Adaptive Testing for Cointegration with Nonstationary Volatility

Boswijk, H.P.; Zu, Yang

Publication date
2019
Document Version
Final published version

TI 2019-043/III
Tinbergen Institute Discussion Paper

Adaptive Testing for Cointegration with Nonstationary Volatility

H. Peter Boswijk[1]
Yang Zu[2]

[1] University of Amsterdam
[2] University of Nottingham

Adaptive Testing for Cointegration with Nonstationary Volatility*

H. Peter Boswijk
Tinbergen Institute & Amsterdam School of Economics,
University of Amsterdam†

Yang Zu
School of Economics, University of Nottingham

June, 2019

Abstract

This paper generalises Boswijk and Zu (2018)'s adaptive unit root test for time series with nonstationary volatility to a multivariate context. 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 by Cavaliere *et al.* (2010, 2014). 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. Two empirical examples illustrate the applicability of the tests.

Key words: Adaptive estimation; Nonparametric volatility estimation; Wild bootstrap.

*An earlier version of this paper was titled "Testing for Cointegration with Nonstationary Volatility".

†Address for correspondence: Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands. E-mail: H.P.Boswijk@uva.nl.

1 Introduction

An important approach to the analysis of cointegrated time series is based on a likelihood analysis of the Gaussian vector autoregressive 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. Thus, for example, the asymptotic critical values for the likelihood ratio 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 (generalised) autoregressive-conditional heteroskedastic ((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 *et al.* (2001), Wong *et al.* (2005) and Seo (2007).

Recent developments in the univariate unit root literature, however, have emphasised 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 loose their asymptotic validity. Possible causes of such nonstationarity include level shifts or other deterministic trending patterns in the volatility, see Kim *et al.* (2002) and Cavaliere (2004), but also (near-) integrated GARCH dynamics, see Boswijk (2001). Cavaliere and Taylor (2007) and Beare (2018) develop two alternative approaches to constructing unit root test statistics with the conventional (Dickey-Fuller) asymptotic null distribution, and Cavaliere and Taylor (2008) show that application of the wild bootstrap leads to asymptotically valid inference. Boswijk and Zu (2018) derive the power envelope for unit root tests with observable (nonstationary) volatility, and show that considerable power gains may be obtained relative to procedures that do not take the heteroskedasticity into account. They also show that when the volatility is unobserved, the power envelope may be reached by an adaptive procedure based on nonparametric volatility estimation.

This paper seeks to extend Boswijk and Zu (2018)'s analysis to a multivariate context, and hence develop efficient tests for cointegration in the presence of nonstationary multivariate (unconditional) heteroskedasticity. First, building on the analysis of Hansen (2003), we derive the likelihood ratio test for cointegration in a vector autoregressive 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, this estimator is consistent and hence the resulting cointegration test has the same asymptotic power function as in the case of known volatility. The asymptotic 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 non-parametric volatility estimate.

In a related line of research, Cavaliere *et al.* (2010, 2014) show that application of the wild bootstrap to the traditional (constant-variance) Gaussian pseudo-likelihood ratio statistic leads to a correctly sized cointegration test in the presence of nonstationary volatility. They allow for discontinuous level shifts in the volatility, which are excluded in the present paper because the nonparametric estimator requires continuous volatility sample paths. Because they focus on the constant-variance pseudo-likelihood ratio test, their analysis does not exploit the power gain potential in the presence of nonstationary volatility.

The plan of the paper is as follows. Section 2 presents the model and assumptions, and characterises the limiting behaviour of the process. In Section 3 we obtain an expression of the likelihood ratio 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 S&P 500 and NASDAQ-100 indices in the 1990s, and to the term structure of interest rates in the US. Section 7 contains some concluding remarks, and proofs of all results are given in the Appendix.

Throughout the paper, 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, 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 $\|A\| = \sqrt{\mathrm{tr}(A'A)}$.

2 The model

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

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

where Π and $\Gamma_j, j = 1, \ldots, k-1$ are $p \times p$ coefficient matrices, and where ε_t is a p-variate disturbance vector with mean zero. The starting values $\{X_{1-k}, \ldots, X_0\}$ 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 α and β 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

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

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

Assumption 1 *In the model (2),* (a) *the $p \times r$ matrices α and β 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 (i.i.d.) Gaussian with mean zero and positive definite variance matrix Σ, the likelihood function for the model under $\mathcal{H}(r)$ is maximised by reduced rank regression. From this, an explicit expression is available for the likelihood ratio test of $\mathcal{H}(r)$ in the unrestricted model $\mathcal{H}(p)$, corresponding to (1); see Johansen (1996). Here we consider a deviation from the i.i.d. assumption, in that we allow for unconditional heteroskedasticity:

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

$$\varepsilon_t = \sigma_t z_t, \qquad t = 1, \ldots, n,$$

where:

(a) $\sigma_t = \sigma(t/n)$, where $\sigma(\cdot)$ is a non-stochastic $p \times p$ matrix-valued function on $[0,1]$, such that $\sigma(u)$ is non-singular for all $u \in [0,1]$, and continuous in $u \in [0,1]$;

(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 $\sup_{t\geq 1} E(\|z_t\|^{4m}) < \infty$ for some $m > 1$.

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 σ_t, a matrix square root of Σ_t, as the *volatility matrix* of ε_t. In the next section, we will analyse 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 Σ as $n \to \infty$ 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-) likelihood ratio test, based on the Gaussian i.i.d. assumption on $\{\varepsilon_t\}_{t\geq 1}$, would retain its usual asymptotic properties, even though more efficient tests may 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 Beare (2018), *inter alia*. The multivariate analog was considered by Cavaliere *et al.* (2010, 2014) and Boswijk *et al.* (2016). In particular, Assumption 2 is a specific version of Assumption 2 of Boswijk *et al.* (2016), who allow for a countable number of discontinuities in $\sigma_{ij}(\cdot)$, and for conditional heteroskedasticity in $\{z_t\}_{t\geq 1}$. The assumption of continuity is made here to facilitate consistent non-parametric estimation of $\sigma(\cdot)$, although Xu and Phillips (2008) show that this could be relaxed for adaptive estimation. For practical purposes, the continuity assumption can still accommodate large changes in volatility, as discussed in Boswijk and Zu (2018) for the univariate case. The analysis in the present paper could be extended to allow for conditional heteroskedasticity, 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 behaviour of the process under the null $\mathcal{H}(r)$, and under a sequence of local alternatives

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

where α and β are the same as before, and α_1 and β_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 likelihood ratio test under the Gaussian i.i.d. assumption and (3). Proofs of all results are given in the Appendix.

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

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

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

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

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

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

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

$$dU_A(u) = AU_A(u)du + \sigma_U(u)dW(u). \tag{5}$$

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 heteroskedastic 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 pseudo-likelihood ratio (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 *et al.*, 2010) that the PLR statistic satisfies, under the null hypothesis,

$$\begin{aligned}&\mathrm{PLR}_n(0)\\ &\xrightarrow{w} \mathrm{tr}\left\{\left(\int_0^1 \Sigma(u)du\right)^{-1}\int_0^1 dM(u)M(u)'\left(\int_0^1 M(u)M(u)'du\right)^{-1}\int_0^1 M(u)dM(u)'\right\}.\end{aligned}$$

If and only if $\sigma(\cdot)$ is a constant matrix σ, such that $M(u) = \sigma W(u)$, the usual limiting distribution tabulated in Johansen (1996) will result.

3 The likelihood ratio test with known volatility

In this section we analyse the likelihood ratio (LR) statistic for $\mathcal{H}(r)$ in the model (2) in the case where $\{\sigma_t\}_{t=1}^n$ is known, and where the standardised innovations $\{z_t\}_{t=1}^n$ are taken to be i.i.d. $N(0, I_p)$. 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 upper bound[1] to 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 $\Psi = [\Gamma_1 : \ldots : \Gamma_{k-1}]$ and $W_t = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}')'$, such that the model (2) under Assumption 2 with Gaussian $\{\sigma_t\}_{t=1}^n$ may be expressed more compactly as

$$\Delta X_t = \alpha\beta' X_{t-1} + \Psi W_t + \varepsilon_t, \qquad \varepsilon_t|\mathcal{X}_{t-1} \sim N(0, \Sigma_t), \qquad t = 1, \ldots, n, \tag{6}$$

where $\Sigma_t = \sigma_t\sigma_t'$ as before, and $\mathcal{X}_{t-1} = \{X_{t-1}, \ldots, X_1, X_0, \ldots, X_{1-k}\}$. 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

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

Maximum likelihood estimation in a closely related class of models was studied by Hansen (2003), who generalised the switching algorithm developed by Boswijk (1995) in various directions, including time-varying variance matrices. The key idea of this so-called *generalised reduced rank regression* procedure is that, although no closed-form expression exists for the maximum likelihood estimator (MLE) $(\widetilde{\alpha}_n, \widetilde{\beta}_n, \widetilde{\Psi}_n)$, the maximization of $\ell_n(\alpha, \beta, \Psi)$ over (α, Ψ)

[1] We confine ourselves to likelihood-ratio-type tests, and hence do not attempt to derive an asymptotic power envelope for *all* possible tests of the null hypothesis, which would be defined as the limiting power of a point optimal invariant test of $\mathcal{H}(r)$ against $\mathcal{H}_n(r, r_1)$.

for fixed β does lead to a closed-form expression, and similarly the MLE of β for fixed (α, Ψ) has a closed-form expression. The likelihood may then be maximised, starting from an initial guess, by switching between maximization over (α, Ψ) and β. 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 β. 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 Φ is a $(p-r) \times r$ matrix of free parameters, such that

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

where $h = \operatorname{vec} \bar{c}$ and $H = I_r \otimes c_\perp$, and $\phi = \operatorname{vec} \Phi$. 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 β.

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

$$\operatorname{vec}[\widetilde{\alpha}_n(\beta) : \widetilde{\Psi}_n(\beta)] = \left(\sum_{t=1}^n \left[Z_t(\beta) Z_t(\beta)' \otimes \Sigma_t^{-1} \right] \right)^{-1} \operatorname{vec} \left(\sum_{t=1}^n \Sigma_t^{-1} \Delta X_t Z_t(\beta)' \right), \tag{9}$$

whereas the MLE of β for fixed (α, Ψ) is given by

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

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

$$\widetilde{\varepsilon}_t = \Delta X_t - [\widetilde{\alpha}_n : \widetilde{\Psi}_n] Z_t(\widetilde{\beta}_n) = \Delta X_t - \widetilde{\alpha}_n \widetilde{\beta}'_n X_{t-1} - \widetilde{\Psi}_n W_t, \qquad t = 1, \ldots, n.$$

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

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 maximised by

$$\operatorname{vec}[\widehat{\Pi}_n : \widehat{\Psi}_n] = \left(\sum_{t=1}^n \left[Z_t Z'_t \otimes \Sigma_t^{-1} \right] \right)^{-1} \operatorname{vec} \left(\sum_{t=1}^n \Sigma_t^{-1} \Delta X_t Z'_t \right), \tag{11}$$

yielding the unrestricted residuals

$$\widehat{\varepsilon}_t = \Delta X_t - [\widehat{\Pi}_n : \widehat{\Psi}_n] Z_t = \Delta X_t - \widehat{\Pi}_n X_{t-1} - \widehat{\Psi}_n W_t, \qquad t = 1, \ldots, n. \tag{12}$$

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

$$\mathrm{LR}_n(r) = -2 \left[\ell_n(\widetilde{\alpha}_n, \widetilde{\beta}_n, \widetilde{\Psi}_n) - \ell_n(\widehat{\Pi}_n, I_{p+1}, \widehat{\Psi}_n) \right] = \sum_{t=1}^n \left(\widetilde{\varepsilon}'_t \Sigma_t^{-1} \widetilde{\varepsilon}_t - \widehat{\varepsilon}'_t \Sigma_t^{-1} \widehat{\varepsilon}_t \right). \tag{13}$$

The limiting behaviour of $\mathrm{LR}_n(r)$ is characterised in Theorem 1. Define

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

and

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

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

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

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 α (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), discretising the integrals and replacing α by $\widetilde{\alpha}_n$. Consistency of $\widetilde{\alpha}_n$ (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 discretisation 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

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

Corollary 1 *Under the conditions of Theorem 1, the likelihood ratio statistic* $\mathrm{LR}_n(0)$ *for $r = 0$ satisfies, as $n \to \infty$,*

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

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

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

where ρ_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 μ is an n-vector and ρ_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 $\operatorname{vec} \beta^{\#} = \operatorname{vec}\left(\bar{c}^{\#} + c_{\perp}^{\#}\Phi^{\#}\right)$ 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)–(10) remains the same, with all parameters and vectors replaced by their "#" counterparts. Without proof, we state the limiting distribution of the resulting LR test statistic in the following corollary.

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

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

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

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

4 Adaptive likelihood ratio test

4.1 Volatility estimation

In the previous section we have developed a likelihood ratio 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 $\{X_t, t = 1-k, \ldots, 0, 1, \ldots, n\}$, 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 σ_t may be estimated either parametrically or nonparametrically. In the presence of conditional heteroskedasticity, possible parametric approaches include multivariate GARCH models, notably the dynamic conditional correlation (DCC) model of Engle (2002). The likelihood ratio 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 $\widehat{\Sigma}_t$ is then substituted for Σ_t in the expressions for the MLE and LR statistic given in the previous section. In this paper we propose to estimate σ_t by a nonparametric kernel estimator, generalising the approach of Boswijk and Zu (2018), which in turn is based on Hansen (1995). It should be noted, however, that as analysed 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) shows 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 extend Hansen (1995)'s nonparametric volatility filter in two directions: we consider a multivariate version of the estimator, but we also propose a version of the variance matrix estimator at time t based on *leads and lags* of the outer product of the residual vector, to increase efficiency of the estimator. Such an approach to adaptive estimation was also considered by Xu and Phillips (2008) and Patilea and Raïssi (2012).

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)–(12) with $\Sigma_t = I_n$. 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 Σ_t is defined as

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

By choosing different kernel functions, one could consider both one-sided smoothing (or filtering, where $\widehat{\Sigma}_t$ is based on lags of $e_t e_t'$ only), or two-sided smoothing (based on leads and lags). Note, however, that consistency 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 below imposes two-sided smoothing.

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

Define $\widehat{\Sigma}_n(u) := \sum_{t=1}^{n} \widehat{\Sigma}_t \mathbf{1}_{[(t-1)/n,t/n)}(u)$ as the variance matrix process implied by the n nonparametrically estimated covariance matrices and recall that $\Sigma(u) := \sigma(u)\sigma(u)'$ is the true variance matrix process. We next show that the estimated variance matrix process is uniformly consistent over the compact interval $[0, 1]$.

Lemma 2 *Consider the model (2) under Assumptions 1–3. If $n \to \infty, h \to 0$ and $nh^2 \to \infty$, then under both the null $\mathcal{H}(r)$ and the local alternatives $\mathcal{H}_n(r, r_1)$,*

$$\sup_{u\in[0,1]} \|\widehat{\Sigma}_n(u) - \Sigma(u)\| \xrightarrow{p} 0. \tag{20}$$

The above uniform consistency result for the estimated process $\widehat{\Sigma}_n(u)$ clearly implies uniform consistency of the nonparametric estimator $\widehat{\Sigma}_t$ over $t = 1, \ldots, n$. These consistent estimators may be used to construct a feasible likelihood ratio test by replacing Σ_t by $\widehat{\Sigma}_t$ in the definition of the likelihood ratio statistic (13). Denoting the resulting statistic by $\widehat{\mathrm{LR}}_n(r)$, the next theorem establishes that the volatility estimation error has an asymptotically negligible effect on the asymptotic distribution of the likelihood ratio test, under both the null and the local alternatives.

Theorem 2 *In the model (2), under Assumptions 1–3 and under $\mathcal{H}_n(r, r_1)$, as $n \to \infty$, $h \to 0$ and $nh^2 \to \infty$, $\widehat{\mathrm{LR}}_n(r)$ has the same limiting distribution as $\mathrm{LR}_n(r)$ 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. Lemma 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 minimising

$$\mathrm{CV}_n(h) = \sum_{t=1}^{n} \|\widehat{\Sigma}_t^{-t}(h) - e_t e_t'\|^2, \tag{21}$$

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

In a similar context of vector autoregressive models, Patilea and Raïssi (2012) show that the result of Lemma 2 holds uniformly over $h \in [\underline{h}_n, \overline{h}_n]$, with the upper and lower bounds satisfying the rate condition Lemma 2. This provides an asymptotic justification of a constrained cross-validation procedure, where $\mathrm{CV}_n(h)$ in (21) is minimised over $[\underline{h}_n, \overline{h}_n]$. 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.

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 *et al.* (2012), one may 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 $\{X_t^* = X_t, t = 1-k, \ldots, 0\}$ as

$$\Delta X_t^* = \widetilde{\alpha}_n \widetilde{\beta}_n' X_{t-1}^* + \sum_{j=1}^{k-1} \widetilde{\Gamma}_{jn} \Delta X_{t-j}^* + \varepsilon_t^*, \qquad t = 1, \ldots, n,$$

where $\widetilde{\alpha}_n$, $\widetilde{\beta}_n$, and $\widetilde{\Gamma}_{jn}$ 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 $\widehat{\varepsilon}_t^*$ and $\widetilde{\varepsilon}_t^*$, $t = 1, \ldots, n$, respectively, such that the bootstrap test statistic is computed as

$$\widehat{\mathrm{LR}}_n^*(r) = \sum_{t=1}^n \left(\widetilde{\varepsilon}_t^{*\prime} \widehat{\Sigma}_t^{-1} \widetilde{\varepsilon}_t^* - \widehat{\varepsilon}_t^{*\prime} \widehat{\Sigma}_t^{-1} \widehat{\varepsilon}_t^* \right).$$

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

The two bootstrap methods we consider differ in the method for drawing the bootstrap errors. In a volatility bootstrap, we take $\varepsilon_t^* = \widehat{\sigma}_t z_t^*$, where $\widehat{\sigma}_t$ is a matrix square root of $\widehat{\Sigma}_t$, and z_t^* is i.i.d. $N(0, I_p)$. 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 $\widehat{\Sigma}_n(u)$ and discretising 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 *et al.* (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 i.i.d. sequence with zero mean and unit variance. (The unrestricted least-squares residuals e_t could be replaced by the restricted or unrestricted ML residuals, $\widetilde{\varepsilon}_t$ or $\widehat{\varepsilon}_t$, without affecting the main properties of the procedure.) The most common implementation in this literature is to take $w_t^* \sim$ i.i.d. $N(0,1)$, 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 $h \to 0$ (for fixed n), such that $\widehat{\Sigma}_t \to e_t e_t'$.

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

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

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 *et al.* (2010, 2014).

The simulation DGP is a VAR(1) process of dimension $p = 2$:

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

with $X_0 = 0$. For $a = 0$, the model is a vector random walk, and this is used to study the size of tests for no cointegration, i.e., $\mathcal{H}(0)$. We also consider the case $a = -30$ to study the power of the test for $\mathcal{H}(0)$ against local alternatives $\mathcal{H}_n(0, 1)$.

The errors are defined as $\varepsilon_t = \sigma_t z_t$, with $z_t \sim$ i.i.d. $N(0, I_2)$. 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 ε_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

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

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 homoskedasticity, and Cases 2 and 3 involve a deterministically changing Σ_t, with a late positive shift in the variances and covariances ($\Sigma_t^{(2)}$), or in the variances only ($\Sigma_t^{(3)}$). Note that Cases 2 and 3 do not satisfy

Assumption 2, because the volatility paths are not continuous. As indicated before, this is not a fundamental problem, because the indicator function $\mathbf{1}_{[s,1]}(u)$ in (22) could be replaced by, e.g., the cumulative distribution function of the $N(s-3\epsilon, \epsilon^2)$ distribution with a very small ϵ. Case 4 corresponds to continuous variation in the variances and covariances, driven by a single realisation of a log-Ornstein-Uhlenbeck stochastic volatility process.

We analyse 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. The window width for the volatility matrix estimation is selected using the leave-one-out cross-validation method. Because this method is rather time-consuming in a Monte Carlo study, we initially simulate 200 paths of the DGP (both under the null and under the alternative hypothesis), calculate the cross-validation window widths, and then use the average of these in all replications. The tables report the empirical size (or actual rejection frequency under the null, $a = 0$) and size-corrected power ($a = -30$) 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 homoskedasticity), respectively, whereas ALR-VBS and ALR-WBS indicate the volatility and wild bootstrap based adaptive LR tests.

Table 1: Size and power, 5% level

Case	n	PLR-VBS	PLR-WBS	ALR-VBS	ALR-WBS
			size		
1	500	0.055	0.053	0.065	0.063
	1000	0.049	0.049	0.053	0.054
2	500	0.043	0.039	0.065	0.080
	1000	0.053	0.044	0.061	0.068
3	500	0.063	0.042	0.051	0.071
	1000	0.062	0.047	0.061	0.065
4	500	0.061	0.052	0.074	0.073
	1000	0.047	0.048	0.058	0.072
			power		
1	500	0.829	0.810	0.826	0.821
	1000	0.846	0.842	0.844	0.844
2	500	0.701	0.697	0.806	0.795
	1000	0.701	0.708	0.825	0.826
3	500	0.618	0.598	0.748	0.736
	1000	0.595	0.603	0.736	0.731
4	500	0.609	0.606	0.795	0.780
	1000	0.637	0.631	0.819	0.818

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 and ALR are the pseudo-LR and adaptive LR tests, and VBS and WBS refer to the volatility and wild bootstrap based tests, respectively.

Table 1 gives the size and size-corrected power of the tests for each of the 4 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 $-30/n$); 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. 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 three tests. On the other hand, the power of the two adaptive tests decreases slightly as the break happens towards 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.

Table 2: Size and power, 5% level, Case 2, varying break times s

s	n	PLR-VBS	PLR-WBS	ALR-VBS	ALR-WBS
			size		
0.2	500	0.064	0.050	0.042	0.064
	1000	0.063	0.051	0.056	0.054
0.5	500	0.068	0.050	0.064	0.068
	1000	0.068	0.048	0.063	0.058
0.8	500	0.043	0.039	0.065	0.080
	1000	0.053	0.044	0.061	0.068
			power		
0.2	500	0.867	0.871	0.847	0.840
	1000	0.880	0.876	0.864	0.864
0.5	500	0.782	0.778	0.826	0.813
	1000	0.801	0.799	0.846	0.845
0.8	500	0.701	0.697	0.806	0.795
	1000	0.701	0.708	0.825	0.826

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 and ALR are the pseudo-LR and adaptive LR tests, and VBS and WBS refer to the volatility and wild bootstrap based tests, respectively.

Finally, we study the sensitivity of the simulation results for different volatility-of-volatility parameter values ζ for the log-volatility $H(u)$ in (23), Case 4; see Table 3. We first note that the size distortion of the wild bootstrap adaptive test becomes more severe as the volatility-of-volatility parameter gets higher; the volatility bootstrap performs more stably, with more modest size distortions. The power of the PLR test is very sensitive to the degree of variation in the volatility: in the high ζ scenario, the power 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 κ.

Table 3: Size and power, 5% level, Case 4, varying volatility of volatility ζ

ζ	n	PLR-VBS	PLR-WBS	ALR-VBS	ALR-WBS
			size		
0.5	500	0.068	0.052	0.075	0.063
	1000	0.059	0.048	0.062	0.059
1.0	500	0.061	0.052	0.074	0.073
	1000	0.047	0.048	0.058	0.072
2.0	500	0.045	0.052	0.062	0.102
	1000	0.044	0.044	0.056	0.108
			power		
0.5	500	0.748	0.723	0.792	0.785
	1000	0.783	0.763	0.827	0.823
1.0	500	0.609	0.606	0.795	0.780
	1000	0.637	0.631	0.819	0.818
2.0	500	0.344	0.331	0.794	0.782
	1000	0.347	0.337	0.806	0.809

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). ζ refers to the volatility of volatility in Case 4, and n is the sample size. PLR and ALR are the pseudo-LR and adaptive LR tests, and VBS and WBS refer to the volatility and wild bootstrap based tests, respectively.

In all cases, we find that the size and power properties of the tests does not seem to be affected too much by varying the sample size, and all tests are considered to perform reasonably well in a sample size typical for macro-economic applications, examples of which are considered in the next section.

In summary, the Monte Carlo simulation results in this section indicate that the volatility bootstrap, based on the non-parametric 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 Revisiting US stock market indices in the 1990s

We illustrate the use of the proposed test in an application to a cointegration analysis of the S&P 500 index and the NASDAQ-100 index in the 1990s. A cointegration model with stochastic volatility (essentially a constant conditional correlation model) was proposed by Duan and Pliska (2004), for the purpose of pricing multi-asset (basket) options. They found evidence of cointegration between the logarithms of these two stock market indices in the period from January 2, 1991 to May 15, 1998, based on the Engle and Granger (1987) residual-based cointegration test. We will reanalyse this relationship over the same sample period (with $n = 1864$) using the Gaussian (constant volatility) likelihood ratio test (the PLR test) and the adaptive test proposed in this paper. Because the previous section has indicated that the volatility bootstrap based adaptive test has better size properties than the wild bootstrap version, we focus on this implementation of the test. More recent data, covering the stock-market run up and decline around the turn of the millennium and the subsequent financial crisis, reveals that a constant linear cointegrating relation between these stock market indices will break down eventually, in line with statistical arbitrage approaches which tend to find evidence of temporary and time-varying cointegrating relations. Therefore, we will focus on the 1990s to illustrate the different outcomes resulting from the treatment of the volatility process.

Figure 1 displays the original data (not in logs). It is seen that both time series display an upward trend over this period, which may be common but with a different slope: the average growth rate of the NASDAQ clearly exceeds that of the S&P 500. To allow for the possibility that this different slope is partly caused by a different deterministic linear trend in the logs, we follow Duan and Pliska (2004) in allowing for a restricted linear trend in the cointegrating relationship (i.e., restricted to exclude the possibility of a quadratic trend in the levels).

Letting $X_t = (X_{1t}, X_{2t})' = (\log(\text{S\&P } 500)_t, \log(\text{NASDAQ-100})_t)'$, a first-order vector autoregressive model with linear trend appears to be dynamically well-specified: residuals do not display significant serial correlation, and the lag length of one is selected by the usual information criteria. However, residual analysis reveals that a Gaussian i.i.d. assumption on the errors is likely to be misspecified, since the residuals display heteroskedasticity and leptokurtosis. Ignoring these results, and using the Johansen trace test as a pseudo-likelihood ratio test for $\mathcal{H}(0)$ leads to a test statistic of 24.37, with an asymptotic p-value of 0.075 (based on the approximation method of Doornik (1998)). However, the wild bootstrap p-value, which as shown by Cavaliere *et al.* (2014) corrects for both conditional and unconditional heteroskedasticity, is 0.14, indicating only very weak evidence for cointegration. The estimated cointegrating relation is (heteroskedasticity-consistent standard errors in parentheses)

$$\widehat{\beta}^{\#\prime} X_t^{\#} = X_{1t} - \underset{(0.079)}{0.922} X_{2t} + \underset{(0.677)}{1.976} \times 10^{-4} t,$$

which includes a linear trend coefficient that is significantly different from zero at the usual significance level. The estimated error correction coefficients are $\widehat{\alpha}_1 = -0.001$ (0.004) and

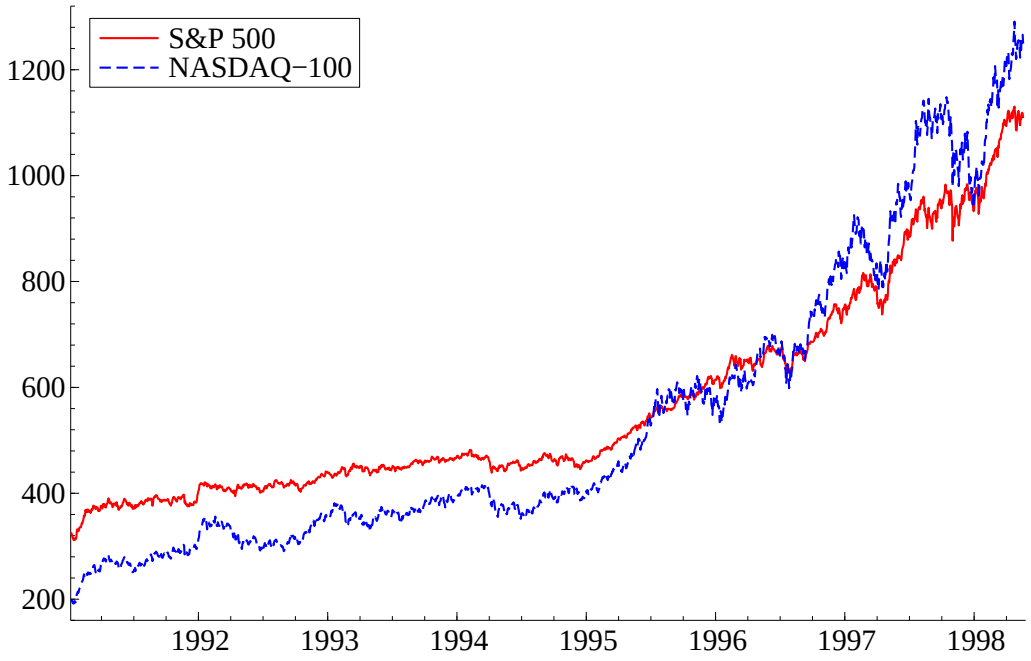

Figure 1: S&P 500 and NASDAQ-100 index, 02/01/1991 – 15/05/1998.

$\widehat{\alpha}_2 = 0.020$ (0.007), suggesting that most of the error correction is done by the NASDAQ.

Figure 2 shows the estimated volatilities and the covariance and correlation based on a Gaussian kernel, with the window width $h = 0.0026$ chosen by leave-one-out cross validation. The resulting estimates might seem undersmoothed; this can be explained by the fact that the errors in this model are likely to contain a combination of long-run (unconditional) volatility changes and short-run (GARCH-type) volatility clustering, and since the latter is not explicitly modelled, it is also picked up by the nonparametric volatility estimates. It should be emphasised, however, that the cross-validation criterion function (21) is quite flat for values of h higher than the minimiser. We will investigate the robustness of the adaptive cointegrating testing result to variations in the window width. From both estimates, we observe that the correlation may display stationary variation around a mean of about 0.75, but the volatilities and covariance appear to have a lower mean-reversion. Therefore it seems reasonable to apply the type of asymptotics implied by Assumption 2.

The adaptive likelihood ratio statistic for no cointegration based on $\{\widehat{\Sigma}_t\}_{t=1}^n$ has a value of 33.66, with a volatility bootstrap p-value of 0.001, indicating much stronger support for the cointegration hypothesis. As mentioned above, the cross-validation criterion function is quite flat, suggesting that the optimal value of h is hard to determine, and may lead to undersmoothing. We have also carried out the test with $h = 0.005$ and $h = 0.01$, leading to p-values of 0.015 and 0.043, respectively, which suggest that the evidence in favour of cointegration is reasonably robust. Increasing the window width further will eventually lead to a flat volatility pattern and hence the same p-value as for the constant-volatility pseudo-likelihood ratio test.

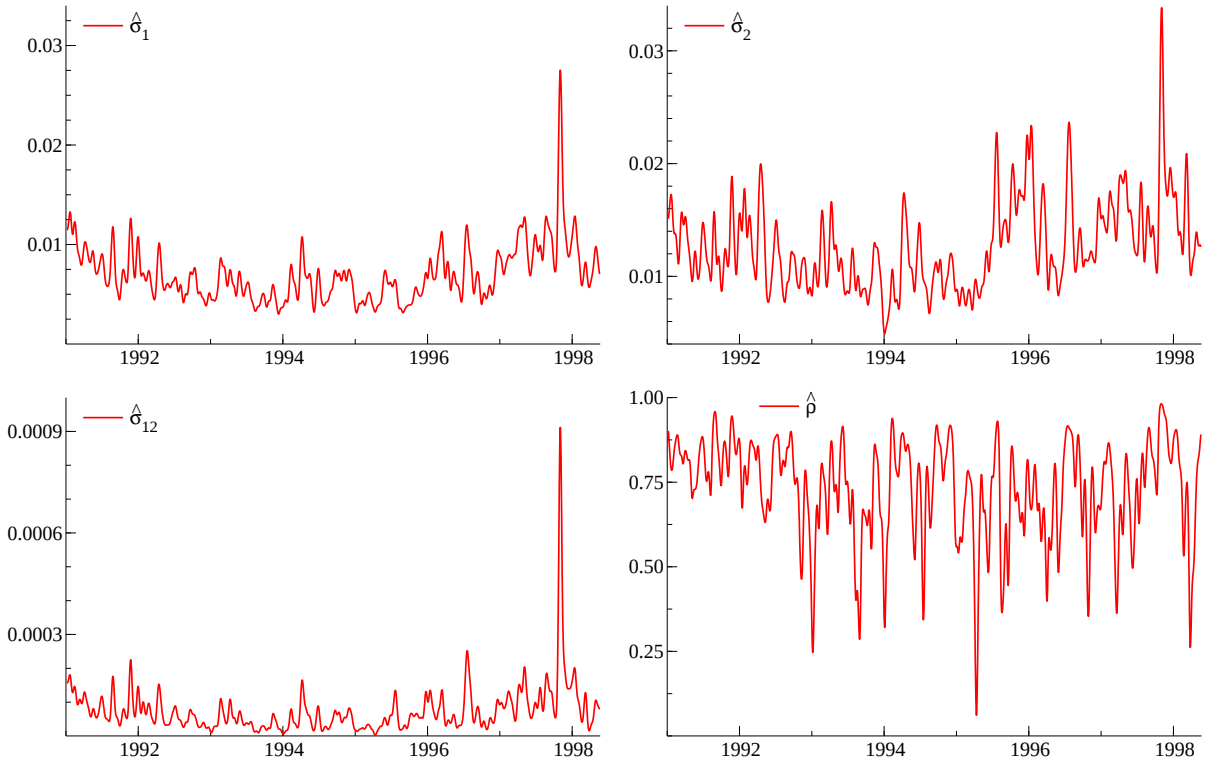

Figure 2: Estimated volatilities, covariance and correlation of e_{1t} and e_{2t}.

The resulting cointegrating relation, based on $\{\widehat{\Sigma}_t\}_{t=1}^n$ now becomes

$$\widehat{\beta}^{\#\prime} X_t^{\#} = X_{1t} - \underset{(0.083)}{0.956} X_{2t} + \underset{(0.656)}{2.227} \times 10^{-4} t,$$

which is quite similar to the relation found from the Gaussian constant-volatility pseudo-likelihood (as well as the estimates based on $\{\widehat{\Sigma}_t\}$). Similarly, the adjustment coefficients $\widehat{\alpha}_1 = -0.00007$ (0.003) and $\widehat{\alpha}_2 = 0.019$ (0.006) are hardly affected.

In summary, this empirical example illustrates that empirically relevant volatility patterns may lead, on the one hand, to size distortions of conventional cointegration tests, and on the other hand, to more efficient estimators and more powerful tests derived from an appropriate likelihood function that allows for time-varying volatilities.

6.2 Revisiting the US term structure of interest rates, 1970–2009

Boswijk *et al.* (2016) provide a cointegration analysis of the term structure of interest rates in the US, showing that allowing for heteroskedasticity 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 i.i.d. pseudo-likelihood. We will investigate to what extent their empirical results change if we use adaptive tests as developed in this paper.

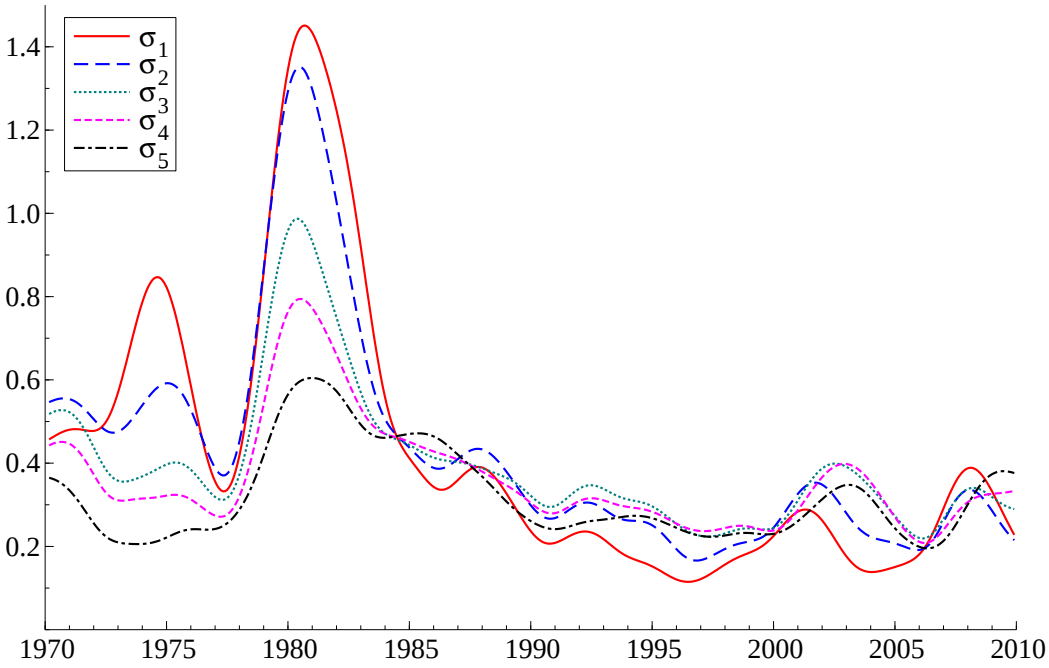

Figure 3: Estimated volatilities based on residuals of an unrestricted VAR(2).

We analyse the same time series $X_t = (X_{1t}, \ldots, X_{5t})'$ of monthly zero 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).

Before we present the cointegration test results, Figure 3 displays the nonparametric estimates of the time-varying volatilities σ_{it} of X_{it}, $i = 1, \ldots, 5$. 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).

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

Table 4: Volatility and wild bootstrap p-values of ALR and PLR tests for cointegration rank.

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 US term structure data.

the conventional 5% significance level. This implies a single stochastic trend driving the five different yields. The adaptive ML estimators of α and β for $r = 4$ (with β normalised on $X_{2t}, \ldots, X_{5t}$) together with QMLE standard errors, are as follows:

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

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

We may be interested in the hypothesis that the first row of β 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 β has not been developed in this paper, but following the analysis of Boswijk *et al.* (2016), we expect the bootstrap to yield asymptotically valid inference again.

7 Discussion

In this paper 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 examples have illustrated that applying our newly developed tests can indeed lead to stronger evidence for cointegration than alternative tests. These applications relate to the equity market and the fixed income market, but the method could also be applied to exchange rate models (spot-forward relations, purchasing power parity). In principle the approach could also be used with macro-data, but the non-parametric kernel estimator cannot be expected to give very accurate estimates when data are observed infrequently, and in such cases it may be advisable to adopt a parametric model for the changing volatilities and correlations.

The theory and methods used in this paper can be extended in various directions. First, we have excluded discontinuities in the time variation of volatilities and correlations, but this has been mainly to simplify the analysis. As discussed by, e.g., Xu and Phillips (2008), adaptive estimation based on nonparametric volatility estimation is still possible in the presence of a finite number of jumps. This is supported by our Monte Carlo simulations, although we do not give an explicit theoretical analysis of this case.

The paper has focussed on time-variation in the unconditional variance matrix. The analysis could be extended to also allow for conditional heteroskedasticity, 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.

Finally, we have considered here tests based on a Gaussian likelihood function. In practice one often observes that standardised financial returns still display excess kurtosis. The asymptotic results in this paper are robust to this type of nonnormality, but in such cases more powerful tests could be derived from, e.g., a Student's t likelihood.

Appendix

Proof of Lemma 1. The starting point is that under Assumption 2, by the multivariate invariance principle for martingale difference sequences (mds),

$$W_n(u) := n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} z_t \xrightarrow{w} W(u), \tag{A.1}$$

with W a p-variate standard Brownian motion. Next, defining the discretised version $\sigma_n(u) = \sum_{t=1}^{n} \sigma(t/n) \mathbf{1}_{[(t-1)/n, t/n)}(u)$ of $\sigma(u)$, we have

$$n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t = n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} \sigma_t z_t = \int_0^u \sigma_n(s) dW_n(s).$$

Under the assumptions on $\sigma(\cdot)$, it follows that as $n \to \infty$, $\sigma_n(\cdot) \to \sigma(\cdot)$ in $\mathcal{D}[0,1]^{p\times p}$, jointly with (A.1). Using a multivariate version of Hansen (1992)'s Theorem 2.1, it follows that $\int_0^u \sigma_n(s)dW_n(s) \xrightarrow{w} \int_0^u \sigma(s)dW(s)$ in $\mathcal{D}[0,1]^p$, jointly with (A.1), because $\sigma_n(\cdot)$ is non-stochastic and $\{z_t\}_{t\geq 1}$ is an mds with constant (conditional) variance and finite fourth moment.

The proof of (4) is based on the moving average representation implied by (1) under $\mathcal{H}_n(r, r_1)$. For the homoskedastic case, this has been analysed in detail in Theorem 14.1 and Exercise 14.1 of Johansen (1996), and the corresponding solution to Exercise 14.1 given in Hansen and Johansen (1998). Consider the model in companion form, for the stacked vector $\mathbb{X}_t = (X_t', \ldots, X_{t-k+1}')'$:

$$\mathbb{X}_t = \left(I_{kp} + \mathbb{A}\mathbb{B}' + n^{-1}\mathbb{A}_1\mathbb{B}_1'\right)\mathbb{X}_{t-1} + F\varepsilon_t, \tag{A.2}$$

where

$$\mathbb{A} = \begin{bmatrix} \alpha & \Gamma_1 & \cdots & \Gamma_{k-1} \\ 0 & I_p & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_p \end{bmatrix}, \qquad \mathbb{B} = \begin{bmatrix} \beta & I_p & 0 & 0 \\ 0 & -I_p & \ddots & 0 \\ \vdots & \ddots & \ddots & I_p \\ 0 & \cdots & 0 & -I_p \end{bmatrix},$$

and where $\mathbb{A}_1 = (\alpha_1', 0_{(k-1)p}')'$, $\mathbb{B}_1 = (\beta_1', 0_{(k-1)p}')'$, and $F = (I_p, 0_{p\times(k-1)p})'$. Note that $\mathbb{A}_\perp' = \alpha_\perp'[I_p : -\Gamma_1 : \cdots : -\Gamma_{k-1}]$ and $\mathbb{B}_\perp' = \beta_\perp'[I_p : I_p : \cdots : I_p]$, and hence

$$\mathbb{A}_\perp'\mathbb{A}_1 = \alpha_\perp'\alpha_1, \qquad \mathbb{B}_1'\mathbb{B}_\perp = \beta_1'\beta_\perp, \qquad \mathbb{A}_\perp'\mathbb{B}_\perp = \alpha_\perp'\Gamma\beta_\perp.$$

Let $\Phi_n = I_{kp} + \mathbb{A}\mathbb{B}' + n^{-1}\mathbb{A}_1\mathbb{B}_1'$. Backward substitution in (A.2) gives the solution

$$\mathbb{X}_t = \Phi_n^t\mathbb{X}_0 + \sum_{j=1}^t \Phi_n^{t-j}F\varepsilon_j.$$

It will be convenient to work with the decomposition $\mathbb{X}_t = \mathbb{A}(\mathbb{B}'\mathbb{A})^{-1}\mathbb{B}'\mathbb{X}_t + \mathbb{B}_\perp(\mathbb{A}_\perp'\mathbb{B}_\perp)^{-1}\mathbb{A}_\perp'\mathbb{X}_t$; we consider the behaviour of $\mathbb{B}'\mathbb{X}_t$ (the stable linear combinations) and $\mathbb{A}_\perp'\mathbb{X}_t$ (the non-stationary linear combinations) separately. Assumption 1 requires that all eigenvalues of the matrix $I_{r+(k-1)p} + \mathbb{B}'\mathbb{A}$ are less than one in absolute value, and this implies for the stable linear combinations, by Theorem 14.1 of Johansen (1996),

$$\mathbb{B}'\mathbb{X}_t = (I_{r+(k-1)p} + \mathbb{B}'\mathbb{A})^t\mathbb{B}'\mathbb{X}_0 + \sum_{j=1}^t (I_{r+(k-1)p} + \mathbb{B}'\mathbb{A})^{t-j}\mathbb{B}'F\varepsilon_j + R_{\vec{\beta}t},$$

with $R_{\vec{\beta}t} = \mathrm{o}_p(1)$, such that $n^{-1/2}\mathbb{B}'\mathbb{X}_{\lfloor sn\rfloor} \xrightarrow{p} 0$. For the nonstationary linear combinations, we find

$$\mathbb{A}_\perp'\mathbb{X}_t = \mathbb{A}_\perp'\Phi_n^t\mathbb{X}_0 + \sum_{j=1}^t \mathbb{A}_\perp'\Phi_n^{t-j}F\varepsilon_j.$$

Defining $\Phi_n(u) = \Phi_n^{\lfloor un\rfloor}$, we therefore find

$$n^{-1/2}\mathbb{A}_\perp'\mathbb{X}_{\lfloor un\rfloor} = n^{-1/2}\mathbb{A}_\perp'\Phi_n(u)\mathbb{X}_0 + \mathbb{A}_\perp'\Phi_n(u)\int_0^u \Phi_n(-s)F\sigma_n(s)dW_n(s).$$

Theorem A.14 of Johansen (1996) implies $\Phi_n(u) \to \mathbb{B}_\perp(\mathbb{A}_\perp'\mathbb{B}_\perp)^{-1}\exp(uA)\mathbb{A}_\perp'$, and hence

$$\Phi_n(-u)F\sigma_n(u) \longrightarrow \mathbb{B}_\perp(\mathbb{A}_\perp'\mathbb{B}_\perp)^{-1}\exp(uA)\sigma_U(u), \qquad u \in [0,1],$$

jointly with $W_n(\cdot) \xrightarrow{w} W(\cdot)$. Because z_t is an mds with constant variance, Hansen (1992)'s Theorem 2.1 again implies

$$n^{-1/2}\mathbb{A}_\perp'\mathbb{X}_{\lfloor un\rfloor} \xrightarrow{w} \int_0^u \exp\left((u-s)A\right)\sigma_U(s)dW(s) = U_A(u), \qquad u \in [0,1].$$

The stochastic differential equation (5) follows from Itô's formula, writing $U_A(u) = \exp(uA)U_A^*(u)$ $= f(u, U_A^*(u))$, with $dU_A^*(u) = \exp(-uA)\sigma_U(u)dW(u)$. Finally,

$$\begin{aligned} n^{-1/2}X_{\lfloor un\rfloor} &= n^{-1/2}F'\mathbb{B}_\perp(\mathbb{A}_\perp'\mathbb{B}_\perp)^{-1}\mathbb{A}_\perp'\mathbb{X}_{\lfloor un\rfloor} + \mathrm{o}_p(1) \\ &= \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}n^{-1/2}\mathbb{A}_\perp'\mathbb{X}_{\lfloor un\rfloor} \\ &\xrightarrow{w} \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}U_A(u), \qquad u \in [0,1], \end{aligned}$$

which concludes the proof of (4). □

Proof of Theorem 1. Let $\theta = \mathrm{vec}[\Pi : \Psi]$, and define the residual function $\varepsilon_t(\theta) = \Delta X_t - [\Pi : \Psi]Z_t = \Delta X_t - [Z_t' \otimes I_p]\theta$. It will be convenient to start analysing (twice) the log-likelihood ratio function relative to the unrestricted estimators, i.e.,

$$\Lambda_n(\theta) = -2\left[\ell_n(\Pi, I_p, \Psi) - \ell_n(\widehat{\Pi}_n, I_p, \widehat{\Psi}_n)\right] = \sum_{t=1}^n \left(\varepsilon_t(\theta)'\Sigma_t^{-1}\varepsilon_t(\theta) - \widehat{\varepsilon}_t'\Sigma_t^{-1}\widehat{\varepsilon}_t\right).$$

Using $\varepsilon_t(\theta) = \widehat{\varepsilon}_t + [Z_t' \otimes I_p](\widehat{\theta}_n - \theta)$, and $\sum_{t=1}^n [Z_t \otimes \Sigma_t^{-1}]\widehat{\varepsilon}_t = 0$, we find

$$\Lambda_n(\theta) = (\widehat{\theta}_n - \theta)' \sum_{t=1}^n [Z_t Z_t' \otimes \Sigma_t^{-1}](\widehat{\theta}_n - \theta). \tag{A.3}$$

Note that $\mathrm{LR}_n = \min_{\theta\in\Theta_r}\Lambda_n(\theta)$, where Θ_r is the restricted parameter space

$$\Theta_r = \{\theta \in \mathbb{R}^{kp^2} : \Pi = \alpha\beta'; \quad (\alpha, \beta, \Psi) \in \mathbb{R}^{p\times r} \times \mathbb{R}^{p\times r} \times \mathbb{R}^{p\times(k-1)p}\}.$$

Let θ_0 denote the true value under $\mathcal{H}(r)$, and let $\theta_n = \theta_0 + D_n\tau$ denote a sequence of parameter values, where $\tau \in \mathbb{R}^{p(kp+1)}$ is a fixed vector and D_n a sequence of non-singular norming matrices, chosen such that the corresponding probability measures $P_{\theta_0}^n$ and $P_{\theta_n}^n$ are contiguous. In the present situation, this requires that $D_n' \sum_{t=1}^n [Z_t Z_t' \otimes \Sigma_t^{-1}] D_n$ and $D_n^{-1}(\widehat{\theta}_n - \theta_0)$ converge in distribution, and a choice of D_n that satisfies this requirement is

$$D_n = \begin{bmatrix} n^{-1}\Gamma'\alpha_\perp & n^{-1/2}\beta & 0 \\ 0 & 0 & n^{-1/2}I_{k(p-1)} \end{bmatrix} \otimes I_p, \tag{A.4}$$

such that

$$D_n'[Z_t \otimes I_p] = n^{-1/2}\begin{pmatrix} n^{-1/2}\alpha_\perp'\Gamma X_{t-1} \\ Z_t(\beta) \end{pmatrix} \otimes I_p.$$

Here $\alpha_\perp$, β and Γ correspond to the true value θ_0.

The local alternative $\mathcal{H}(r, r_1)$ corresponds to a particular choice of the non-centrality parameter τ, which is seen as follows. First, the condition $\text{rank}(\alpha'_\perp \Gamma \beta_\perp) = p - r$ (Assumption 1 (c)) implies that the matrix $[\beta : \Gamma' \alpha_\perp]$ has full rank. A projection of β_1 on $\text{sp}(\beta : \Gamma' \alpha_\perp)$ is given by $\beta_1 = \beta a + \Gamma' \alpha_\perp b$, where $[a' : b']' = [\beta : \Gamma' \alpha_\perp]^{-1} \beta_1$. Because $\Pi_n = \alpha \beta' + n^{-1} \alpha_1 \beta_1' = \alpha_n^\dagger \beta' + n^{-1} \alpha_1 \beta_1^{\dagger\prime}$, with $\alpha_n^\dagger = \alpha + n^{-1} \alpha_1 a'$ and $\beta_1^\dagger = \Gamma' \alpha_\perp b$, the part of β_1 that lies in $\text{sp}(\beta)$ may be absorbed in $\alpha \beta'$. Note also that for the asymptotic analysis, only $\beta'_\perp \beta_1 = \beta'_\perp \Gamma' \alpha_\perp b$ is relevant, cf. Lemma 1. Therefore we may set $a = 0$ and hence confine ourselves to local alternatives with $\beta_1 = \Gamma' \alpha_\perp b$, with $b = (\beta'_\perp \Gamma' \alpha_\perp)^{-1} \beta'_\perp \beta_1$. This implies

$$\theta_n = \text{vec}\,[\Pi_n : \Psi_0] = \text{vec}\,[\Pi_0 : \Psi_0] + \text{vec}\left[n^{-1} \alpha_1 \beta_1' : 0\right] = \theta_0 + D_n \check{\tau}, \tag{A.5}$$

with $\check{\tau} = (\text{vec}(\alpha_1 b')', 0'_{pr+(k-1)p})'$.

Lemma 1 implies that both under the null and under local alternatives, $n^{-1/2} \alpha'_\perp \Gamma X_{\lfloor un \rfloor} \xrightarrow{w} U_A(u)$, $u \in [0, 1]$. Under the null hypothesis, Granger's representation theorem implies that $Z_t(\beta)$ is a mean-zero linear process $\sum_{j=1}^\infty C_j \varepsilon_{t-j}$, with exponentially decaying weight matrices C_j. From Chapter 14 of Johansen (1996) and Hansen and Johansen (1998), we know that under local alternatives, $Z_t(\beta)$ may be decomposed into the same linear process and an additional term, which is asymptotically negligible. Using a multivariate generalization of the asymptotic theory for stationary linear processes with non-stationary volatility, cf. Hansen (1995) and Phillips and Xu (2006), it follows that, under both the null and local alternatives,

$$\begin{aligned} D_n' \sum_{t=1}^n [Z_t Z_t' \otimes \Sigma_t^{-1}] D_n \quad &\xrightarrow{w} \quad \begin{bmatrix} \int_0^1 \left[U_A(u) U_A(u)' \otimes \Sigma(u)^{-1}\right] du & 0 \\ 0 & \int_0^1 \left[\Omega(u) \otimes \Sigma(u)^{-1}\right] du \end{bmatrix} \\ &=: \quad \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} = J, \end{aligned} \tag{A.6}$$

where $\Omega(u) = \sum_{j=1}^\infty C_j \Sigma(u) C_j$.

For the unrestricted estimator $\widehat{\theta}_n$, we find

$$D_n^{-1}(\widehat{\theta}_n - \theta_0) = \left(D_n' \sum_{t=1}^n [Z_t Z_t' \otimes \Sigma_t^{-1}] D_n \right)^{-1} D_n' \sum_{t=1}^n [Z_t \otimes \Sigma_t^{-1}] \varepsilon_t(\theta_0), \tag{A.7}$$

where under $\mathcal{H}(r)$, $\varepsilon_t(\theta_0) = \varepsilon_t$, whereas under the local alternative $\mathcal{H}(r, r_1)$, $\varepsilon_t(\theta_0) = \varepsilon_t + [Z_t' \otimes I_p] D_n \check{\tau}$. Under both hypotheses, we find, again generalising the results of Hansen (1995) and Phillips and Xu (2006),

$$\begin{aligned} D_n' \sum_{t=1}^n [Z_t \otimes \Sigma_t^{-1}] \varepsilon_t(\theta_0) \quad &\xrightarrow{w} \quad \begin{pmatrix} \int_0^1 \left[U_A(u) \otimes \Sigma(u)^{-1}\right] (dM(u) + [U_A(u)' du \otimes I_p] \check{\tau}_1) \\ \sum_{j=1}^\infty \int_0^1 [C_j \sigma(u) \otimes \sigma(u)'^{-1}] dB_j(u) \end{pmatrix} \\ &=: \quad \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = S, \end{aligned} \tag{A.8}$$

where $\check{\tau}_1 = \text{vec}(\alpha_1 b')$, a $p(p-r)$-vector consisting of the first rows of the value of $\check{\tau}$ in (A.5), and where $\{B_j(\cdot)\}_{j=1}^\infty$ are mutually independent p^2-vector Brownian motion processes, obtained

as the limit in distribution of the partial sum processes of $\text{vec}(z_t z'_{t-j})$. Because $\sigma(\cdot)$ is non-stochastic, the bottom right-hand side expression of (A.8) has a normal distribution with mean zero and variance $\int_0^1 \left[\Omega(u) \otimes \Sigma(u)^{-1}\right] du$. Note that under the null, $\check{\tau}_1 = 0$ and $U_A(u) = U_0(u) = \alpha'_\perp M(u)$ in (A.6) and (A.8).

The representation (A.7), together with the limit results (A.6) and (A.8), implies that $D_n^{-1}(\widehat{\theta}_n - \theta_0) = \text{O}_p(1)$ under both the null and local alternatives, such that $\widehat{\theta}_n$ is consistent under both hypotheses. Generalising the argument explained fully in Theorem A1 of Johansen (1997), this implies that the restricted MLE $\widetilde{\theta}_n$ is also consistent, and $D_n^{-1}(\widetilde{\theta}_n - \theta_0) = \text{O}_p(1)$. This implies that for the derivation of the limiting distribution of the LR statistic $\text{LR}_n = \min_{\theta \in \Theta_r} \Lambda_n(\theta)$, we may confine ourselves to the behaviour of $\Lambda_n(\theta)$ for sequences $\theta_n = \theta_0 + D_n \tau$. In particular, let $\mathcal{T}_{r,n} = \{\tau \in \mathbb{R}^{kp^2} : \theta_0 + D_n \tau \in \Theta_r\}$, such that $\text{LR}_n = \min_{\tau \in \mathcal{T}_{r,n}} \Lambda_n(\theta_0 + D_n \tau)$. We will show that $\Lambda_n(\theta_0 + D_n \tau) \xrightarrow{w} \Lambda(\tau)$ uniformly on compact sets, and that the restricted parameter space $\mathcal{T}_{r,n}$ converges to a limit $\mathcal{T}_r$. Because $\widetilde{\tau}_n = D_n^{-1}(\widetilde{\theta}_n - \theta_0)$ is $\text{O}_p(1)$, this will then imply, by the argmax theorem (Van der Vaart, 1998, Corollary 5.58), $D_n^{-1}(\widetilde{\theta}_n - \theta_0) \xrightarrow{w} \arg\min_{\tau \in \mathcal{T}_r} \Lambda(\tau)$ and $\text{LR}_n \xrightarrow{w} \min_{\tau \in \mathcal{T}_r} \Lambda(\tau)$.

For the limit of the log-likelihood ratio, we find

$$\Lambda_n(\theta_0 + D_n \tau) = \left(D_n^{-1}(\widehat{\theta}_n - \theta_0) - \tau\right)' D'_n \sum_{t=1}^{n} [Z_t Z'_t \otimes \Sigma_t^{-1}] D_n \left(D_n^{-1}(\widehat{\theta}_n - \theta_0) - \tau\right),$$

which by (A.6)–(A.8) converges in distribution to $\Lambda(\tau) = (S - J\tau)' J^{-1} (S - J\tau)$. Because both $\Lambda_n(\cdot)$ and $\Lambda(\cdot)$ are quadratic, this convergence is uniform on compact sets.

For the restricted parameter space, we use the fact that $\Pi = \alpha\beta' = \alpha(\bar{c}' + \Phi' c'_\perp)$, such that

$$\begin{aligned}
\text{vec}(\Pi_n - \Pi_0) &= \text{vec}\left(\alpha_n[\bar{c}' + \Phi'_n c'_\perp] - \alpha[\bar{c}' + \Phi' c'_\perp]\right) \\
&= \text{vec}\left([\alpha_n - \alpha]\beta' + \alpha[\Phi_n - \Phi]' c'_\perp + [\alpha_n - \alpha][\Phi_n - \Phi]' c'_\perp\right) \\
&= [\beta \otimes I_p] \,\text{vec}(\alpha_n - \alpha) + [c_\perp \otimes I_p][I_{p-r} \otimes \alpha] \,\text{vec}(\Phi'_n - \Phi') + R_n,
\end{aligned}$$

where the remainder term R_n is $\text{O}(\|\alpha_n - \alpha\| \, \|\Phi_n - \Phi\|)$, and where (α, Φ, β) now denote the true values, corresponding to Π_0. Note that c may be freely chosen, as long as $c'\beta$ is non-singular, which is equivalent to the condition that $[\beta : c_\perp]$ should be of full rank. Using the fact that $[\beta : \Gamma' \alpha_\perp]$ has full rank p, we find that we may choose $c_\perp = \Gamma' \alpha_\perp$. This in turn means that, if we let $\text{vec}(\Psi_n - \Psi_0) = n^{-1}\kappa_\psi$, $\text{vec}(\alpha_n - \alpha) = n^{-1/2}\kappa_\alpha$ and $\text{vec}(\Phi'_n - \Phi') = n^{-1}\kappa_\phi$, then in the restricted parameter space Θ_r, we have

$$\theta_n = \theta_0 + D_n \begin{pmatrix} [I_q \otimes \alpha]\kappa_\phi \\ \kappa_\alpha \\ \kappa_\psi \end{pmatrix} + \begin{pmatrix} R_n \\ 0 \end{pmatrix},$$

such that

$$\tau = D_n^{-1}(\theta_n - \theta_0) = \begin{bmatrix} I_q \otimes \alpha & 0 & 0 \\ 0 & I_r \otimes I_p & 0 \\ 0 & 0 & I_{(k-1)p} \otimes I_p \end{bmatrix} \begin{pmatrix} \kappa_\phi \\ \kappa_\alpha \\ \kappa_\psi \end{pmatrix} + \text{o}(1) = G\kappa + \text{o}(1).$$

where $q = p - r$. Therefore, the limiting null space for τ is the linear subspace $\mathcal{T}_r = \{\tau = G\kappa : \kappa \in \mathbb{R}^l\}$, where $l = qr + pr + (k-1)p^2$, the dimension of the restricted parameter space.

Combining these results, we find

$$\begin{aligned} \mathrm{LR}_n \xrightarrow{w} \min_{\tau \in \mathcal{T}_r} \Lambda(\tau) &= S'J^{-1}S - S'G(G'JG)^{-1}G'S \\ &= S'J^{-1}G_\perp(G'_\perp J^{-1}G_\perp)^{-1}G'_\perp J^{-1}S, \end{aligned}$$

where

$$G_\perp = \begin{bmatrix} I_q \otimes \alpha_\perp \\ 0 \end{bmatrix}.$$

It follows that $LR_n \xrightarrow{w} S'_1 J_1^{-1} G_{\perp 1}(G'_{\perp 1} J_1^{-1} G_{\perp 1})^{-1} G'_{\perp 1} J_1^{-1} S_1$, where $G_{\perp 1} = I_q \otimes \alpha_\perp$.

Let K_{mn} denote the commutation matrix of appropriate order, such that $\mathrm{vec}(A') = K_{mn}\,\mathrm{vec}(A)$ for an $m \times n$ matrix A, see Magnus and Neudecker (1988). We will use the properties $K'_{mn} = K_{mn}^{-1} = K_{nm}$, and $K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$ for matrices A and B of dimensions $m \times n$ and $p \times q$, respectively. We find

$$G_{\perp 1} = I_q \otimes \alpha_\perp = K_{qp}(\alpha_\perp \otimes I_q)K_{qq} = K_{qp}\left([I_q : 0] \otimes I_q\right)\left([\alpha_\perp : \bar{\alpha}] \otimes I_q\right)K_{qq}.$$

Without loss of generality, $\alpha_\perp$ may be chosen such that $\alpha'_\perp \alpha_\perp = I_q$, and hence $[\alpha_\perp : \bar{\alpha}]^{-1} = [\alpha_\perp : \alpha]'$. Therefore, we find

$$\begin{aligned} G'_{\perp 1} J_1^{-1} G_{\perp 1} &= K_{qp}\left([I_q : 0] \otimes I_q\right)\left(\int_0^1 Y(u)Y(u)'du\right)^{-1}\left([I_q : 0]' \otimes I_q\right)K_{pq} \\ &= K_{qp}\left(\int_0^1 Z(u)Z(u)'du\right)^{-1}K_{pq}. \end{aligned}$$

Similarly, using

$$\begin{aligned} S_1 &= K_{qp}\int_0^1 \left[\sigma(u)'^{-1} \otimes U_A(u)\right]\left(dW(u) + [\sigma(u)^{-1} \otimes U_A(u)']du K_{pq}\check{\tau}_1\right) \\ &= K_{qp}\left([\alpha_\perp : \bar{\alpha}] \otimes I_q\right)\int_0^1 Y(u)\left(dW(u) + Y(u)'du\left([\alpha_\perp : \bar{\alpha}]' \otimes I_q\right)K_{pq}\check{\tau}_1\right) \\ &= K_{qp}\left([\alpha_\perp : \bar{\alpha}] \otimes I_q\right)\int_0^1 Y(u)\left(dW(u) + Y(u)'du\,\mathrm{vec}(A' : b\alpha'_1\bar{\alpha})\right), \end{aligned}$$

we obtain

$$\begin{aligned} G'_{\perp 1} J_1^{-1} S_1 &= K_{qp}\left([I_q : 0] \otimes I_q\right)\left(\int_0^1 Y(u)Y(u)'du\right)^{-1} \\ &\quad \times \int_0^1 Y(u)[dW(u) + Y(u)'du\,\mathrm{vec}(A' : b\alpha'_1\bar{\alpha})] \\ &= K_{qp}\left(\int_0^1 Z(u)Z(u)'du\right)^{-1}\int_0^1 Z(u)[dW(u) + Z(u)'\,\mathrm{vec}(A')du]. \end{aligned}$$

This leads to the required result. □

Proof of Lemma 2. We first show the preliminary uniform consistency result $\max_{1 \le t \le n} \|\widehat{\Sigma}_t - \Sigma_t\| \xrightarrow{p} 0$ for the nonparametric estimator $\widehat{\Sigma}_t$, and then strengthen the result to the estimated process over the compact interval $[0, 1]$.

The proof for this preliminary step generalises Theorem 2 in Hansen (1995), in three directions: (i) we use a multivariate instead of univariate volatility estimator, (ii) we consider residuals from a regression with $I(1)$ instead of $I(0)$ regressors, and (iii) a general (double-sided) kernel function is used.

First note that the kernel estimator can be written as $\widehat{\Sigma}_t = \sum_{s=1}^n w_{st} e_s e_s'$, with weights $w_{st} = K_h(t-s)/\left(\sum_{s=1}^n K_h(t-s)\right)$ summing to 1 by construction. It is convenient to decompose the difference $\widehat{\Sigma}_t - \Sigma_t$ into three terms, as follows:

$$\widehat{\Sigma}_t - \Sigma_t = R_t^a + R_t^b + R_t^c,$$

where

$$\begin{aligned} R_t^a &= \sum_{s=1}^n w_{st}(\Sigma_s - \Sigma_t), \\ R_t^b &= \sum_{s=1}^n w_{st}\sigma_s(z_s z_s' - I_p)\sigma_s' \\ R_t^c &= \sum_{s=1}^n w_{st}(e_s e_s' - \varepsilon_s \varepsilon_s'). \end{aligned}$$

The asymptotic negligibility of R_t^a relates to continuity of Σ_t, and R_t^c refers to the estimation error in the residuals. We will show that the maximum over $t = 1, \ldots, n$ of each of these terms will converge in probability to zero (in Frobenius norm), such that $\max_{1\leq t\leq n} \|\widehat{\Sigma}_t - \Sigma_t\| \xrightarrow{p} 0$ (using the triangle inequality).

For the first term, continuity of $\Sigma(\cdot)$ implies $\|R_t^a\| \to 0$ as $n \to \infty$, which follows by the argument of Xu and Phillips (2008, proof of Theorem 2). For the second term, note that $\|z_t z_t' - I_p\|$ is an mds with finite moments up to order $2m > 2$. Following the argument in Hansen (1995) this implies that $\max_{1\leq t\leq n} \sum_{s=1}^n w_{st}\|z_s z_s' - I_p\| = \mathrm{o}_p(1)$. And because $\max_{1\leq t\leq n} \|\sigma_t\| < \max_{0\leq u\leq 1} \|\sigma(u)\|$ is finite, this in turn implies

$$\max_{1\leq t\leq n} \|R_t^b\| \leq \max_{1\leq t\leq n} \|\sigma_t\|^2 \times \max_{1\leq t\leq n} \sum_{s=1}^n w_{st}\|z_s z_s' - I_p\| = \mathrm{o}_p(1).$$

For the third term, we write the unrestricted model as $\Delta X_t = BZ_t + \varepsilon_t$, where in the notation of the proof of Theorem 1, $\theta = \operatorname{vec} B$. The least-squares residuals are given by $e_t = \varepsilon_t - (\bar{B}_n - B)Z_t$, where $\bar{B}_n = B + \sum_{t=1}^n \varepsilon_t Z_t' \left(\sum_{t=1}^n Z_t Z_t'\right)^{-1}$. Expressing D_n in (A.4) as $D_n = [\bar{D}_n \otimes I_p]$, the proof of Theorem 1 implies that $(\bar{D}_n' \sum_{t=1}^n Z_t Z_t' \bar{D}_n)$, $(\bar{D}_n' \sum_{t=1}^n Z_t \varepsilon_t')$ and $\bar{D}_n^{-1}(\bar{B}_n - B)'$ are all $\mathrm{O}_p(1)$. This is useful in the following decomposition of R_t^c:

$$\begin{aligned} R_t^c &= -\left(\sum_{s=1}^n w_s \varepsilon_s Z_s' \bar{D}_n\right) \bar{D}_n^{-1}(\bar{B}_n - B)' - (\bar{B}_n - B)\bar{D}_n'^{-1}\left(\sum_{s=1}^n w_s \bar{D}_n' Z_s \varepsilon_s'\right) \\ &\quad + (\bar{B}_n - B)\bar{D}_n'^{-1}\left(\sum_{s=1}^n w_s \bar{D}_n' Z_s Z_s' \bar{D}_n\right) \bar{D}_n^{-1}(\bar{B}_n - B)'. \end{aligned}$$

Because of the convergence of the sample moments, $\sum_{s=1}^n w_s \bar{D}_n' Z_s \varepsilon_s' = \mathrm{o}_p(1)$ and $\sum_{s=1}^n w_s \bar{D}_n' Z_s Z_s' \bar{D}_n = \mathrm{O}_p(n^{-1}) = \mathrm{o}_p(1)$. This in turn implies that $\max_{1\leq t\leq n} \|R_t^c\| = \mathrm{o}_p(1)$.

The steps above together imply

$$\max_{1\leq t\leq n} \|\widehat{\Sigma}_t - \Sigma_t\| \xrightarrow{p} 0. \tag{A.9}$$

We now strengthen the results to the estimated process $\widehat{\Sigma}_n(u)$, $u \in [0,1]$. Define the discretised version $\Sigma_n(\cdot)$ of $\Sigma(\cdot)$ by $\Sigma_n(u) = \sum_{t=1}^n \Sigma(t/n)\mathbf{1}_{[(t-1)/n,t/n)}(u)$ for $u \in [0,1)$ and $\Sigma_n(1) = \Sigma(1)$. Because a continuous function on a bounded interval is uniformly continuous, the continuity assumption implies

$$\sup_{u\in[0,1]} \|\Sigma_n(u) - \Sigma(u)\| = \max_{1\leq t\leq n} \sup_{u\in[(t-1)/n,t/n)} \|\Sigma(t/n) - \Sigma(u)\| \to 0$$

as $n \to \infty$. The definition of $\Sigma_n(u)$ and $\widehat{\Sigma}_n(u)$ implies that $\|\widehat{\Sigma}_n(u) - \Sigma_n(u)\| = \|\widehat{\Sigma}_t - \Sigma_t\|$ for $u \in [(t-1)/n, t/n)$. Therefore,

$$\begin{aligned}
\sup_{u\in[0,1]} \|\widehat{\Sigma}_n(u) - \Sigma(u)\| &\leq \sup_{u\in[0,1]} \|\widehat{\Sigma}_n(u) - \Sigma_n(u)\| + \sup_{u\in[0,1]} \|\Sigma_n(u) - \Sigma(u)\| \\
&= \max_{1\leq t\leq n} \|\widehat{\Sigma}_t - \Sigma_t\| + \sup_{u\in[0,1]} \|\Sigma_n(u) - \Sigma(u)\| \\
&\xrightarrow{p} 0,
\end{aligned}$$

which completes the proof. □

Proof of Theorem 2. We will show that the convergence results (A.6) and (A.8) still apply when Σ_t is replaced by $\widehat{\Sigma}_t$. Combining these results in the unrestricted log-likelihood ratio (A.3), using (A.7), will imply that $\widehat{\Lambda}_n(\theta_0 + D_n\tau)$ converges to $\Lambda(\tau) = (S_1 - J_1\tau_1)'J_1^{-1}(S_1 - J_1\tau_1) + \Lambda_2(\tau_2)$, uniformly on compact sets. The remainder of the proof is analogous to the proof of Theorem 1.

Define

$$V_{t-1} := \alpha_\perp'\Gamma X_{t-1} \otimes \sigma_t'^{-1},$$

and similarly $\widehat{V}_{t-1}$ (obtained by replacing σ_t by a matrix square root of $\widehat{\Sigma}_t$). Lemma 1 implies

$$n^{-1/2}V_{\lfloor un\rfloor} \xrightarrow{w} V(u) := U_A(u) \otimes \sigma(u)'^{-1}, \quad u \in [0,1],$$

under both the null hypothesis and local alternatives. The continuous mapping theorem, together with Lemma 2, then implies $n^{-1/2}\widehat{V}_{\lfloor un\rfloor} \xrightarrow{w} V(u), u \in [0,1]$. Partitioning $D_n = [D_{1n} : D_{2n}]$ conformably with S and J, the continuous mapping theorem further implies that

$$D_{1n}'\sum_{t=1}^n [Z_tZ_t' \otimes \widehat{\Sigma}_t^{-1}]D_{1n} = \frac{1}{n^2}\sum_{t=1}^n \widehat{V}_{t-1}\widehat{V}_{t-1}' \xrightarrow{w} \int_0^1 V(u)V(u)'du = J_1. \tag{A.10}$$

Furthermore, the stochastic orders of the second diagonal block and the off-diagonal block of (A.6), together with $\widehat{\Sigma}_n(s) = \mathrm{O}_p(1)$, imply

$$D_{2n}'\sum_{t=1}^n [Z_tZ_t' \otimes \widehat{\Sigma}_t^{-1}]D_{2n} = \mathrm{O}_p(1), \qquad D_{1n}'\sum_{t=1}^n [Z_tZ_t' \otimes \widehat{\Sigma}_t^{-1}]D_{2n} = \mathrm{o}_p(1).$$

Given the limiting block-diagonality, the actual limit of the second diagonal block will not be needed.

For the result corresponding to (A.8), the main task is to derive the limit of

$$D_{1n}' \sum_{t=1}^{n} [Z_t \otimes \widehat{\Sigma}_t^{-1}] \varepsilon_t = n^{-1} \sum_{t=1}^{n} \widehat{V}_{t-1}^0 z_t, \tag{A.11}$$

where $\widehat{V}_{t-1}^0 = [\alpha_\perp' \Gamma X_{t-1} \otimes \widehat{\Sigma}_t^{-1} \sigma_t] = \widehat{V}_{t-1}[I_q \otimes \widehat{\sigma}_t^{-1} \sigma_t]$, such that the continuous mapping theorem again implies $n^{-1/2} \widehat{V}_{\lfloor un \rfloor}^0 \xrightarrow{w} V(u), u \in [0,1]$. To show that (A.11) converges weakly to the stochastic integral $\int_0^1 V(u) dW(u)$, we cannot apply Hansen (1992)'s Theorem 2.1, because $\{\widehat{V}_t^0\}_{t \geq 1}$ is not adapted to the filtration $\{\mathcal{F}_t\}_{t \geq 1}$, with respect to which $\{z_t\}_{t \geq 1}$ is an mds. This is caused by the fact that $\widehat{\Sigma}_t$ depends on the full sample. For the same reason, we cannot decompose z_t into a martingale part and a remainder, a technique that is often useful to deal with dependent processes, see Hansen (1992). Instead, we will follow the approach by Chan and Wei (1988), Theorem 2.4, which was extended by Davidson (1994) and De Jong and Davidson (2000).

Let $V_n(u) = n^{-1/2} \widehat{V}_{\lfloor un \rfloor}^0$ and note that $(V_n, W_n) \xrightarrow{w} (V, W)$ in $\mathcal{D}[0,1]^{qp \times p} \times \mathcal{D}[0,1]^p$ where the limit (V, W) has continuous sample paths. The Skorohod representation theorem implies the existence of sequences (V^n, W^n) in $D[0,1]^{qp \times p} \times D[0,1]^p$, defined on an underlying probability space $(\Omega, \mathcal{F}, P)$, such that $(V^n, W^n) \xrightarrow{a.s.} (V, W)$ in $\mathcal{D}[0,1]^{qp \times p} \times \mathcal{D}[0,1]^p$. This implies that, given $\epsilon > 0$, there exists an event $\Omega_\epsilon \subset \Omega$ such that $P(\Omega_\epsilon) \geq 1 - \epsilon$ and

$$\sup_{\omega \in \Omega_\epsilon} d\left((V^n, W^n)(\omega) - (V, W)(\omega)\right) = \delta_n \to 0,$$

where $d(\cdot, \cdot)$ is the uniform metric.

Let $\{k_n, n \in \mathbb{N}\}$ be an increasing integer subsequence, such that $k_n / n \to 0$ and $k_n \delta_n^2 \to 0$. For each k_n, choose integers $0 = n_0 < n_1 < n_2 < \ldots < n_{k_n} = n$, corresponding to a partition

$$0 = u_0 = \frac{n_1}{n} < u_1 = \frac{n_1}{n} < u_2 = \frac{n_2}{n} < \ldots < u_{k_n} = \frac{n_{k_n}}{n} = 1,$$

with $\min_{1 \leq j \leq k_n} |n_j - n_{j-1}| \to \infty$ and $\max_{1 \leq j \leq k_n} |u_j - u_{j-1}| \to 0$ as $n \to \infty$. Consider the decomposition

$$\begin{aligned}
\int_0^1 V_n(u) dW_n(u) &= n^{-1} \sum_{t=1}^{n} \widehat{V}_{t-1}^0 z_t \\
&= \sum_{j=1}^{k_n} V_n(u_{j-1}) \left[W_n(u_j) - W_n(u_{j-1})\right] + \frac{1}{n} \sum_{j=1}^{k_n} \sum_{t=n_{j-1}+2}^{n_j} (\widehat{V}_{t-1}^0 - \widehat{V}_{n_{j-1}}^0) z_t \\
&=: G_n + Q_n.
\end{aligned}$$

Analogous to the arguments in Chan and Wei (1988) and Davidson (1994), $G_n \xrightarrow{w} \int_0^1 V(u) dW(u)$ follows from $G^n \xrightarrow{p} \int_0^1 V(u) dW(u)$, where

$$G^n = \sum_{j=1}^{k_n} V^n(u_{j-1}) \left[W^n(u_j) - W^n(u_{j-1})\right].$$

For the remainder term Q_n, we note that

$$\begin{aligned} Q_n &= \frac{1}{n}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}+2}^{n_j}(\widehat{V}^0_{t-1}-\widehat{V}^0_{n_{j-1}})z_t \\ &= \frac{1}{n}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}+2}^{n_j}(V_{t-1}-V_{n_{j-1}})z_t \\ &\quad +\frac{1}{n}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}+2}^{n_j}\left[(\widehat{V}^0_{t-1}-V_{t-1})-(\widehat{V}^0_{n_{j-1}}-V_{n_{j-1}})\right]z_t \\ &=: Q_{1n}+Q_{2n}. \end{aligned}$$

$Q_{1n} = \mathrm{o}_p(1)$ because it converges to a stochastic integral with respect to $W(u)$, where the integrand is the limit of the difference between $n^{-1/2}V_{\lfloor un\rfloor}$ and a discretised version thereof (which changes values only at times $u = n_j/n$); and this difference converges to 0. Using the uniform consistency result of Lemma 2, $Q_{2n} = \mathrm{o}_p(1)$ because it is the average over n terms, each of which is $o_p(1)$. Therefore, $Q_n = \mathrm{o}_p(1)$, such that

$$\begin{aligned} D'_{1n}\sum_{t=1}^{n}[Z_t\otimes\widehat{\Sigma}_t^{-1}]\varepsilon_t = \int_0^1 V_n(u)dW_n(u) &\xrightarrow{w} \int_0^1 V(u)dW(u) \\ &= \int_0^1 \left[U_A(u)\otimes\Sigma(u)^{-1}\right]dM(u). \quad \text{(A.12)} \end{aligned}$$

The corresponding result for $D'_{1n}\sum_{t=1}^{n}[Z_t\otimes\widehat{\Sigma}_t^{-1}]\varepsilon_t(\theta_0)$ under local alternatives follows from combining (A.10) with (A.12).

The results obtained so far can also be used to show that $D'_{2n}\sum_{t=1}^{n}[Z_t\otimes\widehat{\Sigma}_t^{-1}]\varepsilon_t = \mathrm{O}_p(1)$. Combining these results implies

$$\widehat{\Lambda}_n(\theta_0+D_n\tau)\xrightarrow{w}(S_1-J_1\tau_1)'J_1^{-1}(S_1-J_1\tau)+\Lambda_2(\tau_2),$$

uniformly on compact sets. Following the same steps as the proof of Theorem 1, this implies $\widehat{\mathrm{LR}}_n \xrightarrow{w} \min_{\tau\in\mathcal{T}_r}\Lambda(\tau)$, such that $\widehat{\mathrm{LR}}_n$ has the same limiting distribution as LR_n. (Note that the exact form of the quadratic function $\Lambda_2(\tau_2)$ is irrelevant.) □

Proof of Theorem 3. The starting point is a conditional invariance principle for the bootstrap errors ε_t^*. For the volatility bootstrap, consider

$$M_n^*(u) := n^{-1/2}\sum_{t=1}^{\lfloor un\rfloor}\widehat{\sigma}_t z_t^*, \qquad u\in[0,1],$$

which conditional on the data is a Gaussian process with independent increments and variance matrix $n^{-1}\sum_{t=1}^{\lfloor un\rfloor}\widehat{\Sigma}_t = \int_0^u\widehat{\Sigma}_n(s)ds$. Uniform consistency of $\widehat{\Sigma}_n(\cdot)$ implies that $\int_0^u\widehat{\Sigma}_n(s)ds$ converges in probability to $\int_0^u\Sigma(s)ds$ uniformly in $u\in[0,1]$, and this in turn implies that

$$M_n^*(u)\xrightarrow{w}_p M(u), \qquad u\in[0,1]. \quad \text{(A.13)}$$

For the wild bootstrap, we use $M_n^*(u) = n^{-1/2}\sum_{t=1}^{\lfloor un\rfloor}e_t w_t^*$, and taking w_t^* to be i.i.d. $N(0,1)$, this is a Gaussian process with independent increments and variance matrix $n^{-1}\sum_{t=1}^{\lfloor un\rfloor}e_t e_t'$

conditional on the data. Using the notation and the results from the proof of Lemma 2 regarding the least-squares residuals $e_t = \varepsilon_t - (\bar{B}_n - B)Z_t$, we find

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor} e_t e_t' &= \frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor}\varepsilon_t\varepsilon_t' - (\bar{B}_n - B)\frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor} Z_t\varepsilon_t' - \frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor}\varepsilon_t Z_t'(\bar{B}_n - B)' \\
&\quad + \frac{1}{n}(\bar{B}_n - B)\sum_{t=1}^{\lfloor un\rfloor} Z_t Z_t'(\bar{B}_n - B)' \\
&= \frac{1}{n}\sum_{t=1}^{\lfloor un\rfloor}\varepsilon_t\varepsilon_t' + \mathrm{o}_p(1) \\
&\xrightarrow{p} \int_0^u \Sigma(s)ds,
\end{aligned}$$

uniformly in $u \in [0,1]$; see Lemma A.5 of Cavaliere *et al.* (2010) for the final result. Therefore, (A.13) also applies to the wild bootstrap.

The next step is to prove the bootstrap version of Lemma 1, under the null hypothesis, i.e., with $A = 0$. We follow the approach of the proof of this lemma, noting that the bootstrap observations $\{X_t^*\}_{t\geq 1}$ satisfy

$$\mathbb{X}_t^* = \widetilde{\Phi}_n^t \mathbb{X}_0 + \sum_{j=1}^{t} \widetilde{\Phi}_n^{t-j} F\varepsilon_j^*,$$

where $\mathbb{X}_t^* = (X_t^{*\prime}, \ldots, X_{t-k+1}^{*\prime})'$ and $\widetilde{\Phi}_n = I_{kp} + \widetilde{\mathbb{A}}_n\widetilde{\mathbb{B}}_n'$, with $\widetilde{\mathbb{A}}_n$ and $\widetilde{\mathbb{B}}_n$ the restricted ML estimators of $\mathbb{A}$ and $\mathbb{B}$, respectively. Because it follows from the proof of Theorem 2 that these estimators are consistent (both under $\mathcal{H}(r)$ and under $\mathcal{H}_n(r, r_1)$), it also follows that for n large enough, all eigenvalues of the matrix $I_{r+(k-1)p} + \widetilde{\mathbb{B}}_n'\widetilde{\mathbb{A}}_n$ will be less than one in absolute value, such that

$$\begin{aligned}
n^{-1/2}\widetilde{\mathbb{B}}_n'\mathbb{X}_{\lfloor un\rfloor}^* &= n^{-1/2}(I_{r+(k-1)p} + \widetilde{\mathbb{B}}_n'\widetilde{\mathbb{A}}_n)^{\lfloor un\rfloor}\widetilde{\mathbb{B}}_n'\mathbb{X}_0 \\
&\quad + n^{-1/2}\sum_{j=1}^{\lfloor un\rfloor}(I_{r+(k-1)p} + \widetilde{\mathbb{B}}_n'\widetilde{\mathbb{A}}_n)^{\lfloor un\rfloor - j}\widetilde{\mathbb{B}}_n' F\varepsilon_j^* + \mathrm{o}_p(1) \\
&\xrightarrow{p} 0,
\end{aligned}$$

and, using consistency of $(\widetilde{\alpha}_n, \widetilde{\Psi}_n)$ and hence $\widetilde{\mathbb{A}}_{n\perp}$,

$$\begin{aligned}
n^{-1/2}\widetilde{\mathbb{A}}_{n\perp}'\mathbb{X}_{\lfloor un\rfloor}^* &= n^{-1/2}\widetilde{\mathbb{A}}_{n\perp}'\widetilde{\Phi}_n^{\lfloor un\rfloor}\mathbb{X}_0 + n^{-1/2}\sum_{j=1}^{\lfloor un\rfloor}\widetilde{\mathbb{A}}_{n\perp}'\widetilde{\Phi}_n^{\lfloor un\rfloor - j}F\varepsilon_j^* \\
&\xrightarrow{w} U_0(u) = \alpha_\perp' M(u), \qquad u \in [0,1], \qquad \text{(A.14)}
\end{aligned}$$

analogously to the proof of Lemma 1.

From these results, the proof of the theorem proceeds by following the steps of the proofs of Theorems 1–2. Because the bootstrap statistics $\widehat{\mathrm{LR}}_n^*$ are based on the volatility matrix estimates $\{\widehat{\Sigma}_t\}_{t=1}^n$ obtained from the the original data, the proof of the bootstrap versions of (A.10) and the limit of (A.11) follows directly from combining these steps with (A.13) and (A.14). Results for the score and information for the parameters characterising the stable dynamics (S_2 and J_2) follow analogously to the proof of Proposition 4.1 of Patilea and Raïssi (2012). □

References

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (2003), "Modeling and forecasting realized volatility," *Econometrica* 71, 579–625.

Beare, B. K. (2018), "Unit root testing with unstable volatility," *Journal of Time Series Analysis* 39, 816–835.

Boswijk, H. P. (1995), "Identifiability of cointegrated systems," Tinbergen Institute Discussion Paper # 95-78, `http://dare.uva.nl/document/2/163397`.

Boswijk, H. P. (2001), "Testing for a unit root with near-integrated volatility," Tinbergen Institute Discussion Paper # 01-077/4, `http://papers.tinbergen.nl/01077.pdf`.

Boswijk, H. P. and Y. Zu (2018), "Adaptive wild bootstrap tests for a unit root with non-stationary volatility," *Econometrics Journal* 21, 87–113.

Boswijk, H. P., G. Cavaliere, A. Rahbek and A. M. R. Taylor (2016), "Inference on co-integration parameters in heteroskedastic vector autoregressions," *Journal of Econometrics* 192, 64–85.

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

Cavaliere, G. and A. M. R. Taylor (2007), "Testing for unit roots in time series models with nonstationary volatility," *Journal of Econometrics* 140, 919–947.

Cavaliere, G. and A. M. R. Taylor (2008), "Bootstrap unit root tests for time series with nonstationary volatility," *Econometric Theory* 24, 34–71.

Cavaliere, G., A. Rahbek and A. M. R. Taylor (2010), "Testing for cointegration in vector autoregressions with non-stationary volatility," *Journal of Econometrics* 158, 7–24.

Cavaliere, G., A. Rahbek and A. M. R. Taylor (2014), "Bootstrap determination of the co-integration rank in heteroskedastic VAR models," *Econometric Reviews* 33, 606–650.

Chan, N. H. and C. Z. Wei (1988), "Limiting distributions of least squares estimates of unstable autoregressive processes," *Annals of Statistics* 16, 367–401.

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

De Jong, R. M. and J. Davidson (2000), "The functional central limit theorem and weak convergence to stochastic integrals I: Weakly dependent processes," *Econometric Theory* 16, 621–642.

Doornik, J. A. (1998), "Approximations to the asymptotic distribution of cointegration tests," *Journal of Economic Surveys* 12, 573–593.

Duan, J.-C. and S. R. Pliska (2004), "Option valuation with cointegrated asset prices," *Journal of Economic Dynamics and Control* 28, 727–754.

Engle, R. F. (2002), "Dynamic conditional correlation: A simple class of multivariate GARCH models," *Journal of Business & Economic Statistics* 20, 339–350.

Engle, R. F. and C. W. J. Granger (1987), "Cointegration and error correction: Representation, estimation and testing," *Econometrica* 55, 251–276.

Fan, J., Y. Li and K. Yu (2012), "Vast volatility matrix estimation using high-frequency data for portfolio selection," *Journal of the American Statistical Association* 107, 412–428.

Giné, E. and J. Zinn (1990), "Bootstrapping general empirical measures," *Annals of Probability* 18, 851–869.

Hansen, B. E. (1992), "Convergence to stochastic integrals for dependent heterogeneous processes," *Econometric Theory* 8, 489–500.

Hansen, B. E. (1995), "Regression with nonstationary volatility," *Econometrica* 63, 1113–1132.

Hansen, P. R. (2003), "Structural changes in the cointegrated vector autoregressive model," *Journal of Econometrics* 114, 261–295.

Hansen, P. R. and S. Johansen (1998), *Workbook on Cointegration*. Oxford: Oxford University Press.

Johansen, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed. Oxford: Oxford University Press.

Johansen, S. (1997), "Likelihood analysis of the $I(2)$ model," *Scandinavian Journal of Statistics* 24, 433–462.

Kim, T.-H., S. Leybourne and P. Newbold (2003), "Unit root tests with a break in innovation variance," *Journal of Econometrics* 109, 365–387.

Ling, S. and W. K. Li (1998), "Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with GARCH errors," *Annals of Statistics* 26, 84–125.

Ling, S. and W. K. Li (2003), "Asymptotic inference for unit root processes with GARCH(1,1) errors," *Econometric Theory* 19, 541–564.

Li, W. K., S. Ling and H. Wong (2001), "Estimation for partially nonstationary multivariate autoregressive models with conditional heteroskedasticity," *Biometrika* 88, 1135–1152.

Magnus, J. R. and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: John Wiley & Sons.

Nelson, D. B. (1996), "Asymptotic filtering theory for multivariate ARCH models," *Journal of Econometrics* 71, 1–47.

Oberhofer W. and J. Kmenta (1974), "A general procedure for obtaining maximum likelihood estimates in generalized regression models," *Econometrica* 42, 579–590.

Patilea, V. and H. Raïssi (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.

Phillips, P. C. B. and K.-L. Xu (2006), "Inference in autoregression under heteroskedasticity," *Journal of Time Series Analysis* 27, 289–308.

Sargan, J. D., "Wages and prices in the United kingdom: A study in econometric methodology," in Hart, P. E., G. Mills and J. K. Whitaker (eds.), *Econometric Analysis for National Economic Planning*, pp. 25–63. London: Butterworth.

Seo, B. (1999), "Distribution theory for unit root tests with conditional heteroskedasticity," *Journal of Econometrics* 91, 113–144.

Seo, B. (2007), "Asymptotic distribution of the cointegrating vector estimator in error correction

models with conditional heteroskedasticity," *Journal of Econometrics* 137, 68–111.

Van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Wong, H., W. K. Li and S. Ling (2005), "Joint modeling of cointegration and conditional heteroskedasticity with applications," *Annals of the Institute of Statistical Mathematics* 57, 83–103.

Xu, K.-L. and P. C. B. Phillips (2008), "Adaptive estimation of autoregressive models with time-varying variances," *Journal of Econometrics* 142, 265–280.