol{\theta}_0 = \left(\boldsymbol{\theta}_{1,0}', \boldsymbol{\theta}_{2,0}'\right)'$ denotes the true parameter vector and λ_0 the true change point location.

We only model one break to avoid notational overhead, but note that this section easily extends to several breaks. 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

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

and, for $t \geq \lambda n$,

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

where the quantities $A_{j,n}, B_{j,n}$ and W_n are taken to obey Assumption 2.1 for the two subsamples, $1 \leq t < \lambda_0 n$ and $\lambda_0 n < t \leq n$. Since, in this formulation, the parameter vector is $\boldsymbol{\theta}$, one obtains a specific structure of the quantities which appear in the limit distribution displayed in Proposition 3.1. For example, $\boldsymbol{\Psi}_\lambda$, the analog of $\boldsymbol{\Psi}$ for the break case, 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

$$\Pi_\lambda(s) = \begin{pmatrix} \Pi(s)\mathbb{I}(s<\lambda) + \Pi(\lambda)\mathbb{I}(s \geq \lambda) & 0 \\ 0 & (\Pi(s) - \Pi(\lambda))\mathbb{I}(s \geq \lambda) \end{pmatrix}$$

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

$$W_{n,\lambda} = \begin{pmatrix} W_n & 0 \\ 0 & W_n \end{pmatrix} \to_p \begin{pmatrix} W & 0 \\ 0 & W \end{pmatrix} =: W_\lambda$$

Moreover,

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

where $\boldsymbol{\tau}_\lambda(s)$ is a $L \times (2M)$ matrix for all s, and obvious notation $\boldsymbol{\tau}^{\boldsymbol{\theta}_i}(s)$, $i = 1, 2$.

The true break date λ_0 is either known or not. 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 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 $\boldsymbol{h}(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \boldsymbol{\theta}) \neq \boldsymbol{Z}_t$ if $\boldsymbol{X}_t$ comes from the wrong regime, but we require minimal regularity conditions that would help control for this technical problem if the estimated break time is close enough to the true one.

ASSUMPTION 3.1. *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 *For some* $0 < \epsilon < \min\{\lambda_0, 1-\lambda_0\}$, $\sqrt{n}\sup_{s \in [\epsilon, \lambda_0] \cup [\lambda_0+\epsilon, 1]} \left|R_{[sn],n}\right| \xrightarrow{p} 0$, *and* $\sqrt{n}\sup_{s \in [\lambda_0, \lambda_0+\epsilon]} \left|R_{[sn],n} - R_{[\lambda_0 n],n}\right| \xrightarrow{p} 0$

3 *For* $\bar{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}_i, i = 1, 2\}$, $\max_{t=1,\ldots,n} \left\|\boldsymbol{g}\left(\boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \bar{\boldsymbol{\theta}}\right)\right)\right\| = o_p(\sqrt{n})$ *and* $\max_{t=1,\ldots,n} \left\| \frac{\partial g_l}{\partial \boldsymbol{z}}\Big|_{\boldsymbol{z}=\boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, t/n; \bar{\boldsymbol{\theta}}\right)} \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \right\| = o_p(n)$

4 *For* $\bar{\Phi}_n = \left\{\boldsymbol{\theta}^* : \left\|\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}\right\| < Cn^{-1/2+\epsilon}, 0 < \epsilon < 1/2, C > 0\right\}$, $\bar{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}_i, i = 1, 2\}$,

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

We also introduce extra notation: $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ depend on the assumed break time, so we make this explicit by writing $\hat{\boldsymbol{\theta}}_1(\lambda)$ etc. for $\lambda = \lambda_0$ or $\lambda = \hat{\lambda}$. They lead to residuals

$$\hat{\boldsymbol{Z}}_t(\lambda) := \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, {}^t\!/\!_n; \hat{\boldsymbol{\theta}}_1\right)(1 - D_{t,\lambda}) + \boldsymbol{h}\left(\boldsymbol{X}_t, \boldsymbol{X}_{t-1}, \ldots, {}^t\!/\!_n; \hat{\boldsymbol{\theta}}_2\right) D_{t,\lambda}$$

and similarly for $\tilde{\boldsymbol{Z}}_t(\lambda)$.

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 3.2. *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 2.1 and 3.1, it holds as $n \to \infty$, on $[0,1]$,*

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

A similar result holds for the case of recursive residuals.

Remark 3.3. *If the true break date is known, i.e., $\hat{\lambda} = \lambda_0$, it follows directly from Proposition 3.1 that Proposition 3.2 holds and we do not need Assumption 3.1. Thus, the effect of plugging in an estimated break time is asymptotically negligible. Note that the assumptions on $\hat{\lambda}$ are mild. On the one hand, there must be a minimal convergence rate, which is usually fulfilled, see Dette and Wied (2016). On the other hand, it must hold that $0 < \underline{\lambda} \leq \hat{\lambda} \leq \overline{\lambda} < 1$. This can be ensured for every break point estimator $\tilde{\lambda}$ by setting*

$$\hat{\lambda} = \tilde{\lambda}\mathbb{I}\{\underline{\lambda} \leq \tilde{\lambda} \leq \overline{\lambda}\} + \underline{\lambda}\mathbb{I}\{\tilde{\lambda} < \underline{\lambda}\} + \overline{\lambda}\mathbb{I}\{\tilde{\lambda} > \overline{\lambda}\}.$$

Remark 3.4. *Should there be no break and λ converges in distribution to a random variable, which is not concentrated on one point (see Dette and Wied, 2016 for an example), the weak limit in Proposition 3.2 changes. Since we explicitly model a break (and, in practice, one would test for the presence of breaks anyway), we do not pursue this topic here.*

4. TESTING FOR CONSTANT CORRELATION UNDER BREAKS IN THE MARGINAL DISTRIBUTION

We now turn our attention to the main question of testing the constancy of correlations under possible changes in the marginal distributions. After proposing the new test and arguing in favor of a bootstrap implementation, as its limiting distribution depends on several nuisance parameters, we illustrate the robustness properties of the new test in finite-sample experiments.

4.1. A robustified constant correlation test

Note that, although the marginal distributions may change in a number of ways, the ones relevant for testing the correlation in a nonparametric fashion are changes in marginal means and variances.

We therefore examine the effect *piecewise* standardization has on the relevant cross-product moment. We shall not consider recursive parameter estimation as the presence of breaks in the marginal means and variances complicates this approach without obvious advantages. The calculations can be found in the supplement (Section B) for the case of one break in mean or variance. They allow us to put forward a test for constant correlations that takes changes in the marginal means and variances into account by using the setup presented in this paper.

The robustified constant-correlation test is based on $\hat{Q}_n$ from (3.3) with

$$g\left(\hat{Z}_{t1}, \hat{Z}_{t2}\right) = \hat{Z}_{t1}\hat{Z}_{t2} \qquad \text{and} \qquad \hat{Z}_{ti} = \frac{X_{ti} - \hat{\mu}_{1,i}\left(1 - D_{t,\lambda}\right) - \hat{\mu}_{2,i} D_{t,\lambda}}{\sqrt{\hat{\sigma}^2_{i,1}\left(1 - D_{t,\lambda}\right) + \hat{\sigma}^2_{i,2} D_{t,\lambda}}}. \tag{4.4}$$

The limiting distribution (and the corresponding asymptotic correction) results in a straightforward manner, given the results of the previous section.

COROLLARY 4.1. *Under the Assumptions 2.1, 2.2 and 3.1, it holds as $n \to \infty$ that*

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

provided that $\hat{\lambda} = \lambda_0 + O_p(n^{-1})$, where

$$\hat{\mathbf{\Gamma}}\left(s\right) = \mathbf{\Gamma}\left(s\right) + \Omega^{-1/2}\boldsymbol{\tau}_{\lambda_0}\left(s\right)\left(\Pi'_{\lambda_0}(1)\, W\, \Pi_{\lambda_0}(1)\right)^{-1}\Pi'_{\lambda_0}(1)\, W\, \Sigma^{1/2}_{\lambda_0}\mathbf{\Theta}_{\lambda_0}\left(1\right).$$

Its form is however not particularly suitable for applied work, as it depends on nuisance parameters (in particular the true change point λ_0). While these may be estimated, tabulating the resulting critical values is a complication that may be avoided by using resampling methods. The following section describes the basic structure of the bootstrap algorithm employed in this work.

Note however that, for each specific application of the generic theory derived in this paper, a suitable bootstrap scheme needs to be established whenever the limiting null distribution is not immediately analytically accessible. However, no generic schemes seem to exists. For instance, preliminary simulations revealed that it is not possible to simply use a wild bootstrap (which may have accounted for changes in the marginal variances) in our testing setup. A discussion of these aspects is given in Section D in the supplement.

REMARK 4.1. *The extension to multiple breaks is straightforward: standardize the series in each regime separately to obtain the corresponding $\hat{\boldsymbol{Z}}_t$, and take the full number of regimes into account when constructing the resampled series for bootstrapping. Note that the breaks in the means and the variances may have different timing, so by regime we understand here the time segments where both mean and variance of either component of $\boldsymbol{X}_t$ are constant.*

REMARK 4.2. *A variant of the robust test consists of first applying a test for constant variances and, depending on the test's decision, either the robust test with estimated change-point or the non-robust test. Such two-stage pretest-based procedures are popular in econometric practice.*

4.2. Experimental evidence on the robustified constant correlation test

4.2.1. Setup In this subsection, we analyze the finite-sample behavior of the test for constant correlation if the marginal moments are time-varying. We compare the performance of the robust test and its two-stage (pre-test) version with that of the non-robust Wied et al. (2012) test and that a (non-robust test) constructed in the framework of Andrews (1993); concretely, we use the sup-Wald test as a representative his partial-sample GMM (PS-GMM) based tests. The sup-Wald test of Andrews (1993) offers in principle an alternative way of testing for constant correlations. The relevant moment conditions are

$$\begin{aligned} \mathrm{E}\left(X_k\right) = \mu_k, \quad \mathrm{E}\left(X_k^2\right) = \mu_k^2 + \sigma_k^2, \qquad \text{for } k = 1, 2; \\ \mathrm{E}\left(X_1 X_2\right) = \sigma_1\sigma_2\rho + \mu_1\mu_2 \end{aligned}$$

Andrews' nuisance parameter vector $\boldsymbol{\delta}$ stacks here μ_k and σ_k^2, and the sup-Wald statistic tests the constancy of ρ (β in Andrews' notation). However, his approach is not directly applicable here, because (as Andrews directly writes after equation (2.1) in his paper) the nuisance parameters are assumed constant under both the null and the alternative. This is exactly the assumption that we relax. Stacking means and variances in $\boldsymbol{\beta}$ alongside ρ leads to a test that will also consistently reject if the means or variances change but correlations don't. However, we are interested in tests which keep the size when the correlations are constant but means or variances may change. Therefore, using the PS-GMM estimators by (falsely) maintaining the assumption of constant nuisance parameters yields a non-robust one-step procedure; the sup-Wald statistic is easily checked to be

$$\sup_{s\in\Pi}\left(\frac{1}{[sn]}\sum_{i=1}^{[sn]}\hat{z}_{i1}\hat{z}_{i2}-\frac{1}{n-[sn]}\sum_{i=[sn]+1}^{n}\hat{z}_{i1}\hat{z}_{i2}\right)^2\left(\frac{1}{[sn]}\sum_{i=1}^{[sn]}\hat{z}_{i1}^2\hat{z}_{i2}^2+\frac{1}{n-[sn]}\sum_{i=[sn]+1}^{n}\hat{z}_{i1}^2\hat{z}_{i2}^2\right)^{-1},$$

in which the residuals z_{i1} and z_{i2} are obtained from the original observations by demeaning and dividing by its estimated standard deviations. The estimators for means and variances are obtained from the full sample. The interval Π is a subset of $[0,1]$, we consider the case $\Pi = [0.1, 0.9]$, and we use the critical values tabulated in Andrews (1993).

The test of Wied et al. (2012) is given by

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

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).

The new robust test is based on (3.3) with (4.4). The two-stage version is obtained by first employing the variance constancy test in Wied et al. (2012) (see equation (2) in that paper) on the individual series.

Throughout, the sample size is 500 and we use 10000 Monte Carlo replications and test at 5% nominal level.

4.2.2. Baseline bootstrap implementation To obtain critical values for the robust test, we resort to a bootstrap procedure. The change point is either known, $\lambda = \lambda_0$, or can be estimated super-consistently, $\lambda := \hat{\lambda}$. (Since only the convergence rate matters, we refrain from recommending a particular choice.) 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 piecewise demeaned X_{t1} and X_{t2}. (For the application, we shall resort to a block bootstrap in this step to allow for serial correlation.) We may then use as standardizing matrix $\hat{\Omega}$ the sample variance of $\hat{Z}_{t1}\hat{Z}_{t2}$. This is based on the implicit assumption that the only breaks relevant are in the means and the variances, which is reasonable for a wide variety of applications. 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 rescaled such that they have the same empirical variance as the original series. Shifting to match the original means is not required since the effect of demeaning vanishes in the limit and the bootstrap series need not replicate that. Then apply the test on the B bootstrapped series, and use the relevant empirical quantile of the bootstrap realizations of the statistic of interest as critical value. We use 199 bootstrap repetitions to keep the computational effort to a minimum.

We employ the same bootstrap critical values for the second stage of the two-stage robust variant.

4.2.3. Robustness with respect to non-constant variances First, we present evidence for the case of constant means and possibly changing marginal variances. For analyzing the size properties, we generate independent data from a bivariate normal distribution with mean zero and 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. We consider both the case of estimated and of true variance change point locations. In the first case, we estimate the breakpoint by applying the argmax estimator based on the variance constancy test in Wied et al. (2012). The new robust test is based on (3.3) with (4.4), but without demeaning in the numerator as we generate the series with zero mean and we focus on variances for this batch of simulations.

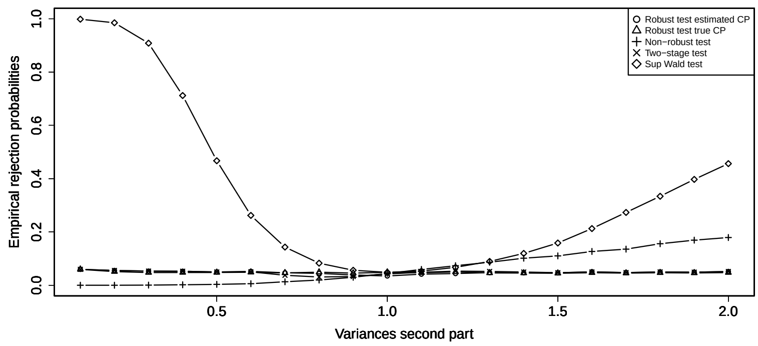

Figure 1. Empirical rejection probabilities of the non-robust and the robust tests in a setting with constant cross-correlations and non-constant marginal variances

The plot of the empirical sizes is given in Figure 1. One sees that our robust 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 figure also shows that the two-stage procedure keeps the size. It is marginally more conservative in the cases with slight variance decreases, though.

The size of the nonrobust Wied et al. (2012)-test is smaller than the nominal level in the case of decreasing variances and larger 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.

The sup-Wald test also reacts to increasing variances by over-rejecting, but more strongly so than the Wied et al. (2012)-test. In contrast to the latter, the sup-Wald test is extremely oversized for decreasing variances as well. This shows that the sup-Wald constant correlation statistic picks any change in the marginal moments as a change in correlations.

Figure 2.a shows the empirical power of the tests considered in the previous setting in a setting

under which the Wied et al. (2012) and the sup-Wald tests 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 the robust tests and that of the Wied et al. (2012)-test 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. The two-step procedure is as powerful as the Wied et al. (2012)-test. At the same time, we note that the sup-Wald test exhibits disadvantages in terms of power compared to all other tests, so one may see little reason to choose it over CUSUM-type constant-correlation tests in practice.

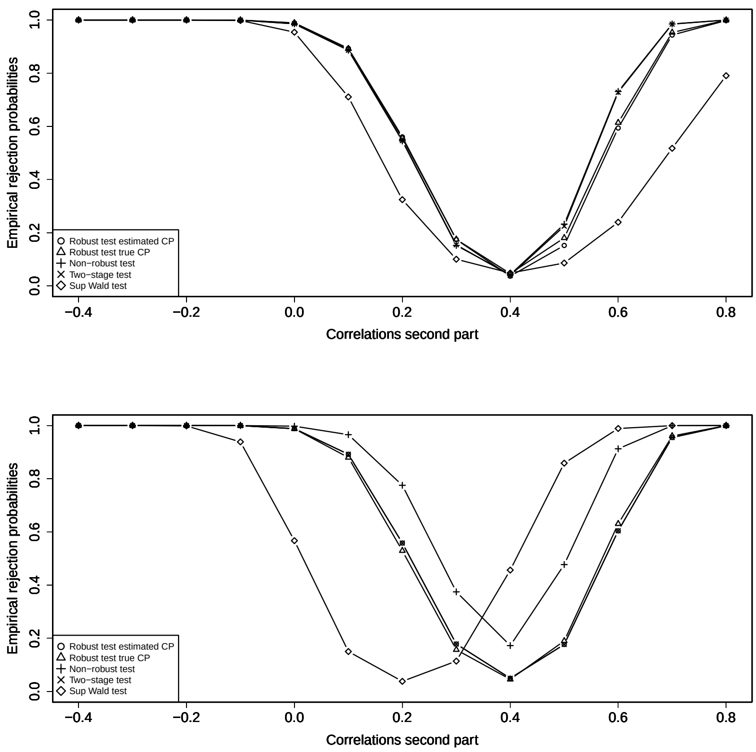

Figure 2. Empirical rejection probabilities of the non-robust and the robust tests in a setting with changing cross-correlations and constant (a) / changing (b) marginal variances (Note: the correlation in the first sample part is 0.4, which then marks the null on the x-axes.

Figure 2.b shows the empirical power of the tests in a setting under which the non-robust 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 of Wied et al. (2012) has higher rejection frequencies than the new test but, of course, it must be taken into account that it is quite oversized. As the test for constant variances always rejects in this setting, the power of the two-stage

procedure is the same as that of the robust test. In the case of nonconstant variances, the power curve of the sup-LM is shifted; under breaks in variances, a correlation of 0.2 in the second sample part most likely mimics the null of no change in correlation.

4.2.4. 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 by filtering out change points in the first moment. Since this does not induce a residual effect (see Section B in the supplement), 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 a significance level of 0.05, the critical value is 1.358.

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

We do not include the two-stage procedure in the comparison since it follows closely the behavior of the robust tests; we also exclude the sup-Wald test as it does not exhibit reliable behavior in the changing-variances case and we see no gain if further considering it.

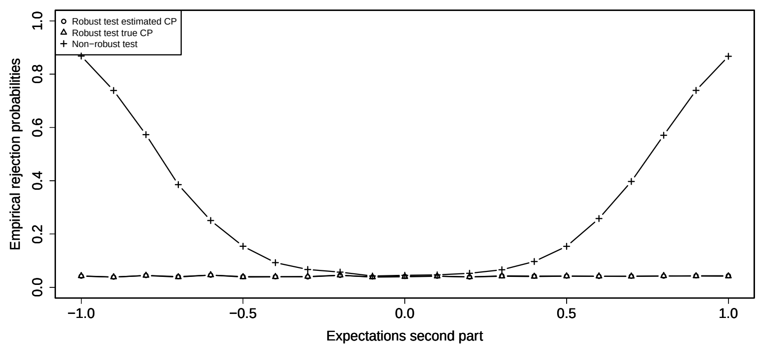

Figure 3. Empirical rejection probabilities of the non-robust and the robust test in a setting with constant cross-correlations and non-constant marginal expectations

The results are plotted in Figure 3. More so than in the breaking variance case (cf. Figure 1), estimating the change point makes no difference in the robust test's behavior: it is slightly conservative in both cases. The Wied et al. (2012)-test is oversized if the expectations change.

Figure 4.a compares (in a way similar to Figure 2.a) the robust and nonrobust tests in a setting with constant expectation zero. As in Figure 3, estimating change point locations does not make any difference compared to using the true change point locations. The Wied et al. (2012)-test performs however somewhat better for upward changes in the correlation (cf. Figure 2).

Finally, Figure 4.b (in a similar way as Figure 2.b) 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, the power curve of the Wied et al. (2012)-test has a minimum at 0.1. This is reminiscent of the behavior of the sup-Wald test under nonconstant variances. One must of course consider that the Wied et al. (2012)-test rejects almost every time

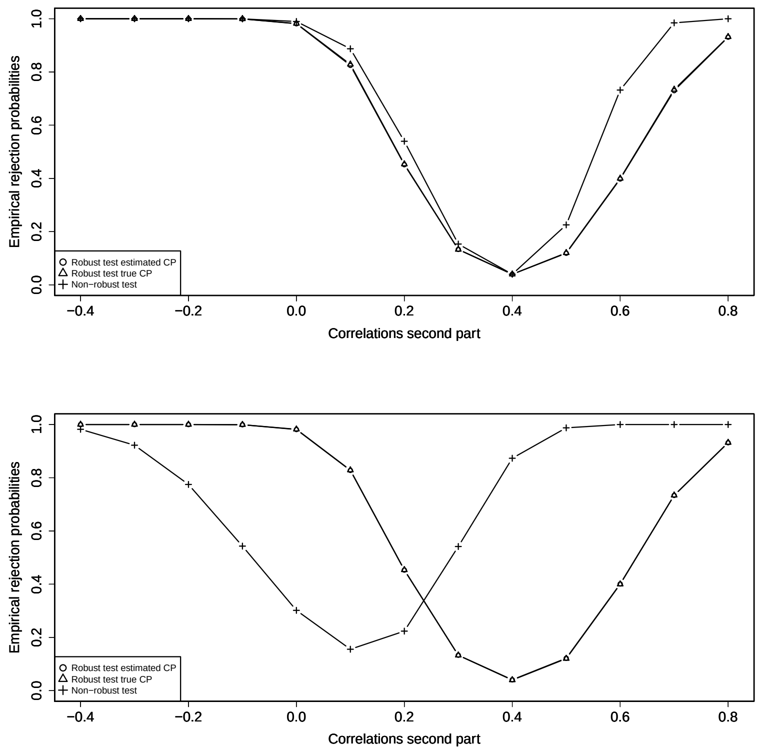

Figure 4. Empirical rejection probabilities of the non-robust and the robust test in a setting with changing cross-correlations and constant (a) / changing (b) marginal expectations

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

Summing up, out robust tests exhibit satisfactory size and power properties.

5. 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 5.a gives some evidence for increasing correlations around the climax in the spirit of the "diversification meltdown"-hypothesis. This is supported by the test of Wied et al. (2012), plotted in Figure 5.b, and it is clearly seen that the maximum is larger than the 5% critical value of 1.358. The (argmax) estimator for the break date is February 20th, 2008.

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.

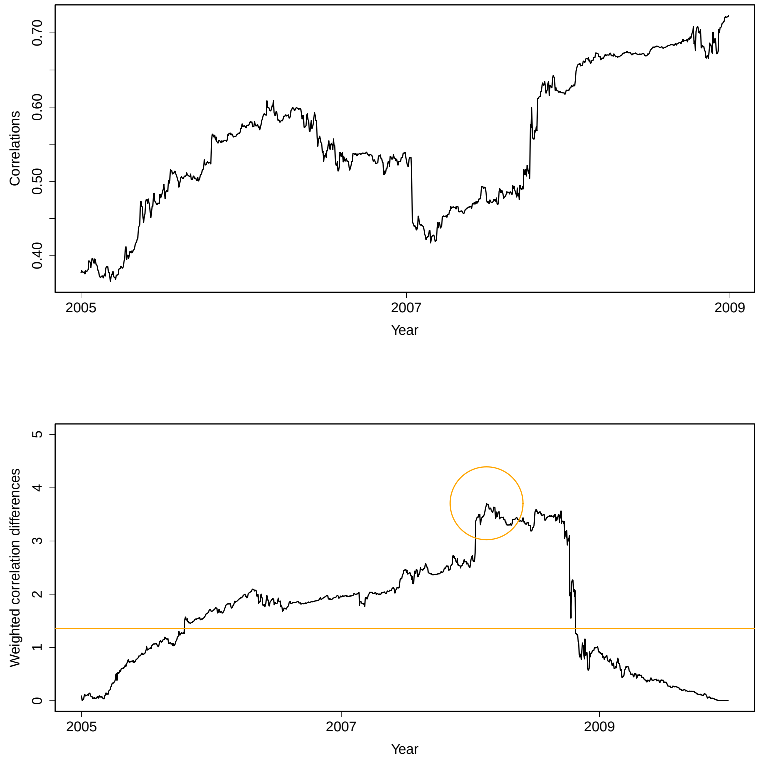

Figure 5. (a) Rolling correlations / (b) weighted differences of successively calculated correlations

Figures 6.a and 6.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 $1 - 0.95^{1/2}$. 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 $0.05 \cdot T$. Finally, 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.[7]

[7] We neglect potential estimation errors regarding the number and location in the subsequent calculations.

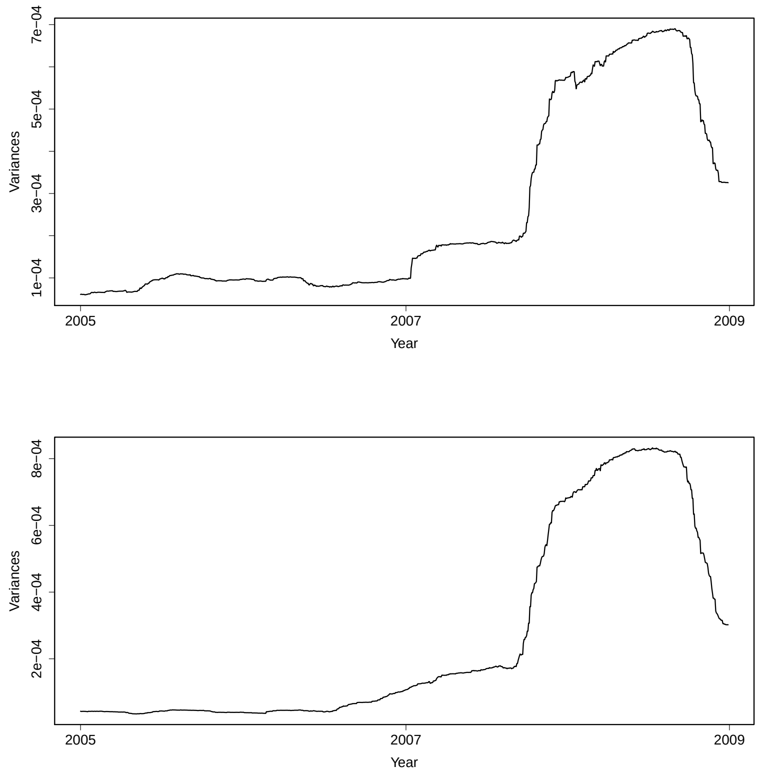

Figure 6. Rolling variances of (a) the DAX and (b) the S&P500 returns

We apply the test from (3.3) in combination with (4.4) 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 4.2.3, with one modification: we resort to a block bootstrap (as in Wied, 2017), as the ACF of the product of the residuals $\hat{Z}_{t,1}\hat{Z}_{t,2}$ from (4.4) reveals autocorrelation (see Figure 7) (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 8 shows a similar graph as Figure 5.b for (3.3). 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.

Moreover, Figure 8 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

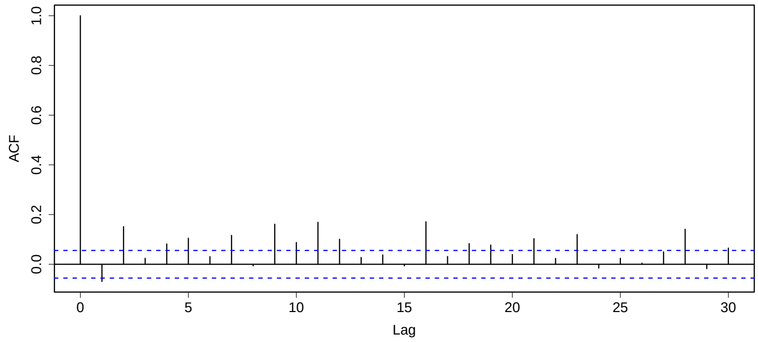

Figure 7. ACF of product $\hat{Z}_{t,1}\hat{Z}_{t,2}$ of piecewise standardized series

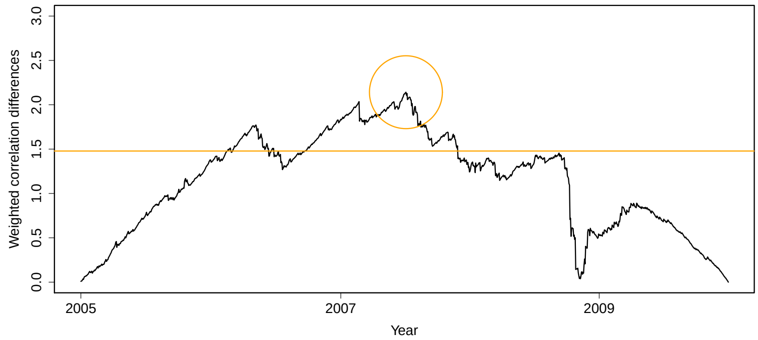

Figure 8. Weighted differences of successively calculated correlations (without the assumption of constant variances)

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.

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 1 gives an overview of the estimated regimes and corresponding correlations. We find that the correlation severely increases at the end of September 2008, corresponding quite closely to the Lehman bankruptcy, and drops somewhat in 2009 as the crisis appears to be under control.

6. CONCLUDING REMARKS

The paper tackled inference about moments of series that have been filtered using estimated filter parameters, with a direct application to testing pairwise correlations of series with unknown and possibly non-constant variances. We discussed conditions under which the filtering effect does not appear, and addressed the issue of breaks at unknown time in the parameters of the filter. For future research, it would be of interest to analyze the case where the function whose expectation is of interest is not smooth itself.

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.

Berens, T., G. Weiß, and D. Wied (2015). Testing for structural breaks in correlations: Does it improve value-at-risk forecasting? *Journal of Empirical Finance 32*, 135–152.

Born, B. and M. Demetrescu (2015). Recursive adjustment for general deterministic components and improved cointegration rank tests. *Journal of Time Series Econometrics 7*(2), 143–179.

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.

Bücher, A., S. Jäschke, and D. Wied (2015). Nonparametric tests for constant tail dependence with an application to energy and finance. *Journal of Econometrics 187*(1), 154–168.

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.

Gagliardinia, P., T. F., and G. Urga (2005). Robust gmm tests for structural breaks. *Journal of Econometrics 129*, 139–182.

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.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics 3*, 37–98.

Lee, S., J. Ha, O. Na, and S. Na (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics 96*, 782–796.

Li, H. and U. K. Müller (2009). Valid inference in partially unstable generalized method of moments models. *The Review of Economic Studies 76*(1), 343–365.

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

Newey, W. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica 53*(5), 1047–1070.

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., 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.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica 75*(2), 459–502.

Rapach, D. E. and J. K. Strauss (2008). Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics 23*(1), 65–90.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics 30*(1–2), 315–443.

Wied, D. (2013). Cusum-type testing for changing parameters in a spatial autoregressive model for stock returns. *Journal of Time Series Analysis 23*, 221–229.

Wied, D. (2017). A nonparametric test for a constant correlation matrix. *Econometric Reviews 36*, 1157–1172.

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.

Zeileis, A. (2005). A unified approach to structural change tests based on ml scores, f statistics, and ols residuals. *Econometric Reviews 24*, 445–466.

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