

# Adaptive Testing for Cointegration With Nonstationary Volatility

H. Peter Boswijk & Yang Zu

To cite this article: H. Peter Boswijk & Yang Zu (2021): Adaptive Testing for Cointegration With Nonstationary Volatility, Journal of Business & Economic Statistics, DOI: [10.1080/07350015.2020.1867558](https://doi.org/10.1080/07350015.2020.1867558)

To link to this article: <https://doi.org/10.1080/07350015.2020.1867558>



© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#) 



Published online: 03 Feb 2021.



[Submit your article to this journal](#) 



Article views: 141



[View related articles](#) 



[View Crossmark data](#) 

# Adaptive Testing for Cointegration With Nonstationary Volatility

H. Peter Boswijk<sup>a</sup> and Yang Zu<sup>b</sup>

<sup>a</sup>Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam, Amsterdam, The Netherlands; <sup>b</sup>School of Economics, University of Nottingham, Nottingham, UK

## ABSTRACT

This article develops a class of adaptive cointegration tests for multivariate time series with nonstationary volatility. Persistent changes in the innovation variance matrix of a vector autoregressive model lead to size distortions in conventional cointegration tests, which may be resolved using the wild bootstrap, as shown in recent work by Cavaliere, Rahbek, and Taylor. We show that it also leads to the possibility of constructing tests with higher power, by taking the time-varying volatilities and correlations into account in the formulation of the likelihood function and the resulting likelihood ratio test statistic. We find that under suitable conditions, adaptation with respect to the volatility process is possible, in the sense that nonparametric volatility matrix estimation does not lead to a loss of asymptotic local power relative to the case where the volatilities are observed. The asymptotic null distribution of the test is nonstandard and depends on the volatility process; we show that various bootstrap implementations may be used to conduct asymptotically valid inference. Monte Carlo simulations show that the resulting test has good size properties, and higher power than existing tests. Empirical analyses of the U.S. term structure of interest rates and purchasing power parity illustrate the applicability of the tests.

## ARTICLE HISTORY

Received June 2019  
Accepted December 2020

## KEYWORDS

Adaptive estimation;  
Nonparametric volatility  
estimation; Wild bootstrap

## 1. Introduction

An important approach to the analysis of cointegrated time series is based on a likelihood analysis of the Gaussian vector autoregressive (VAR) model, as developed by Johansen (1996). The resulting estimators and test statistics, although derived under the assumption that the disturbances are independent and identically normally distributed, can be shown to retain their asymptotic properties in more general circumstances. For example, the asymptotic critical values for the likelihood ratio (LR) test for the cointegration rank are still valid in the presence of leptokurtosis and time-varying volatilities, commonly observed in daily financial time series, as long as the invariance principle holds. Clearly, the resulting analysis is then based on a misspecified model and hence on a pseudo-likelihood, such that more efficient procedures may be based on the true likelihood function, which incorporates these characteristics. For the case of stationary (generalized) autoregressive-conditional heteroscedastic ((G)ARCH) processes, such procedures have been developed in the univariate case by Ling and Li (1998, 2003) and Seo (1999), and for the multivariate (cointegration) case by Li, Ling, and Wong (2001), Wong, Li, and Ling (2005), and Seo (2007).

Recent developments in the univariate unit root literature, however, have emphasized that volatility processes may display nonstationary variation, such that the disturbances no longer satisfy the conditions of an invariance principle, and hence standard unit root tests lose their asymptotic validity. Possible

causes of such nonstationarity include level shifts or other deterministic trending patterns in the volatility, see Kim, Leybourne, and Newbold (2002) and Cavaliere (2004), but also (near-)integrated GARCH dynamics, see Boswijk (2001). Cavaliere and Taylor (2007) and Beare (2018) developed two alternative approaches to constructing unit root test statistics with the conventional (Dickey–Fuller) asymptotic null distribution, and Cavaliere and Taylor (2008) showed that application of the wild bootstrap leads to asymptotically valid inference. Boswijk and Zu (2018) derived the power envelope for unit root tests with observable (nonstationary) volatility, and showed that considerable power gains may be obtained relative to procedures that do not take the heteroscedasticity into account. They also showed that when the volatility is unobserved, the power envelope may be reached by an adaptive procedure based on nonparametric volatility estimation.

This article seeks to extend Boswijk and Zu's (2018) analysis to a multivariate context, and hence develop powerful tests for cointegration in the presence of nonstationary multivariate (unconditional) heteroscedasticity. First, building on the analysis of Hansen (2003), we derive the LR test for cointegration in a VAR model with observed time-varying variance matrices and Gaussian errors. Next, we consider the case of unknown volatility, and propose a two-step procedure where the volatility process is estimated nonparametrically. Under suitable conditions, the resulting cointegration test has the same asymptotic power function as in the case of known volatility. The asymptotic

**CONTACT** H. Peter Boswijk  [h.p.boswijk@uva.nl](mailto:h.p.boswijk@uva.nl)  Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands.

 Supplementary materials for this article are available online. Please go to [www.tandfonline.com/UBES](http://www.tandfonline.com/UBES).

© 2021 The Authors. Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

null distribution of the test is nonstandard and depends on the volatility function, such that asymptotic  $p$ -values have to be obtained by Monte Carlo simulation or a bootstrap method; we develop the theory for the wild bootstrap, as well as for the volatility bootstrap (Boswijk and Zu 2018), where the bootstrap errors are based on the nonparametric volatility estimate.

In a related line of research, Cavaliere, Rahbek, and Taylor (2010, 2014) showed that application of the wild bootstrap to the traditional (constant-variance) Gaussian pseudo-likelihood ratio (PLR) statistic leads to a correctly sized cointegration test in the presence of nonstationary volatility. Cavaliere et al. (2018) showed that this result also holds if the lag order is unknown and a consistent lag selection criterion is used. However, their analysis does not exploit the power gain potential in the presence of nonstationary volatility.

The plan of the article is as follows. Section 2 presents the model and assumptions, and characterizes the limiting behavior of the process. In Section 3, we obtain an expression of the LR statistic for the cointegration rank for the case of a known volatility process, and we derive its limiting distribution, both under the null hypothesis and under a sequence of local alternatives. Section 4 discusses estimation of the volatility matrix, and its impact on the resulting test for cointegration rank. Section 5 gives Monte Carlo evidence about the finite sample performance of the test. Section 6 contains empirical applications of the test to the term structure of interest rates in the United States, and to purchasing power parity. Section 7 contains some concluding remarks, and proofs of all results are given in the online Appendix.

Throughout the article, we use the notation  $X_n \xrightarrow{p} X$  for convergence in probability,  $X_n \xrightarrow{w} X$  for convergence in distribution, and  $X_n \xrightarrow{w_p} X$  for weak convergence in probability, see Giné and Zinn (1990).  $X_n(u) \xrightarrow{w} X(u)$ ,  $u \in [0, 1]$  denotes weak convergence in  $\mathcal{D}[0, 1]^k$ , the product space of right-continuous functions with finite left limits (càdlàg), under the Skorohod metric. The notation  $\lfloor x \rfloor$  is used for the largest integer less than or equal to  $x$ . For any  $n \times m$  matrix  $A$  of full column rank  $m < n$ ,  $A_\perp$  denotes an  $n \times (n - m)$  matrix of full column rank such that  $A'_\perp A = 0$ , and  $\bar{A} = A(A'A)^{-1}$ . The Euclidean norm of a column vector  $x$  is denoted  $\|x\| = \sqrt{x'x}$ , and similarly the Frobenius norm of a matrix  $A$  is denoted  $\|A\| = \sqrt{\text{tr}(A'A)}$ .

## 2. The Model

Consider the VAR model of order  $k$ , written in error correction form, for a  $p$ -variate time series  $\{X_t, t = 1, \dots, n\}$ :

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t, \quad (1)$$

where  $\Pi$  and  $\Gamma_j, j = 1, \dots, k-1$  are  $p \times p$  coefficient matrices, and where  $\varepsilon_t$  is a  $p$ -variate disturbance vector with mean zero. The starting values  $\{X_{1-k}, \dots, X_0\}$  are considered fixed. For ease of exposition, we first consider the model with no deterministic components such as a constant or linear trend; extensions in this direction are discussed at the end of Section 3.

We wish to test the null hypothesis:

$$\mathcal{H}(r) : \Pi = \alpha\beta',$$

where  $\alpha$  and  $\beta$  are  $p \times r$  matrices,  $0 \leq r < p$ . Note that  $\mathcal{H}(r)$  may be equivalently formulated as  $\text{rank}(\Pi) \leq r$ . Under this hypothesis, the vector error correction model (VECM) becomes

$$\Delta X_t = \alpha\beta'X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t. \quad (2)$$

This implies that  $X_t$  is integrated of order 1, with cointegration rank  $r$  and cointegration matrix  $\beta$ , provided that the following assumption is satisfied (Johansen 1996, Theorem 4.2):

**Assumption 1.** In model (2), (a) the  $p \times r$  matrices  $\alpha$  and  $\beta$  are of full column rank  $r$ , (b) the characteristic equation  $\left| I_p(1 - z) - \Pi z - \sum_{j=1}^{k-1} \Gamma_j z^j(1 - z) \right| = 0$  has all its roots equal to one or outside the unit circle, and (c)  $\text{rank}(\alpha'_\perp \Gamma \beta_\perp) = p - r$ , where  $\Gamma = I_p - \sum_{j=1}^{k-1} \Gamma_j$ .

Under the assumption that the disturbances  $\{\varepsilon_t\}_{t \geq 1}$  are independent and identically distributed (iid) Gaussian with mean zero and positive definite variance matrix  $\Sigma$ , the likelihood function for the model under  $\mathcal{H}(r)$  is maximized by reduced rank regression. From this, an explicit expression is available for the LR test of  $\mathcal{H}(r)$  in the unrestricted model  $\mathcal{H}(p)$ , that is, against the alternative  $\mathcal{H}(p) \setminus \mathcal{H}(r) : r < \text{rank}(\Pi) \leq p$ ; see Johansen (1996). Here we consider a deviation from the iid assumption, allowing for unconditional heteroscedasticity:

**Assumption 2.** In model (2), the disturbances satisfy

$$\varepsilon_t = \sigma_t z_t, \quad t = 1, \dots, n,$$

where

- (a)  $\sigma_t = \sigma(t/n)$ , where  $\sigma(\cdot)$  is a nonstochastic, nonsingular  $p \times p$  matrix valued function on  $[0, 1]$ ; its elements  $\sigma_{ij}(\cdot)$  are, for all  $i, j = 1, \dots, p$ , piecewise Lipschitz continuous except on a finite number  $d$  of discontinuity points  $u_1, \dots, u_d \in [0, 1]$ .
- (b)  $\{z_t\}_{t \geq 1}$  satisfies  $E(z_t | \mathcal{F}_{t-1}) = 0$  and  $E(z_t z'_t | \mathcal{F}_{t-1}) = I_p$  for all  $t \geq 1$ , where  $\mathcal{F}_t = \sigma(\{z_s\}_{s=1}^t)$ , and  $\sup_{t \geq 1} E(\|z_t\|^8) < \infty$ .

The assumption directly implies  $E(\varepsilon_t) = 0$  and  $\text{var}(\varepsilon_t) = \sigma_t \sigma'_t =: \Sigma_t$ , a positive definite variance matrix. We will refer to  $\sigma_t$ , a matrix square root of  $\Sigma_t$ , as the *volatility matrix* of  $\varepsilon_t$ . In the next section, we will analyze the likelihood function derived from the stronger assumption  $\varepsilon_t \sim N(0, \Sigma_t)$ , but the asymptotic properties of the resulting procedures will continue to hold under Assumption 2 with non-Gaussian  $\{z_t\}_{t \geq 1}$ .

If the volatility process were such that the partial averages  $(un)^{-1} \sum_{t=1}^{\lfloor un \rfloor} \Sigma_t$  converge to the same positive definite matrix  $\Sigma$  as  $n \rightarrow \infty$  for all  $u \in [0, 1]$ , then, under suitable technical conditions, the invariance principle would apply to  $\{\varepsilon_t\}_{t \geq 1}$ . This in turn would imply that Johansen's (pseudo-)LR test, based on the Gaussian iid assumption on  $\{\varepsilon_t\}_{t \geq 1}$ , would retain its usual asymptotic properties, even though more efficient tests may be obtained from an analysis of the true likelihood function. Instead, the formulation in Assumption 2 is motivated by the notion that persistent changes in the volatility should be preserved in the limit. In the univariate context, this assumption was used by Cavaliere (2004), Cavaliere and Taylor (2007), and

Bearé (2018), inter alia. The multivariate analog was considered by Cavaliere, Rahbek, and Taylor (2010, 2014) and Boswijk et al. (2016). Assumption 2 is also related to Assumption 2 of Boswijk et al. (2016), who assumed  $\sigma_{ij}(\cdot)$  to be càdlàg, and allow for conditional heteroscedasticity in  $\{z_t\}_{t \geq 1}$ . Càdlàg functions can only have a finite number of discontinuities, as in Assumption 2. The assumption of Lipschitz continuity is made to facilitate consistent nonparametric estimation of  $\sigma(\cdot)$  on its continuity points, as considered in Xu and Phillips (2008). The analysis in the present article could be extended to allow for conditional heteroscedasticity, but this is not considered here to simplify the analysis.

Before we consider likelihood-based testing for  $\mathcal{H}(r)$  in the model (1) under Assumption 2, we conclude this section with a characterization of the limiting behavior of the process under the null  $\mathcal{H}(r)$ , and under a sequence of local alternatives

$$\mathcal{H}_n(r, r_1) : \Pi_n = \alpha\beta' + n^{-1}\alpha_1\beta_1', \quad (3)$$

where  $\alpha$  and  $\beta$  are the same as before, and  $\alpha_1$  and  $\beta_1$  are  $p \times r_1$  matrices of full column rank,  $r_1 \leq p - r$ , such that  $[\alpha : \alpha_1]$  and  $[\beta : \beta_1]$  are both of rank  $r + r_1$ . See Chapter 14 of Johansen (1996) and Hansen and Johansen (1998) for the analysis of the asymptotic local power of the LR test under the Gaussian iid assumption and (3). Proofs of all results are given in the online Appendix.

**Lemma 1.** In model (2) under Assumptions 1 and 2 and under  $\mathcal{H}_n(r, r_1)$ , we have

$$n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t \xrightarrow{w} \int_0^u \sigma(s) dW(s) =: M(u), \quad u \in [0, 1],$$

where  $W(\cdot)$  is a  $p$ -variate standard Brownian motion process, and

$$n^{-1/2} X_{\lfloor un \rfloor} \xrightarrow{w} \beta_{\perp} (\alpha'_{\perp} \Gamma \beta_{\perp})^{-1} U_A(u) =: X_A(u), \quad u \in [0, 1], \quad (4)$$

where the  $(p - r)$ -variate process  $U_A(\cdot)$  is given by

$$U_A(s) = \int_0^s \exp((u - s)A) \sigma_U(u) dW(u), \quad u \in [0, 1],$$

with  $A = \alpha'_{\perp} \alpha_1 \beta_1' \beta_{\perp} (\alpha'_{\perp} \Gamma \beta_{\perp})^{-1}$  and  $\sigma_U(u) = \alpha'_{\perp} \sigma(u)$ , such that  $U_A(\cdot)$  satisfies the stochastic differential equation

$$dU_A(u) = AU_A(u)du + \sigma_U(u)dW(u). \quad (5)$$

The limit  $X_A(\cdot)$  of  $n^{-1/2} X_{\lfloor \cdot n \rfloor}$  is a  $p$ -variate process, but of rank  $p - r$ , in the sense that  $\beta' X_A(u) = 0$  (a.s.). Note that  $U_A(\cdot)$  may be interpreted as a multivariate heteroscedastic Ornstein–Uhlenbeck process. The limit theory under  $\mathcal{H}(r)$  is obtained by setting  $r_1 = 0$  and hence  $A = 0$ , such that  $U_A(u)$  reduces to  $U_0(u) = \alpha'_{\perp} M(u)$ .

Lemma 1 implies that the PLR statistic, derived under the constant-variance assumption, will have a limiting distribution that depends on  $\sigma(\cdot)$ . In particular, in the simple case where  $k = 1$ , and we wish to test  $\mathcal{H}(0)$ , then it follows fairly directly from

Lemma 1 (see also Cavaliere, Rahbek, and Taylor 2010) that the PLR statistic satisfies, under the null hypothesis,

$$\text{PLR}_n(0) \xrightarrow{w} \text{tr} \left\{ \bar{\Sigma}^{-1} \int_0^1 dM(u)M(u)' \left( \int_0^1 M(u)M(u)' du \right)^{-1} \int_0^1 M(u)dM(u)' \right\},$$

where  $\bar{\Sigma} = \int_0^1 \Sigma(u)du$ , with  $\Sigma(\cdot) = \sigma(\cdot)\sigma(\cdot)'$ . If and only if  $\sigma(\cdot)$  is a constant matrix  $\sigma$ , such that  $M(u) = \sigma W(u)$  and  $\bar{\Sigma} = \sigma\sigma'$ , the usual limiting distribution tabulated in Johansen (1996) will result.

### 3. The Likelihood Ratio Test With Known Volatility

In this section, we analyze the LR statistic for  $\mathcal{H}(r)$  in model (2) in the case where  $\{\sigma_t\}_{t=1}^n$  is known, and where the standardized innovations  $\{z_t\}_{t=1}^n$  are taken to be iid  $N(0, I_p)$ . Although the assumption that  $\{\sigma_t\}_{t=1}^n$  is observed is unrealistic in practice, the asymptotic local power of such a test provides an optimal benchmark for the local power of tests in case  $\{\sigma_t\}_{t=1}^n$  is unknown and hence has to be estimated, either based on a parametric model or nonparametrically.

Define  $\Psi = [\Gamma_1 : \dots : \Gamma_{k-1}]$  and  $W_t = (\Delta X'_{t-1}, \dots, \Delta X'_{t-k+1})'$ , such that the model (2) under Assumption 2 with Gaussian  $\{\sigma_t\}_{t=1}^n$  may be expressed more compactly as

$$\begin{aligned} \Delta X_t &= \alpha\beta' X_{t-1} + \Psi W_t + \varepsilon_t, \\ \varepsilon_t | \mathcal{X}_{t-1} &\sim N(0, \Sigma_t), \\ t &= 1, \dots, n, \end{aligned} \quad (6)$$

where  $\Sigma_t = \sigma_t \sigma_t'$  as before, and  $\mathcal{X}_{t-1} = \{X_{t-1}, \dots, X_1, X_0, \dots, X_{1-k}\}$ . Recall that the starting values  $\mathcal{X}_0$ , and hence  $W_1$ , are observed but treated as fixed. The volatility matrices  $\{\sigma_t\}_{t=1}^n$  are also observed, but no specific model (such as multivariate GARCH) is assumed; they are treated as given. Under this condition, the log-likelihood function is given by

$$\begin{aligned} \ell_n(\alpha, \beta, \Psi) &= -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n \log |\Sigma_t| \\ &\quad - \frac{1}{2} \sum_{t=1}^n (\Delta X_t - \alpha\beta' X_{t-1} - \Psi W_t)' \Sigma_t^{-1} \\ &\quad (\Delta X_t - \alpha\beta' X_{t-1} - \Psi W_t). \end{aligned} \quad (7)$$

Maximum likelihood estimation in a closely related class of models was studied by Hansen (2003), who generalized the switching algorithm developed by Boswijk (1995) in various directions, including time-varying variance matrices. The key idea of this so-called *generalized reduced rank regression* procedure is that, although no closed-form expression exists for the maximum likelihood estimator (MLE)  $(\tilde{\alpha}_n, \tilde{\beta}_n, \tilde{\Psi}_n)$ , the maximization of  $\ell_n(\alpha, \beta, \Psi)$  over  $(\alpha, \Psi)$  for fixed  $\beta$  does lead to a closed-form expression, and similarly the MLE of  $\beta$  for fixed  $(\alpha, \Psi)$  has a closed-form expression. The likelihood may then be maximized, starting from an initial guess, by switching between maximization over  $(\alpha, \Psi)$  and  $\beta$ . Convergence properties of such switching algorithms have been studied by Sargan (1964) and Oberhofer and Kmenta (1974).

The algorithm requires that just-identifying restrictions are imposed on  $\beta$ . We formulate these as  $c'\beta = I_r$ , for some known  $p \times r$  matrix  $c$  of full column rank. An equivalent formulation is  $\beta = \bar{c} + c_\perp \Phi$ , where  $\Phi$  is a  $(p-r) \times r$  matrix of free parameters, such that

$$\text{vec}\beta = \text{vec}(\bar{c} + c_\perp \Phi) = h + H\phi, \quad (8)$$

where  $h = \text{vec}\bar{c}$  and  $H = I_r \otimes c_\perp$ , and  $\phi = \text{vec}\Phi$ . Other restrictions are also possible, as long as they are just-identifying, which implies  $r^2$  restrictions and hence  $r(p-r)$  free parameters in  $\beta$ .

Let  $Z_t(\beta) = (X'_{t-1}\beta, W'_t)'$ . Maximization of  $\ell_n(\alpha, \beta, \Psi)$  over  $(\alpha, \Psi)$  for fixed  $\beta$  leads to Hansen (2003, Theorem 2)

$$\begin{aligned} \text{vec}[\tilde{\alpha}_n(\beta) : \tilde{\Psi}_n(\beta)] &= \left( \sum_{t=1}^n [Z_t(\beta)Z_t(\beta)' \otimes \Sigma_t^{-1}] \right)^{-1} \\ &\text{vec} \left( \sum_{t=1}^n \Sigma_t^{-1} \Delta X_t Z_t(\beta)' \right), \end{aligned} \quad (9)$$

whereas the MLE of  $\beta$  for fixed  $(\alpha, \Psi)$  is given by

$$\begin{aligned} \text{vec}\tilde{\beta}_n(\alpha, \Psi) &= h + H \left( H' \sum_{t=1}^n [\alpha' \Sigma_t^{-1} \alpha \otimes X_{t-1} X'_{t-1}] H \right)^{-1} H' \\ &\times \sum_{t=1}^n \left\{ \text{vec} (X_{t-1} (\Delta X_t - \Psi W_t)' \Sigma_t^{-1} \alpha) \right. \\ &\quad \left. - [\alpha' \Sigma_t^{-1} \alpha \otimes X_{t-1} X'_{t-1}] h \right\}. \end{aligned} \quad (10)$$

Upon convergence of the switching algorithm, this yields the MLE  $(\tilde{\alpha}_n, \tilde{\beta}_n, \tilde{\Psi}_n)$ , and hence the residuals

$$\begin{aligned} \tilde{\varepsilon}_t &= \Delta X_t - [\tilde{\alpha}_n : \tilde{\Psi}_n] Z_t(\tilde{\beta}_n) = \Delta X_t - \tilde{\alpha}_n \tilde{\beta}'_n X_{t-1} - \tilde{\Psi}_n W_t, \\ t &= 1, \dots, n. \end{aligned}$$

In the special case  $r = 0$  (no cointegration), corresponding to  $\Pi = \alpha\beta' = 0$ , this reduces to  $\tilde{\varepsilon}_t = \Delta X_t - \tilde{\Psi}_n W_t$ , with  $\text{vec}\tilde{\Psi}_n = \left( \sum_{t=1}^n [W_t W'_t \otimes \Sigma_t^{-1}] \right)^{-1} \text{vec} \left( \sum_{t=1}^n \Sigma_t^{-1} \Delta X_t W'_t \right)$ .

The unrestricted model (1), corresponding to  $\mathcal{H}(p)$ , may be expressed as  $\Delta X_t = [\Pi : \Psi] Z_t + \varepsilon_t$ , where  $Z_t = Z_t(I_p) = (X'_{t-1}, W'_t)'$ . The corresponding log-likelihood is maximized by

$$\begin{aligned} \text{vec}[\hat{\Pi}_n : \hat{\Psi}_n] &= \left( \sum_{t=1}^n [Z_t Z'_t \otimes \Sigma_t^{-1}] \right)^{-1} \\ &\text{vec} \left( \sum_{t=1}^n \Sigma_t^{-1} \Delta X_t Z'_t \right), \end{aligned} \quad (11)$$

yielding the unrestricted residuals

$$\begin{aligned} \hat{\varepsilon}_t &= \Delta X_t - [\hat{\Pi}_n : \hat{\Psi}_n] Z_t = \Delta X_t - \hat{\Pi}_n X_{t-1} - \hat{\Psi}_n W_t, \\ t &= 1, \dots, n. \end{aligned} \quad (12)$$

Using these, the LR statistic for  $\mathcal{H}(r)$  against  $\mathcal{H}(p) \setminus \mathcal{H}(r)$ , with known volatility matrix, is

$$\begin{aligned} \text{LR}_n(r) &= -2 \left[ \ell_n(\tilde{\alpha}_n, \tilde{\beta}_n, \tilde{\Psi}_n) - \ell_n(\hat{\Pi}_n, I_{p+1}, \hat{\Psi}_n) \right] \\ &= \sum_{t=1}^n (\tilde{\varepsilon}'_t \Sigma_t^{-1} \tilde{\varepsilon}_t - \hat{\varepsilon}'_t \Sigma_t^{-1} \hat{\varepsilon}_t). \end{aligned} \quad (13)$$

The limiting behavior of  $\text{LR}_n(r)$  is characterized in [Theorem 1](#). Define

$$Y_A(u) = \begin{pmatrix} Y_{A,1}(u) \\ Y_{A,2}(u) \end{pmatrix} = \begin{pmatrix} \alpha'_\perp \sigma(u)^{-1} \\ \alpha' \sigma(u)^{-1} \end{pmatrix} \otimes U_A(u), \quad (14)$$

and

$$\begin{aligned} Z_A(u) &= Y_{A,1}(u) - \int_0^1 Y_{A,1}(s) Y_{A,2}(s)' ds \\ &\left[ \int_0^1 Y_{A,2}(s) Y_{A,2}(s)' ds \right]^{-1} Y_{A,2}(u). \end{aligned} \quad (15)$$

**Theorem 1.** In model (2), under [Assumptions 1–2](#) and under  $\mathcal{H}_n(r, r_1)$ , the LR statistic (13) satisfies, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \text{LR}_n(r) &\xrightarrow{w} \left( \int_0^1 Z_A(s) [dW(u) + Z_A(u)' \text{vec}(A') du] \right)' \\ &\left( \int_0^1 Z_A(u) Z_A(u)' du \right)^{-1} \\ &\left( \int_0^1 Z_A(u) [dW(u) + Z_A(u)' \text{vec}(A') du] \right). \end{aligned} \quad (16)$$

We observe that the limiting distribution under the null hypothesis  $\mathcal{H}(r)$ , such that  $A = 0$ , depends on (the process generating)  $\sigma(u)$ , and on  $\alpha$  (and hence  $\alpha_\perp$ ). Therefore, no uniformly applicable tables of critical values can be constructed. Quantiles and  $p$ -values of the limiting distribution can be obtained by Monte Carlo simulation of the limiting expression in (16), discretizing the integrals and replacing  $\alpha$  by  $\tilde{\alpha}_n$ . Consistency of  $\tilde{\alpha}_n$  (which follows from the proof of [Theorem 1](#)) guarantees the asymptotic validity of such  $p$ -values, as the sample size, the number of steps in the discretization and the number of Monte Carlo replications tend to infinity. Alternatively, bootstrap-based approaches to approximate the asymptotic null distribution are discussed in the next section.

In the special case of the null hypothesis  $\mathcal{H}(0)$  (no cointegration), the expression for the limiting distribution of the LR statistic simplifies somewhat. The representation in [Corollary 1](#) follows directly from (16), with  $Z_A(u) = \sigma(u)^{-1} \otimes U_A(u)$  and

$$\begin{aligned} dU_A(u) &= \sigma(u) [dW(u) + \sigma(u)^{-1} A U_A(u) du] \\ &= \sigma(u) [dW(u) + Z_A(u)' \text{vec}(A') du]. \end{aligned}$$

**Corollary 1.** Under the conditions of [Theorem 1](#), the LR statistic  $\text{LR}_n(0)$  for  $r = 0$  satisfies, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \text{LR}_n(0) &\xrightarrow{w} \int_0^1 dU_A(u)' [\Sigma(u)^{-1} \otimes U_A(u)'] \\ &\left( \int_0^1 [\Sigma(u)^{-1} \otimes U_A(u) U_A(u)'] du \right)^{-1} \\ &\int_0^1 [\Sigma(u)^{-1} \otimes U_A(u)] dU_A(u). \end{aligned}$$

We conclude this section with a discussion of the adjustments needed to accommodate a constant or linear trend term in the model. We focus on models where the process has either a constant mean or a linearly trending mean in both the stationary and the nonstationary directions. As is well known (Johansen

1996, chaps. 5 and 6), this is accomplished by considering the following two extensions of (2). To allow for a constant mean, the model becomes

$$\begin{aligned} \Delta X_t &= \alpha(\beta' X_{t-1} + \rho_0) + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t \\ &= \alpha \beta^\# X_{t-1}^\# + \Psi W_t + \varepsilon_t, \end{aligned} \tag{17}$$

where  $\rho_0$  is an  $r$ -vector, and where  $\beta^\# = (\beta', \rho_0)'$  and  $X_{t-1}^\# = (X'_{t-1}, 1)'$ . A linear trend is included via

$$\begin{aligned} \Delta X_t &= \mu + \alpha(\beta' X_{t-1} + \rho_1 t) + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t \\ &= \alpha \beta^\# X_{t-1}^\# + \Psi^\# W_t^\# + \varepsilon_t, \end{aligned} \tag{18}$$

where  $\mu$  is an  $n$ -vector and  $\rho_1$  is an  $r$ -vector, and where now  $\beta^\# = (\beta', \rho_1)'$ ,  $X_{t-1}^\# = (X'_{t-1}, t)'$ ,  $\Psi^\# = [\mu : \Psi]$  and  $W_t^\# = (1, W'_t)'$ . The log-likelihood function under Assumption 2 is analogous to (7), with parameters and regressors replaced by their “#” counterparts.

Adjusting the identification restrictions (8) accordingly, such that  $\text{vec} \beta^\# = \text{vec}(\tilde{c}^\# + c^\#_\perp \Phi^\#)$  with  $c^\#$  of dimensions  $(p+1) \times r$  and hence  $c^\#_\perp$  and  $\Phi^\#$  of dimensions  $(p+1) \times (p+1-r)$  and  $(p+1-r) \times r$ , respectively, the switching algorithm based on (9) and (10) remains the same, with all parameters and vectors replaced by their “#” counterparts. The limiting distribution of the resulting LR test statistic is stated in the following corollary (a sketch of its proof is provided in the online Appendix).

**Corollary 2.** In models (17) and (18), under Assumptions 1 and 2 and under  $\mathcal{H}_n(r, r_1)$ , the LR statistic (13) satisfies, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \text{LR}_n(r) &\xrightarrow{w} \left( \int_0^1 Z_A^\#(u) [dW(u) + Z_A^\#(u)' \text{vec}(A^\#) du] \right)' \\ &\quad \left( \int_0^1 Z_A^\#(u) Z_A^\#(u)' du \right)^{-1} \\ &\quad \left( \int_0^1 Z_A^\#(u) [dW(u) + Z_A^\#(u)' \text{vec}(A^\#) du] \right), \end{aligned}$$

where  $A^\# = [A : 0]$  and  $Z_A^\#(u)$  is defined analogously to (14) and (15), with  $U_A(u)$  replaced by  $U_A^\#(u) = (U_A(u)', 1)'$  in model (17), whereas in model (18),  $U_A(u)$  is replaced by  $U_A^\#(u) = (U_A(u)', u)'$  and  $Y_A(u)$  is replaced by

$$\begin{aligned} Y_A^\#(u) &= \begin{pmatrix} \alpha'_\perp \sigma(u)^{\prime-1} \\ \alpha' \sigma(u)^{\prime-1} \end{pmatrix} \otimes U_A^\#(u) \\ &\quad - \int_0^1 \left( \begin{pmatrix} \alpha'_\perp \Sigma(s)^{-1} \\ \alpha' \Sigma(s)^{-1} \end{pmatrix} \otimes U_A^\#(s) \right) ds \\ &\quad \left[ \int_0^1 \Sigma(s)^{-1} ds \right]^{-1} \sigma(u)^{\prime-1}. \end{aligned}$$

## 4. Adaptive Likelihood Ratio Test

### 4.1. Volatility Estimation

In the previous section, we have developed a LR test for cointegration when the volatility process  $\sigma(\cdot)$  is known. In specific

applications to financial data, the assumption that the volatility is observed with negligible measurement error may not be entirely unrealistic, since high-frequency intra-day data may be used to estimate the daily or weekly volatility with a high degree of precision (see, e.g., Andersen et al. 2003). In this section, however, we consider the case where the only data available is  $\{X_t, t = 1 - k, \dots, 0, 1, \dots, n\}$ , and hence an estimator of the volatility matrix has to be obtained from the data at the same observation frequency as used to construct the likelihood function and hence the cointegration test.

The volatility matrix  $\sigma_t$  may be estimated either parametrically or nonparametrically. In the presence of conditional heteroscedasticity, possible parametric approaches include multivariate GARCH models, notably the dynamic conditional correlation (DCC) model of Engle (2002). The LR test statistic may then be obtained by full maximization of the likelihood function for the Gaussian VAR-DCC model, with and without the reduced rank restriction. The obvious disadvantage of such a parametric approach is that it relies on the assumption of correct specification of the volatility process.

Alternatively, a two-step approach may be used, where the volatility matrix is estimated based on the residuals from least-squares estimation of the unrestricted VAR model, and the resulting estimator  $\hat{\Sigma}_t$  is then substituted for  $\Sigma_t$  in the expressions for the MLE and LR statistic given in the previous section. In this article, we propose to estimate  $\sigma_t$  by a nonparametric kernel estimator, generalizing the approach of Boswijk and Zu (2018), which in turn is based on Hansen (1995). However, as analyzed by Nelson (1996), multivariate GARCH models (with deterministic parameter sequences instead of estimated parameters) may also be interpreted as nonparametric filters of continuous-time multivariate stochastic volatility processes. Indeed, Engle (2002) showed via Monte Carlo simulations that the DCC model is rather successful in recovering time-varying correlation paths that are not generated by a DCC process. Therefore, in the continuous-time asymptotic framework of Assumption 2, the difference between parametric and nonparametric approaches is not as essential as it may appear at first sight.

We closely follow the approach to nonparametric volatility estimation and adaptive testing developed by Xu and Phillips (2008) and Patilea and Raïssi (2012), who extended Hansen’s (1995) approach to include leads and lags of the outer product of the residual vector, and to allow for discontinuities in the underlying volatility functions.

Let  $\{e_t\}_{t=1}^n$  denote the least-squares residual vectors of the model (6) (or of the extended models (17) or (18)) with  $r = p$ ; or equivalently the residual vector based on the unrestricted ML estimator (11) and (12) with  $\Sigma_t = I_n$  (where the model is extended by a constant and/or linear trend if applicable). Let  $K(\cdot)$  be a kernel function and define  $K_h(x) = K(x/h)/h$  with  $h > 0$  a window width. The kernel estimator for  $\Sigma_t$  is defined as

$$\hat{\Sigma}_t = \frac{\sum_{s=1}^n K_h\left(\frac{t-s}{n}\right) e_s e_s'}{\sum_{s=1}^n K_h\left(\frac{t-s}{n}\right)}. \tag{19}$$

By choosing different kernel functions, one could consider both one-sided smoothing (or filtering, where  $\hat{\Sigma}_t$  is based on lags of

$e_t e_t'$  only), or two-sided smoothing (based on leads and lags). Note, however, that the use of a one-sided filter for small  $t$  would require a boundary value adjustment to (19), as in Hansen (1995). We do not consider this explicitly, and correspondingly Assumption 3 imposes two-sided smoothing.

**Assumption 3.**  $K$  is a bounded, continuous and nonnegative function defined on the real line, satisfying  $\int_{-\infty}^{\infty} K(x)dx = 1$ ,  $\int_{-\infty}^0 K(x)dx > 0$  and  $\int_0^{\infty} K(x)dx > 0$ .

Recall that Assumption 2 allows for a finite number of discontinuities in the volatility matrix function. Although the estimator defined above is consistent over the intervals where this function is continuous, we cannot expect consistency at the discontinuity points. Despite the potential problem, the volatility estimator may still be used to construct a feasible likelihood ratio test, replacing  $\Sigma_t$  by  $\hat{\Sigma}_t$  in the definition of the LR statistic (13). Denoting the resulting statistic by  $\hat{L}R_n(r)$ , the next theorem establishes that the volatility estimation error has an asymptotically negligible effect on the asymptotic distribution of the LR test statistic, both under the null and under local alternatives.

**Theorem 2.** In model (2), under Assumptions 1–3 and under  $\mathcal{H}_n(r, r_1)$ , as  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $nh^2 \rightarrow \infty$ ,  $\hat{L}R_n(r)$  has the same limiting distribution as  $LR_n(r)$  as given in Theorem 1.

This theorem implies that under the stated conditions, adaptive testing is possible: the fact that the unknown volatility process is not observed but estimated nonparametrically entails no loss of efficiency.

We now discuss the selection of the window width  $h$ . Theorem 2 requires  $h$  to decrease with the sample size at a certain rate, but does not guide us in selecting a window width for a particular sample. A leave-one-out cross-validation technique may be defined as the  $h$  minimizing

$$CV_n(h) = \sum_{t=1}^n \|\hat{\Sigma}_t^{-t}(h) - e_t e_t'\|^2, \quad (20)$$

where  $\hat{\Sigma}_t^{-t}(h)$  is given by (19), but with  $K_h(0)$  replaced by 0, such that  $e_t e_t'$  does not enter the expression for  $\hat{\Sigma}_t^{-t}(h)$ .

In the context of stable VAR models, Patilea and Raïssa (2012) showed that an adaptivity result similar to Theorem 2 holds uniformly over  $h \in [\underline{h}_n, \bar{h}_n]$ , with the upper and lower bounds satisfying the rate condition Theorem 2. This provides an asymptotic justification of a constrained cross-validation procedure, where  $CV_n(h)$  in (20) is minimized over  $[\underline{h}_n, \bar{h}_n]$ . However, in practice such rate conditions impose very little on the interval for fixed  $n$ , and hence in the applications below we use unconstrained cross-validation. As discussed in Härdle, Hall, and Marron (1988) and Wasserman (2006), under suitable conditions, (unconstrained) leave-one-out cross-validation has the optimality property to deliver the bandwidth minimizing the average squared error associated with a nonparametric kernel estimator.

The above estimator uses a common window width to smooth all the elements of the matrix  $e_s e_s'$ . In practice, when different components of the underlying volatility matrix have

different degrees of smoothness, it may be preferable to use different window widths for different elements. In this case, the resulting covariance estimator is still symmetric, but not necessarily positive definite. As discussed in Fan, Li, and Yu (2012), one may then use a projection method to obtain a positive definite matrix.

## 4.2. Bootstrap

The limiting null distribution of the adaptive test depends on the volatility path. In this sub-section we consider two bootstrap implementations to approximate this distribution. In both cases, a bootstrap sample  $\{X_t^*\}_{t=1}^n$  is generated from the starting values  $\{X_t^* = X_t, t = 1 - k, \dots, 0\}$  as

$$\Delta X_t^* = \tilde{\alpha}_n \tilde{\beta}_n' X_{t-1}^* + \sum_{j=1}^{k-1} \tilde{\Gamma}_{jn} \Delta X_{t-j}^* + \varepsilon_t^*, \quad t = 1, \dots, n, \quad (21)$$

where  $\tilde{\alpha}_n$ ,  $\tilde{\beta}_n$ , and  $\tilde{\Gamma}_{jn}$  are the estimated parameter matrices from the model under the reduced rank restriction. Based on this bootstrap sample, one then estimates the unrestricted model and the restricted model, to get corresponding residuals  $\hat{\varepsilon}_t^*$  and  $\tilde{\varepsilon}_t^*$ ,  $t = 1, \dots, n$ , respectively, such that the bootstrap test statistic is computed as

$$\hat{L}R_n^*(r) = \sum_{t=1}^n \left( \tilde{\varepsilon}_t^{*'} \hat{\Sigma}_t^{-1} \tilde{\varepsilon}_t^* - \hat{\varepsilon}_t^{*'} \hat{\Sigma}_t^{-1} \hat{\varepsilon}_t^* \right).$$

Following Boswijk and Zu (2018), we propose to use the estimates  $\{\hat{\Sigma}_t\}_{t=1}^n$  from the original data in the calculation of  $\hat{L}R_n^*(r)$ , that is, we do not re-estimate  $\{\Sigma_t\}_{t=1}^n$  for each bootstrap sample.

In models with a constant term or time trend, the unrestricted and restricted models need be estimated using the methods discussed in Section 3. The bootstrap data-generating process (21) is then extended using the restricted estimates of  $\rho_0$  or  $(\mu, \rho_1)$ , see (17) and (18).

The two bootstrap methods we consider differ in the method for drawing the bootstrap errors. In a volatility bootstrap, we take  $\varepsilon_t^* = \hat{\sigma}_t z_t^*$ , where  $\hat{\sigma}_t$  is a matrix square root of  $\hat{\Sigma}_t$ , and  $z_t^*$  is iid  $N(0, I_p)$ . When  $k = 1$ ,  $X_0 = 0$  and the restricted model has rank  $r = 0$ , then it can be shown that this volatility bootstrap may be interpreted as a Monte Carlo simulation of the asymptotic null distribution of the test statistic, replacing the unknown  $\Sigma(u)$  by its estimate  $\hat{\Sigma}_n(u)$  and discretizing the continuous time processes and integrals using  $n + 1$  equidistant points in the unit time interval.

Alternatively, we consider the wild bootstrap, which has been considered in the literature on unit root and cointegration inference with nonstationary volatility by Cavaliere and Taylor (2008), Cavaliere, Rahbek, and Taylor (2010, 2014), Boswijk and Zu (2018), and Boswijk et al. (2016), among others. Here the bootstrap errors are constructed as  $\varepsilon_t^* = e_t w_t^*$ , where  $w_t^*$  is a scalar iid sequence with zero mean and unit variance. (The unrestricted least-squares residuals  $e_t$  could be replaced by the restricted or unrestricted ML residuals,  $\tilde{\varepsilon}_t$  or  $\hat{\varepsilon}_t$ , without affecting the main properties of the procedure.) The most common implementation in this literature is to take  $w_t^* \sim \text{iid } N(0, 1)$ ,

such that  $\varepsilon_t^* \sim N(0, e_t e_t')$ . This shows that the wild bootstrap may be seen as the limiting case of the volatility bootstrap as  $h \rightarrow 0$  (for fixed  $n$ ), such that  $\hat{\Sigma}_t \rightarrow e_t e_t'$ . Comparison of the two bootstrap implementations should therefore provide information about the usefulness of variance smoothing for simulation of the null distribution of the test statistic.

The asymptotic validity of both bootstrap procedures is stated in the next theorem.

**Theorem 3.** In model (2), under Assumptions 1–3 and under both the null  $\mathcal{H}(r)$  and local alternatives  $\mathcal{H}_n(r, r_1)$ , as  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $nh^2 \rightarrow \infty$ ,

$$\widehat{\text{LR}}_n^*(r) \xrightarrow{w} \left( \int_0^1 Z_0(s) dW(s) \right)' \left( \int_0^1 Z_0(s) Z_0(s)' ds \right)^{-1} \left( \int_0^1 Z_0(s) dW(s) \right).$$

The results in this section have been obtained under the assumption that the VAR lag order  $k$  is known. In practice, following Cavaliere et al. (2018), we propose to estimate  $k$  using the Bayesian information criterion (BIC) in the VAR model with no cointegration imposed. Because Cavaliere et al. (2018) worked under effectively the same nonstationary volatility assumptions as we do, we can directly apply their results; in particular Cavaliere et al.’s (2018, Lemma 3), proving consistency of the BIC-based estimator of  $k$ , and the proof of their Theorem 2, indicating that estimation uncertainty in  $k$  has no effect on the subsequent (bootstrap-based) determination of the cointegration rank.

### 5. Monte Carlo Simulation

In this section, we use Monte Carlo simulation methods to compare the finite sample performance of the two bootstrap versions of the adaptive likelihood ratio test with that of the wild bootstrap PLR test of Cavaliere, Rahbek, and Taylor (2010, 2014).

The simulation DGP is a VAR(2) process of dimension  $p = 2$ :

$$\begin{aligned} \Delta X_t &= n^{-1} \alpha_1 \beta_1' X_{t-1} + \Psi_1 \Delta X_{t-1} + \varepsilon_t, \\ \alpha_1 &= \begin{pmatrix} a \\ 0 \end{pmatrix}, \\ \beta_1 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ \Psi_1 &= \begin{pmatrix} 0 & 0.5 \\ 0 & 0.5 \end{pmatrix}, \end{aligned}$$

with  $X_0 = 0$ . For  $a = 0$ , the process  $X_t$  is integrated of order one ( $\Delta X_t$  is a stationary VAR(1) process), and this is used to study the size of tests for no cointegration, that is,  $\mathcal{H}(0)$ . We also consider the cases  $a = -15$  and  $a = -30$  to study the power of the tests for  $\mathcal{H}(0)$  against  $\mathcal{H}(2) \setminus \mathcal{H}(0)$ , under local alternatives  $\mathcal{H}_n(0, 1)$ .

The errors are defined as  $\varepsilon_t = \sigma_t z_t$ , with  $z_t \sim \text{iid } N(0, I_2)$ . Four versions of the unconditional variance matrix  $\Sigma_t = \sigma_t \sigma_t'$

are considered:

$$\begin{aligned} \Sigma_t^{(1)} &= \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \\ \Sigma_t^{(2)} &= v_t^{(2)} \Sigma, \\ \Sigma_t^{(3)} &= (v_t^{(3)} - 1)I_2 + \Sigma, \\ \Sigma_t^{(4)} &= v_t^{(4)} \Sigma \end{aligned}$$

which we label Case 1, 2, 3, and 4, respectively. We set  $\rho = 0.4$  (implying a moderate (average) degree of correlation between the components of  $\varepsilon_t$ ),

$$v_t^{(2)} = v_t^{(3)} = 0.5 + 2.5 \times \mathbf{1}_{[s,1]}(t/n), \tag{22}$$

with  $s = 0.8$ , and

$$\begin{aligned} v_t^{(4)} &= \exp(2H(t/n)), \\ dH(u) &= -\kappa H(u)du + \zeta dB(u), \\ u &\in [0, 1], \end{aligned} \tag{23}$$

with  $B(\cdot)$  a standard Brownian motion,  $\kappa = 1$  and  $\zeta = 1$ . Cases 1–3 are inspired by the simulations in Boswijk et al. (2016); Case 1 corresponds to homoscedasticity, and Cases 2 and 3 involve a deterministically changing  $\Sigma_t$ , with a late positive shift in the variances and covariances ( $\Sigma_t^{(2)}$ ), or in the variances only ( $\Sigma_t^{(3)}$ ). Case 4 corresponds to continuous variation in the variances and covariances, driven by a single realization of a log-Ornstein–Uhlenbeck stochastic volatility process.

We analyze results for sample sizes  $n \in \{500, 1000\}$ . A restricted constant term is included in the estimation. All experiments are run over 5000 Monte Carlo replications using  $B = 499$  bootstrap replications. In each replication, the VAR lag order is selected by the BIC in the unrestricted model with maximal lag order 5, and the window width for the volatility matrix estimation is selected using the leave-one-out cross-validation method. The tables report the empirical size (or actual rejection frequency under the null,  $a = 0$ ) and size-corrected power ( $a = -15$  and  $a = -30$ ) of all tests at the 5% nominal level.

In all tables, PLR-VBS and PLR-WBS indicate the volatility bootstrap and wild bootstrap versions of the pseudo-LR test (imposing homoscedasticity), respectively; ALR-VBS and ALR-WBS indicate the volatility and wild bootstrap based adaptive LR tests; and LR-VBS indicates the infeasible LR test based on known volatility (used both in computing the LR test statistic and in the volatility bootstrap), included to assess the effect of volatility estimation on test size and power.

Table 1 gives the size and size-corrected power of the tests for each of the four cases. Size distortions appear to be slightly larger for the adaptive test than for the PLR test, but are moderate in all cases, and clearly decrease with the sample size. The two versions of the bootstrap seem to lead to similar size and power, although the size distortions appear to be smallest for the volatility bootstrap. With a few exceptions, the power of the test slightly increases with the sample size, despite the fact that we are considering local alternatives (the error correction coefficient in the first equation is  $a/n$ ); this may be explained by the reduced volatility estimation error in larger samples. The adaptive tests are more powerful than the PLR test when the volatility is time-varying; their power is close to but falls slightly

**Table 1.** Size and power, 5% level.

Case	$n$	PLR-VBS	PLR-WBS	ALR-VBS	ALR-WBS	LR-VBS
Size						
1	500	0.046	0.045	0.057	0.051	0.037
	1000	0.051	0.050	0.056	0.057	0.047
2	500	0.053	0.061	0.089	0.073	0.041
	1000	0.059	0.048	0.074	0.070	0.039
3	500	0.058	0.073	0.075	0.067	0.048
	1000	0.057	0.060	0.069	0.058	0.055
4	500	0.055	0.086	0.099	0.092	0.040
	1000	0.052	0.075	0.084	0.068	0.049
Power $a = -15$						
1	500	0.252	0.248	0.235	0.239	0.265
	1000	0.241	0.248	0.235	0.244	0.262
2	500	0.209	0.219	0.200	0.215	0.263
	1000	0.201	0.197	0.242	0.244	0.260
3	500	0.156	0.156	0.597	0.601	0.666
	1000	0.157	0.154	0.659	0.671	0.720
4	500	0.089	0.089	0.175	0.175	0.242
	1000	0.077	0.078	0.195	0.207	0.246
Power $a = -30$						
1	500	0.747	0.750	0.732	0.737	0.772
	1000	0.770	0.783	0.762	0.776	0.797
2	500	0.595	0.607	0.641	0.673	0.756
	1000	0.614	0.611	0.746	0.758	0.791
3	500	0.459	0.462	0.969	0.970	0.979
	1000	0.498	0.497	0.991	0.992	0.994
4	500	0.285	0.284	0.584	0.606	0.729
	1000	0.296	0.309	0.678	0.698	0.766

NOTES: This table displays rejection frequencies under the null hypothesis (size) and under the alternative (power), using critical values simulated under the null hypothesis (size-corrected power). "Case" refers to the four different volatility specifications explained in the text, and  $n$  is the sample size. PLR, ALR, and LR are the pseudo-LR, the adaptive LR, and the (true-volatility) LR test, respectively; VBS and WBS refer to the volatility and wild bootstrap based tests, respectively.

short of the benchmark optimal power of the LR test with known volatility. In Case 1 (constant volatility), the adaptive tests and PLR tests have similar power, suggesting that there is no serious disadvantage to using the adaptive test even in such cases.

In the remainder of this section, we investigate the sensitivity of the size and power for parameter variations in Cases 2 and 4. For Case 2, we consider three possible values of the break time  $s$ ; see Table 2. We observe that the break time has little effect on the size of the PLR tests, but the size of the ALR tests increases slightly as the break happens toward the end of the sample, although the size distortion is still mild. On the other hand, the power of the two adaptive tests decreases slightly as the break happens toward the end of the sample, while the power of the PLR test decreases more. This illustrates that the potential for increasing the power relative to the PLR test is highest for the late positive break, which as shown by Cavaliere (2004) also has the largest effect on the asymptotic null distribution of PLR tests for a unit root.

Finally, we study the sensitivity of the simulation results for different volatility-of-volatility parameter values  $\zeta$  for the log-volatility  $H(u)$  in (23), Case 4; see Table 3. We first note that the size distortion of the bootstrap adaptive tests becomes more severe as the volatility-of-volatility parameter gets higher, with the volatility bootstrap performing slightly better than the wild bootstrap.

The power of the PLR test is very sensitive to the degree of variation in the volatility: in the high  $\zeta$  scenario, the power

**Table 2.** Size and power, 5% level, Case 2, varying break times  $s$ .

$s$	$n$	PLR-VBS	PLR-WBS	ALR-VBS	ALR-WBS	LR-VBS
Size						
0.2	500	0.045	0.043	0.063	0.042	0.037
	1000	0.053	0.048	0.062	0.048	0.043
0.5	500	0.049	0.055	0.074	0.062	0.048
	1000	0.053	0.053	0.063	0.061	0.046
0.8	500	0.053	0.061	0.089	0.073	0.041
	1000	0.059	0.048	0.074	0.070	0.039
Power $a = -15$						
0.2	500	0.336	0.331	0.290	0.276	0.279
	1000	0.308	0.309	0.269	0.279	0.280
0.5	500	0.259	0.253	0.248	0.229	0.284
	1000	0.266	0.261	0.271	0.269	0.281
0.8	500	0.209	0.219	0.200	0.215	0.263
	1000	0.201	0.197	0.242	0.244	0.260
Power $a = -30$						
0.2	500	0.820	0.825	0.784	0.778	0.787
	1000	0.837	0.843	0.796	0.804	0.816
0.5	500	0.692	0.689	0.724	0.706	0.784
	1000	0.725	0.720	0.783	0.783	0.810
0.8	500	0.595	0.607	0.641	0.673	0.756
	1000	0.614	0.611	0.746	0.758	0.791

NOTES: This table displays rejection frequencies under the null hypothesis (size) and under the alternative (power), using critical values simulated under the null hypothesis (size-corrected power).  $s$  refers to the fraction of the sample at which the break occurs in Case 2, and  $n$  is the sample size. PLR, ALR, and LR are the pseudo-LR, the adaptive LR, and the (true-volatility) LR test, respectively; VBS and WBS refer to the volatility and wild bootstrap based tests, respectively.

**Table 3.** Size and power, 5% level, Case 4, varying volatility of volatility  $\zeta$ .

$\zeta$	$n$	PLR-VBS	PLR-WBS	ALR-VBS	ALR-WBS	LR-VBS
Size						
0.5	500	0.051	0.072	0.071	0.070	0.033
	1000	0.050	0.068	0.066	0.067	0.048
1	500	0.055	0.086	0.099	0.092	0.040
	1000	0.052	0.075	0.084	0.068	0.049
2	500	0.067	0.093	0.184	0.117	0.040
	1000	0.060	0.074	0.139	0.079	0.063
Power $a = -15$						
0.5	500	0.147	0.149	0.204	0.205	0.241
	1000	0.151	0.155	0.200	0.209	0.244
1	500	0.089	0.089	0.175	0.175	0.242
	1000	0.077	0.078	0.195	0.207	0.246
2	500	0.040	0.040	0.111	0.121	0.356
	1000	0.032	0.034	0.209	0.208	0.374
Power $a = -30$						
0.5	500	0.513	0.522	0.663	0.673	0.739
	1000	0.567	0.585	0.710	0.719	0.775
1	500	0.285	0.284	0.584	0.606	0.729
	1000	0.296	0.309	0.678	0.698	0.766
2	500	0.087	0.084	0.270	0.332	0.733
	1000	0.079	0.090	0.557	0.567	0.789

NOTES: This table displays rejection frequencies under the null hypothesis (size) and under the alternative (power), using critical values simulated under the null hypothesis (size-corrected power).  $\zeta$  refers to the volatility of volatility in Case 4, and  $n$  is the sample size. PLR, ALR, and LR are the pseudo-LR, the adaptive LR, and the (true-volatility) LR test, respectively; VBS and WBS refer to the volatility and wild bootstrap based tests, respectively.

of the PLR test is rather low. Still, the power of the two adaptive tests seems stable, staying at a high level throughout all scenarios. Therefore, the high volatility-of-volatility scenario is a clear example where the adaptive tests outperform the PLR

test. In unreported additional simulations, we found similar but less pronounced effects from varying the mean-reversion parameter  $\kappa$ .

In all cases, the size and power properties of the tests does not seem to be affected too much by varying the sample size, and all tests perform reasonably well in a sample size typical for macroeconomic applications, an example of which is considered in the next section.

In summary, the Monte Carlo simulation results in this section indicate that the volatility bootstrap, based on the nonparametric volatility estimator, performs slightly better than the wild bootstrap in controlling the size of the adaptive test. Furthermore, we have seen that for various scenarios, the adaptive tests outperform the wild bootstrap based pseudo-LR test in terms of power.

## 6. Empirical Applications

### 6.1. U.S. Term Structure of Interest Rates

Boswijk et al. (2016) provided a cointegration analysis of the term structure of interest rates in the United States, showing that allowing for heteroscedasticity affects inference on the cointegration rank, as well as on structural hypotheses on the cointegrating vectors and the adjustment coefficients. Their analysis is based on wild bootstrap versions of Wald and likelihood ratio tests based a Gaussian iid pseudo-likelihood. We will investigate to what extent their empirical results change if we use adaptive tests as developed in this article.

We analyze the same time series  $X_t = (X_{1t}, \dots, X_{5t})'$  of monthly zero-coupon treasury yields, 1970:1–2009:12, for maturities equal to 3 months ( $X_{1t}$ ), 1 year ( $X_{2t}$ ), 3 years ( $X_{3t}$ ), 5 years ( $X_{4t}$ ), and 10 years ( $X_{5t}$ ). Following Boswijk et al. (2016), we estimate a VAR(2) model with a constant term for  $X_t$ , using observations on the first two months of 1970 as starting values; hence  $n = 478$ . The lag order  $k = 2$  is selected by the Hannan–Quinn information criterion, and supported by (wild bootstrap) residual serial correlation tests. Further details on the source of the data are provided by Boswijk et al. (2016).

**Table 4.** Cointegration test results for the U.S. term structure data.

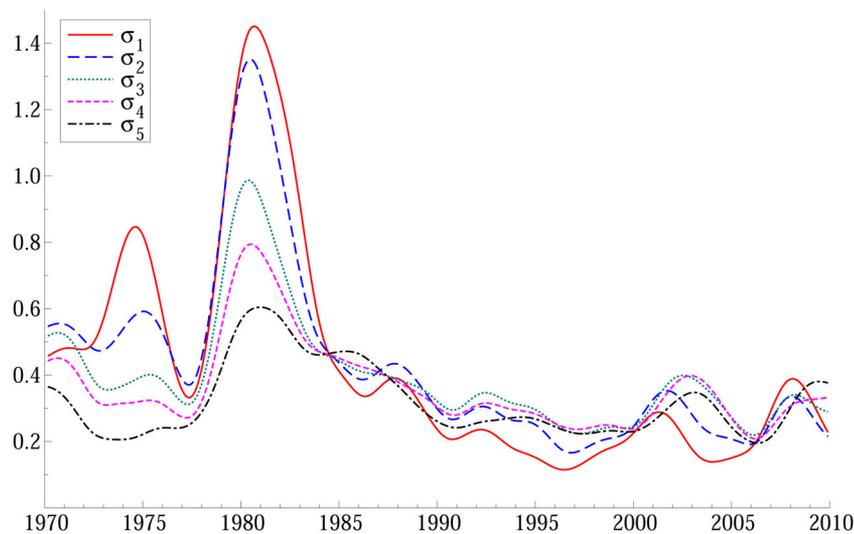
$r$	ALR-VBS	ALR-WBS	PLR-WBS
0	0.000	0.000	0.000
1	0.000	0.000	0.000
2	0.000	0.001	0.087
3	0.038	0.011	0.286
4	0.172	0.124	0.795

NOTES: This table displays  $p$ -values based on the volatility bootstrap (VBS) and wild bootstrap (WBS) of the adaptive (ALR) and pseudo-likelihood ratio (PLR) tests for cointegration rank in the U.S. term structure data.

Before we present the cointegration test results, Figure 1 displays the nonparametric estimates of the time-varying volatilities  $\sigma_{it}$  of  $X_{it}$ ,  $i = 1, \dots, 5$ . The window width chosen by cross-validation is  $h = 0.0217$ . We observe similar patterns in all five volatilities, with the most pronounced variation in the short-maturity interest rate. Most striking is the high volatility period around 1980, and the lower volatility after 1985, the period known as the Great Moderation.

To calculate the cointegration test statistics and their bootstrap  $p$ -values, we have implemented the switching algorithm discussed in Section 3 to implement the restricted maximum likelihood estimation. The tolerance level used for the switching algorithm is  $10^{-6}$ , which means that we stop the algorithm when the increase in the likelihood function is smaller than  $10^{-6}$ . The results, based on  $B = 999$  bootstrap replications, are given in Table 4; the final column (the wild bootstrap  $p$ -value of the constant-variance PLR statistic) is taken from Boswijk et al. (2016). As considered in the previous sections, all test are for  $\mathcal{H}(r) : \text{rank}(\Pi) \leq r$  against the alternative  $\mathcal{H}(p) \setminus \mathcal{H}(r) : r < \text{rank}(\Pi) \leq p$ .

We observe that using the adaptive test leads to a higher cointegration rank: whereas the PLR test would lead us to select a cointegrating rank  $r = 3$  only if we are willing to use a significance level of 10%, the adaptive tests lead to the conclusion of  $r = 4$  even if we use the conventional 5% significance level. This implies a single stochastic trend driving the five different yields. The adaptive ML estimators of  $\alpha$  and  $\beta$  for  $r = 4$  (with  $\beta$  normalized on  $X_{2t}, \dots, X_{5t}$ ) together with QMLE standard



**Figure 1.** Estimated volatilities, U.S. 3-month and 1-, 3-, 5-, and 10-year treasury yields.

errors, are as follows:

$$\hat{\alpha} = \begin{pmatrix} 0.282 & -0.030 & -0.019 & 0.003 \\ (0.085) & (0.143) & (0.131) & (0.051) \\ -0.234 & 0.356 & -0.231 & 0.035 \\ (0.092) & (0.162) & (0.149) & (0.055) \\ -0.125 & -0.022 & 0.009 & 0.040 \\ (0.083) & (0.157) & (0.143) & (0.050) \\ -0.178 & 0.113 & -0.111 & 0.058 \\ (0.077) & (0.149) & (0.133) & (0.044) \\ -0.149 & -0.114 & 0.218 & -0.085 \\ (0.071) & (0.134) & (0.121) & (0.041) \end{pmatrix},$$

$$\hat{\beta} = \begin{pmatrix} -1.092 & -1.213 & -1.238 & -1.227 \\ (0.040) & (0.128) & (0.173) & (0.213) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We may be interested in the hypothesis that the first row of  $\beta$  are all  $-1$ 's; this corresponds to the hypothesis that the single stochastic trend affects only the height, and not the slope or curvature of the yield curve (so that all spreads are stationary). The adaptive likelihood ratio statistic for this hypothesis is 14.650 with volatility and wild bootstrap  $p$ -values of 0.030 and 0.012, respectively, indicating that this hypothesis is rejected. Note that the asymptotic theory for adaptive likelihood ratio tests for hypotheses on  $\beta$  has not been developed in this article, but following the analysis of Boswijk et al. (2016), we expect the bootstrap to yield asymptotically valid inference again.

## 6.2. Purchasing Power Parity

As discussed in Section 2.4 of Froot and Rogoff (1995), long-run purchasing power parity (PPP) can be formulated as the hypothesis of cointegration between the log-exchange rate  $s_t$  of

a currency against the U.S. dollar and the log relative price level  $p_t - p_t^*$  (domestic relative to the United States). The cointegrating vector  $\beta$  for  $X_t = (s_t, p_t - p_t^*)'$  implied by PPP is  $\beta = (1, -1)'$ , implying stationarity of the real exchange rate. Boswijk and Klaassen (2012) showed that allowing for conditional heteroscedasticity can lead to stronger evidence in favor of long-run PPP from unit root tests applied to real exchange rates (especially at higher observation frequencies). It is therefore of interest to see if a similar conclusion holds based on multivariate tests allowing for unconditional heteroscedasticity, as developed in this article.

We investigate the cointegration hypothesis for the United Kingdom, Germany, and Japan, all relative to the United States. Nominal exchange rate and Consumer Price Index data have been obtained from the Bank of International Settlements' website (<https://www.bis.org/statistics/xrusd.htm> and <https://www.bis.org/statistics/cp.htm>, respectively). We use monthly observations from January 1968 to August 2019, in total 620 observations. For Germany, the exchange rate after 2001 has been converted from the EUR/USD exchange rate.

Separate bivariate VAR models with a constant and linear trend have been selected and estimated for each of the three currencies. The linear trend has been included primarily to allow for divergent inflation rates and hence linear trends in the relative log-price data. The VAR order selected by BIC is  $k = 1$  in all cases. The residuals from this model have been used for nonparametric estimation of the time-varying exchange rate and price volatilities ( $\sigma_1$  and  $\sigma_2$ , respectively) and their correlation  $\rho$ , with the bandwidths selected by cross-validation. These are displayed in Figure 2. We observe an initial increase in exchange rate volatility (as expected from the collapse of the Bretton Woods system in the early 1970s), followed by a slight decrease in recent decades. The relative price volatilities show a downward trend, in particular for the United Kingdom

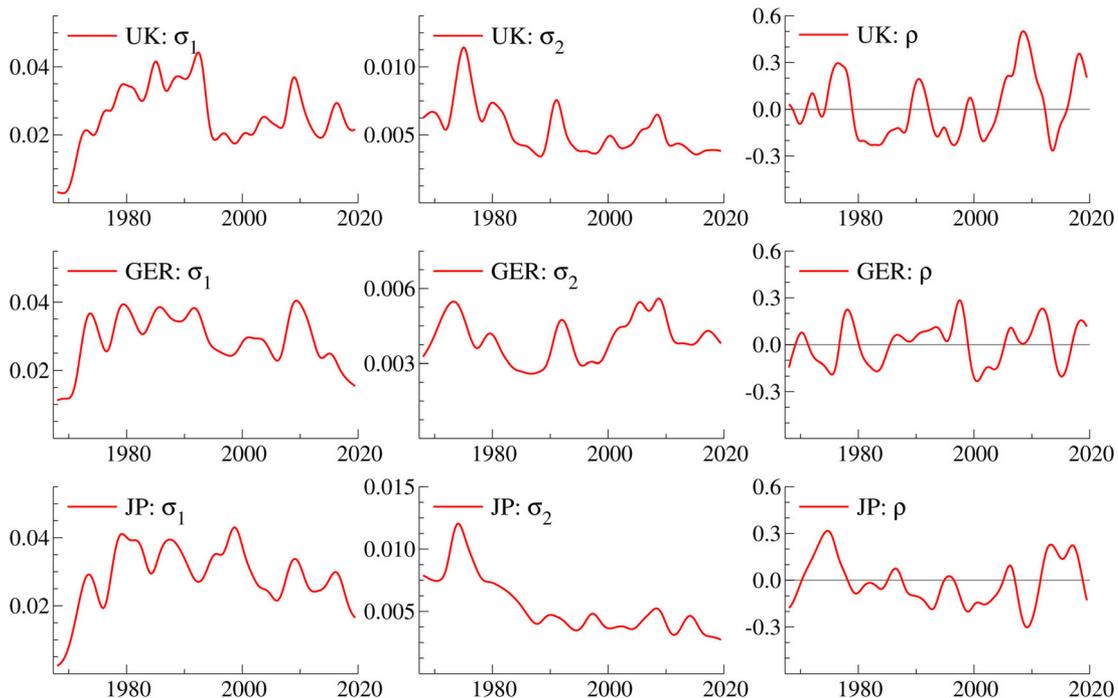


Figure 2. Estimated volatilities ( $\sigma_1, \sigma_2$ ) and correlations ( $\rho$ ) of effective exchange rates and relative prices of the United Kingdom, Germany (GER), and Japan (JP) vis-à-vis the United States.

**Table 5.** Cointegration test results for the PPP data.

Country	PLR	ALR	PLR-VBS	PLR-WBS	ALR-VBS	ALR-WBS
United Kingdom	33.836	23.754	0.006	0.007	0.047	0.024
Germany	30.663	32.406	0.010	0.004	0.004	0.001
Japan	34.055	25.453	0.009	0.004	0.034	0.013

Notes: This table displays test statistics and  $p$ -values based on the volatility bootstrap (VBS) and wild bootstrap (WBS) for the adaptive (ALR) and pseudo-likelihood ratio (PLR) tests for the null hypothesis of no cointegration in the PPP data.

and Japan, which may be related to the changes in monetary policy and the resulting Great Moderation since the mid-1980s. Correlations appear to fluctuate around zero, without a clearly interpretable pattern.

The cointegration test results for no cointegration ( $\mathcal{H}(0)$ ) against the alternative of at least one cointegrating vector, based on 999 bootstrap replications, are presented in Table 5. We observe that both the PLR tests and the ALR tests reject the no-cointegration hypothesis in all cases at the 5% level (although the volatility bootstrap-based adaptive test for the United Kingdom has a  $p$ -value close to 5%).

In all three cases, the estimated cointegrating vectors differ substantially but not significantly from the theoretically expected  $\beta = (1, -1)'$ . This is a reflection of the fact that standard errors are rather large, due to the presence of a restricted trend in the cointegrating relation (which itself has a coefficient not significantly different from zero in all cases). In summary, we find that both testing approaches find evidence in favor of long-run PPP. We observe that in this application, the adaptive tests do not yield more evidence for cointegration than the nonadaptive test; this may be related to the fact that the time variation in volatility in this example is less pronounced than in the term structure example.

## 7. Discussion

In this article, we have proposed a new class of cointegration tests, which have higher power than existing tests by exploiting time variation in the unconditional error variance matrix. Monte Carlo simulations have indicated that a bootstrap implementation of the test has good size and power properties for moderately sized samples. Two empirical examples have illustrated that applying our newly developed tests can indeed lead to stronger evidence for cointegration than alternative tests. These applications are related to the fixed income and foreign exchange markets, but the methods could also be applied to equity markets (analyzing statistical arbitrage) or macro-economic relations.

The theory and methods used in this article can be extended in various directions. First, we have considered time-variation in the unconditional variance matrix only. The analysis could be extended to also allow for conditional heteroscedasticity, but this would lead to a more complicated likelihood analysis (e.g., of a DCC-VAR model), and furthermore one would need to allow for the time-varying unconditional variance matrix in the estimation and identification of the conditional variance process. Second, our tests are based on a Gaussian likelihood function. In practice, one often observes that standardized financial returns still display excess kurtosis. The asymptotic results in this article are robust to this type of nonnormality,

but in such cases more powerful tests could be derived from, for example, a Student's  $t$  likelihood.

## Supplementary Materials

Ox code and data files that can be used to replicate the empirical and simulation results are provided as supplementary materials. Further details are provided in the readme.txt file.

## Acknowledgments

Helpful comments and suggestions from the associate editor and three anonymous reviewers are gratefully acknowledged.

## References

- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625. [5]
- Beare, B. K. (2018), "Unit Root Testing With Unstable Volatility," *Journal of Time Series Analysis*, 39, 816–835. [1,3]
- Boswijk, H. P. (1995), "Identifiability of Cointegrated Systems," Tinbergen Institute Discussion Paper # 95-78, available at <http://dare.uva.nl/document/2/163397>. [3]
- (2001), "Testing for a Unit Root With Near-Integrated Volatility," Tinbergen Institute Discussion Paper # 01-077/4, <http://papers.tinbergen.nl/01077.pdf>. [1]
- Boswijk, H. P., Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2016), "Inference on Co-Integration Parameters in Heteroskedastic Vector Autoregressions," *Journal of Econometrics*, 192, 64–85. [3,6,7,9,10]
- Boswijk, H. P., and Klaassen, F. (2012), "Why Frequency Matters for Unit Root Testing in Financial Time Series," *Journal of Business & Economic Statistics*, 30, 351–357. [10]
- Boswijk, H. P., and Zu, Y. (2018), "Adaptive Wild Bootstrap Tests for a Unit Root With Non-Stationary Volatility," *Econometrics Journal*, 21, 87–113. [1,2,5,6]
- Cavaliere, G. (2004), "Unit Root Tests Under Time-Varying Variances," *Econometric Reviews*, 23, 259–292. [1,2,8]
- Cavaliere, G., De Angelis, L., Rahbek, A., and Taylor, A. M. R. (2018), "Determining the Cointegration Rank in Heteroskedastic VAR Models of Unknown Order," *Econometric Theory*, 34, 349–382. [2,7]
- Cavaliere, G., Rahbek, A., and Taylor, A. M. R. (2010), "Testing for Cointegration in Vector Autoregressions With Non-Stationary Volatility," *Journal of Econometrics*, 158, 7–24. [2,3,6,7]
- (2014), "Bootstrap Determination of the Co-Integration Rank in Heteroskedastic VAR Models," *Econometric Reviews*, 33, 606–650. [2,3,6,7]
- Cavaliere, G., and Taylor, A. M. R. (2007), "Testing for Unit Roots in Time Series Models With Nonstationary Volatility," *Journal of Econometrics*, 140, 919–947. [1,2]
- (2008), "Bootstrap Unit Root Tests for Time Series With Nonstationary Volatility," *Econometric Theory*, 24, 34–71. [1,6]
- Engle, R. F. (2002), "Dynamic Conditional Correlation: A Simple Class of Multivariate GARCH Models," *Journal of Business & Economic Statistics*, 20, 339–350. [5]
- Fan, J., Li, Y., and Yu, K. (2012), "Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection," *Journal of the American Statistical Association*, 107, 412–428. [6]
- Froot, K. A., and Rogoff, K. (1995), "Perspectives on PPP and Long-Run Real Exchange Rates," *Handbook of International Economics*, 3, 1647–1688. [10]
- Giné, E., and Zinn, J. (1990), "Bootstrapping General Empirical Measures," *The Annals of Probability*, 18, 851–869. [2]
- Hansen, B. E. (1992), "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.
- (1995), "Regression With Nonstationary Volatility," *Econometrica*, 63, 1113–1132. [5,6]

- (2003), “Structural Changes in the Cointegrated Vector Autoregressive Model,” *Journal of Econometrics*, 114, 261–295. [1,3,4]
- Hansen, P. R., and Johansen, S. (1998), *Workbook on Cointegration*, Oxford: Oxford University Press. [3]
- Härdle, W., Hall, P., and Marron, J. S. (1988), “How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?,” *Journal of the American Statistical Association*, 83, 86–95. [6]
- Johansen, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (2nd ed.), Oxford: Oxford University Press. [1,2,3,5]
- (1997), “Likelihood Analysis of the  $I(2)$  Model,” *Scandinavian Journal of Statistics*, 24, 433–462.
- Kim, T.-H., Leybourne, S., and Newbold, P. (2003), “Unit Root Tests With a Break in Innovation Variance,” *Journal of Econometrics*, 109, 365–387. [1]
- Ling, S., and Li, W. K. (1998), “Limiting Distributions of Maximum Likelihood Estimators for Unstable Autoregressive Moving-Average Time Series With GARCH Errors,” *The Annals of Statistics*, 26, 84–125. [1]
- (2003), “Asymptotic Inference for Unit Root Processes With GARCH(1, 1) Errors,” *Econometric Theory*, 19, 541–564. [1]
- Li, W. K., Ling, S., and Wong, H. (2001), “Estimation for Partially Nonstationary Multivariate Autoregressive Models With Conditional Heteroskedasticity,” *Biometrika*, 88, 1135–1152. [1]
- Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus With Applications in Statistics and Econometrics*, Chichester: Wiley.
- Nelson, D. B. (1996), “Asymptotic Filtering Theory for Multivariate ARCH Models,” *Journal of Econometrics*, 71, 1–47. [5]
- Oberhofer W., and Kmenta, J. (1974), “A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models,” *Econometrica*, 42, 579–590. [3]
- Patilea, V., and Raïssi, H. (2012), “Adaptive Estimation of Vector Autoregressive Models With Time-Varying Variance: Application to Testing Linear Causality in Mean,” *Journal of Statistical Planning and Inference*, 142, 2891–2912. [5,6]
- Phillips, P. C. B., and Xu, K.-L. (2006), “Inference in Autoregression Under Heteroskedasticity,” *Journal of Time Series Analysis*, 27, 289–308.
- Sargan, J. D. (1964), “Wages and Prices in the United Kingdom: A Study in Econometric Methodology,” in *Econometric Analysis for National Economic Planning*, eds. P. E. Hart, G. Mills, and J. K. Whitaker, London: Butterworth, pp. 25–63. [3]
- Seo, B. (1999), “Distribution Theory for Unit Root Tests With Conditional Heteroskedasticity,” *Journal of Econometrics*, 91, 113–144. [1]
- (2007), “Asymptotic Distribution of the Cointegrating Vector Estimator in Error Correction Models With Conditional Heteroskedasticity,” *Journal of Econometrics*, 137, 68–111. [1]
- Van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.
- Wasserman, L. (2006), *All of Nonparametric Statistics*, New York: Springer-Verlag. [6]
- Wong, H., Li, W. K., and Ling, S. (2005), “Joint Modeling of Cointegration and Conditional Heteroskedasticity With Applications,” *Annals of the Institute of Statistical Mathematics*, 57, 83–103. [1]
- Xu, K.-L., and Phillips, P. C. B. (2008), “Adaptive Estimation of Autoregressive Models With Time-Varying Variances,” *Journal of Econometrics*, 142, 265–280. [3,5]