UvA ECONOMETRICS

Discussion Paper: 2007/06

Testing for cointegration with nonstationary volatility

H. Peter Boswijk and Yang Zu

www.feb.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Testing for Cointegration with Nonstationary Volatility

H. Peter Boswijk* & Yang Zu
Tinbergen Institute & Amsterdam School of Economics
Universiteit van Amsterdam

December 19, 2007

Abstract

The paper generalises recent unit root tests for nonstationary volatility to a multivariate context. Persistent changes in the innovation variance matrix lead to size distortions in conventional cointegration tests, and possibilities of increased power by taking the time-varying volatilities and correlations into account. The testing procedures are based on a likelihood analysis of the vector autoregressive model with a conditional covariance matrix that may be estimated nonparametrically. We find that under suitable conditions, adaptation with respect to the volatility matrix process is possible, in the sense that nonparametric volatility estimation does not lead to a loss of asymptotic local power.

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 (1995). 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 quasi-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.

*Corresponding author: H. Peter Boswijk, Department of Quantitative Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: H.P.Boswijk@uva.nl.

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 (2008) and Beare (2004) develop alternative approaches to constructing unit root test statistics with the conventional (Dickey-Fuller) asymptotic null distribution, whereas Cavaliere and Taylor (2007) propose to apply the standard Dickey test but with simulated p-values to obtain the correct asymptotic size. Boswijk (2005) derives the power envelope for unit root tests with observable (nonstationary) volatility, and shows that considerable power gains may be obtained relative to procedures that do not take the heteroskedasticity into account. He also shows 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 (2005)'s analysis to a multivariate context, and hence develop efficient tests for cointegration in the presence of a nonstationary multivariate heteroskedasticity. First, building on the analysis of Hansen (2002, 2003), we derive the likelihood ratio test for cointegration in a vector autoregressive model with observed time-varying variance matrices and conditionally Gaussian disturbances. The asymptotic null distribution of this test is nonstandard and depends on the realization of the volatility process, such that asymptotic p-values have to be obtained by Monte Carlo simulation. Next, we consider the case of unknown volatility, and propose a two-step procedure where the volatility process is estimated nonparametrically. Under suitable conditions, these estimators are consistent and hence the resulting cointegration tests have the same asymptotic power function as in the case of known volatility.

In a recent paper, Cavaliere *et al.* (2007) develop another approach to the same problem, which involves applying the wild bootstrap to the traditional (constant-variance) Gaussian likelihood ratio test. Their approach is very effective in solving the size distortions, but does not exploit the power gain potential in the presence of nonstationary volatility. However, they require less stringent assumptions on the volatility process than the present paper. In particular, they allow for level shifts in the volatility process, which are excluded here because the nonparametric estimator requires continuous volatility sample paths.

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 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 contains an empirical application of the test to the S&P 500 and NASDAQ-100 indices in the 1990s. Section 6 contains some concluding remarks, and proofs of all results are given in an appendix.

We use the notation $X_n \xrightarrow{\mathcal{L}} X$ to denote convergence in distribution for sequences of random variables or vectors, and $X_n(s) \xrightarrow{\mathcal{L}} X(s), s \in [0,1]$ to denote weak convergence in $D[0,1]^k$, the product space of right-continuous functions with finite left limits, under the uniform 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 $\bar{A} = A(A'A)^{-1}$.

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:

$$\mathcal{H}(r) : \Pi = \alpha\beta', \tag{2}$$

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

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t. \tag{3}$$

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, 1995, Theorem 4.2):

Assumption 1 *In the model (3),* (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) $\mathrm{rank}(\alpha'_\perp \Gamma \beta_\perp) = p - r$*, where* $\Gamma = I_p - \sum_{j=1}^{k-1} \Gamma_j$.

Under the assumption that the disturbances $\{\varepsilon_t\}$ 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 (1995). Here we consider a deviation from the i.i.d. assumption, in that we allow for heteroskedasticity:

Assumption 2 *In the model (3), the disturbances satisfy*

$$\varepsilon_t = \sigma_t \eta_t, \qquad t = 1, 2, \ldots, \tag{4}$$

where $\{\sigma_t\}_{t \geq 1}$ is a sequence of possibly stochastic non-singular $p \times p$ matrices, and $\{\eta_t\}_{t \geq 1}$ is an i.i.d. sequence, with

$$E(\eta_1) = 0, \qquad \mathrm{var}(\eta_1) = I_p, \qquad t = 1, 2, \ldots, \tag{5}$$

and with η_t independent of $\mathcal{F}_{t-1} = \sigma(\{\eta_{t-1-j}, \sigma_{t-j}\}_{j \geq 0}), t = 1, 2, \ldots$.

The assumption directly implies $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $\mathrm{var}(\varepsilon_t|\mathcal{F}_{t-1}) = \sigma_t\sigma_t' =: \Sigma_t$, a positive definite conditional variance matrix. We will refer to σ_t, a matrix square root of Σ_t, as the *volatility matrix* of ε_t. Note that the definition of the filtration $\{\mathcal{F}_t\}_{t\geq 0}$ allows for deterministic volatility, but also for multivariate GARCH processes, where σ_t is a function of $\{\varepsilon_{t-j}\}_{j\geq 1}$, or stochastic volatility specifications, where σ_t is a stochastic process driven by its own innovations, possibly depending on $\{\varepsilon_{t-j}\}_{j\geq 1}$ but independent of η_t. In the next section, we will analyse the likelihood function derived from the stronger assumption $\varepsilon_t|\mathcal{F}_{t-1} \sim N(0, \Sigma_t)$, but the asymptotic properties of the resulting procedures will continue to hold under Assumption 2 with non-Gaussian $\{\eta_t\}$.

If the volatility process is stationary and ergodic, such that the average variance matrix $n^{-1}\sum_{t=1}^{n}\Sigma_t$ converges to a finite and positive definite matrix Σ as $n \to \infty$, then, under suitable technical conditions, the invariance principle will apply to $\{\varepsilon_t\}$. This in turn would imply that Johansen's (quasi-) likelihood ratio test, based on the Gaussian i.i.d. assumption on $\{\varepsilon_t\}$, would retain its usual asymptotic properties, even though more efficient tests may obtained from an analysis of the true likelihood function. In this paper, we consider the case where $\{\sigma_t\}$ displays nonstationary variation, such that the invariance principle does not apply. Define, for $n = 1, 2, \ldots$,

$$W_n(s) = n^{-1/2}\sum_{t=1}^{\lfloor sn\rfloor}\eta_t, \qquad s \in [0,1], \tag{6}$$

a vector-valued process in $D[0,1]^p$, and

$$\sigma_n(s) = \begin{cases} \sigma_{\lfloor sn\rfloor+1}, & s \in [0,1), \\ \sigma_n, & s = 1, \end{cases} \tag{7}$$

a matrix-valued process in $D[0,1]^{p\times p}$.

Assumption 3 *As $n \to \infty$,*

$$(W_n(s), \sigma_n(s)) \xrightarrow{\mathcal{L}} (W(s), \sigma(s)), \qquad s \in [0,1], \tag{8}$$

where $W(\cdot)$ is a standard p-variate Brownian motion process, and $\sigma(\cdot)$ is a $p \times p$ matrix-valued process on $[0,1]$, independent of $W(\cdot)$. The elements $\sigma_{ij}(\cdot)$ of $\sigma(\cdot)$ have continuous sample paths and satisfy $E\left(\int_0^1 \sigma_{ij}(s)^2 ds\right) < \infty$ for all $i, j = 1, \ldots, p$.

This assumption is a multivariate generalization of Assumption 1 in Boswijk (2005), which in turn was inspired by Hansen (1995). Note that the invariance principle for $\{\eta_t\}$ follows directly from Assumption 2, but it is included because joint convergence will be needed. The condition on the limiting behaviour of $\{\sigma_t\}$ is motivated by the notion that persistent changes in the volatility should be preserved in the limit. Beare (2004), Cavaliere (2004), Cavaliere and Taylor (2007, 2008) and Phillips and Xu (2006) consider, in the univariate context, the closely related assumption $\sigma_t = \sigma(t/n)$ for some deterministic function $\sigma(\cdot)$, which implies $\sigma_n(s) \to \sigma(s), s \in [0,1]$. One instance where the assumption arises naturally is in the context of continuous-record asymptotics, where the number of observations is increased by considering an increasing sampling frequency over a fixed time span; see

also Nelson (1990). This also clarifies that $\{\sigma_t\}$, and hence $\{\varepsilon_t\}$ and $\{X_t\}$, are in fact triangular arrays $\{(X_{nt}, \varepsilon_{nt}, \sigma_{nt}), t = 1, \ldots, n; n = 1, 2, \ldots\}$, although we suppress the double index notation for simplicity.

Before we consider likelihood-based testing for $\mathcal{H}(r)$ the model (3) 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

$$\mathcal{H}_n(r, r_1) : \Pi_n = \alpha\beta' + n^{-1}\alpha_1\beta_1', \tag{9}$$

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 (1995) 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 (9). Proofs of all results are given in the Appendix.

Lemma 1 *In the model (3) under Assumptions 1–3 and under $\mathcal{H}_n(r, r_1)$, we have*

$$n^{-1/2} \sum_{t=1}^{\lfloor sn \rfloor} \varepsilon_t \xrightarrow{\mathcal{L}} \int_0^s \sigma(u) dW(u) =: U(s), \qquad s \in [0, 1], \tag{10}$$

and

$$n^{-1/2} X_{\lfloor sn \rfloor} \xrightarrow{\mathcal{L}} \beta_\perp (\alpha_\perp' \Gamma \beta_\perp)^{-1} K(s) =: X(s), \qquad s \in [0, 1], \tag{11}$$

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

$$K(s) = \int_0^s \exp\left((s - u)A\right) \sigma_K(u) dW(u), \qquad s \in [0, 1], \tag{12}$$

with $A = \alpha_\perp' \alpha_1 \beta_1' \beta_\perp (\alpha_\perp' \Gamma \beta_\perp)^{-1}$ and $\sigma_K(s) = \alpha_\perp' \sigma(s)$, such that $K(\cdot)$ satisfies the stochastic differential equation

$$dK(s) = AK(s)ds + \sigma_K(s)dW(s). \tag{13}$$

The limit $X(\cdot)$ of $n^{-1/2} X_{\lfloor \cdot n \rfloor}$ is a p-variate process, but of rank $p - r$, in the sense that $\beta' X(s) = 0$ (a.s.). Note that $K(\cdot)$ 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 $K(s)$ reduces to $\alpha_\perp' U(s)$.

Lemma 1 implies that the limiting null distribution of the quasi-likelihood ratio (QLR) 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 that the QLR statistic satisfies, under the null hypothesis,

$$QLR_n(0) \xrightarrow{\mathcal{L}} \operatorname{tr}\left\{ \int_0^1 dU(s)U(s)' \left(\int_0^1 U(s)U(s)'ds \right)^{-1} \int_0^1 U(s)dU(s)' \left(\int_0^1 \Sigma(s)ds \right)^{-1} \right\}, \tag{14}$$

with $\Sigma(s) = \sigma(s)\sigma(s)'$. If and only if $\sigma(\cdot)$ is a constant matrix σ, such that $U(s) = \sigma W(s)$, the usual limiting distribution tabulated in Johansen (1995) will result.

3 The likelihood ratio rest with known volatility

In this section we analyse the likelihood ratio (LR) statistic for $\mathcal{H}(r)$ in the model (3) in the case where $\{\sigma_t\}$ is known, and where the standardized innovations $\{\eta_t\}$ are taken to be i.i.d. $N(0, I_p)$. Although the assumption that $\{\sigma_t\}$ 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 $\{\sigma_t\}$ is unknown and hence has to be estimated, either based on a parametric model or nonparametrically.

Define $\Psi = [\Gamma_1 : \cdots : \Gamma_{k-1}]$ and $W_t = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1})'$, such that the model (3) under Assumption 2 with Gaussian $\{\eta_t\}$ may be expressed more compactly as

$$\Delta X_t = \alpha\beta' X_{t-1} + \Psi W_t + \varepsilon_t, \qquad \varepsilon_t|\mathcal{F}_{t-1} \sim N(0, \Sigma_t), \qquad t = 1, \ldots, n, \tag{15}$$

where $\Sigma_t = \sigma_t\sigma'_t$ as before. Recall that the starting values $\{X_{1-k}, \ldots, X_0\}$, 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. The only assumption we make is that $\{\sigma_t\}$ is either fixed or weakly exogenous, in the sense of Engle *et al.* (1983), for the parameters (α, β, Ψ). This means that the joint density of (X_t, σ_t) given the past $\mathcal{G}_{t-1} = \sigma(\{X_{t-j}, \sigma_{t-j}\}_{j>0})$ factorises into the conditional Gaussian distribution of X_t given σ_t and $\mathcal{G}_{t-1}$, with parameters (α, β, Ψ), and the marginal distribution of σ_t given $\mathcal{G}_{t-1}$, the (possibly infinite-dimensional) parameters of which are variation independent of (α, β, Ψ). Under this condition, the log-likelihood function is given by

$$\begin{aligned} \ell_n(\alpha, \beta, \Psi) \quad = \quad & -\frac{Tp}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^n \log|\Sigma_t| \\ & -\frac{1}{2}\sum_{t=1}^n (\Delta X_t - \alpha\beta' X_{t-1} + \Psi W_t)'\Sigma_t^{-1}(\Delta X_t - \alpha\beta' X_{t-1} + \Psi W_t). \end{aligned} \tag{16}$$

Maximum likelihood estimation in a closely related class of models was studied by Hansen (2002, 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) $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Psi}_n)$, the maximization of $\ell_n(\alpha, \beta, \Psi)$ over (α, Ψ) 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. Properties of such switching algorithms have been studied by Oberhofer and Kmenta (1974): since the maximised likelihood is nondecreasing in each step, it does eventually converge, although not necessarily very fast, and not necessarily to a global maximum.

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 of full column rank. An equivalent formulation is $\beta = \bar{c} + c_\perp\Phi$,

[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 $\mathcal{H}_n(r, r_1)$.

where Φ is a $(p-r)\times r$ matrix of free parameters, such that

$$\operatorname{vec}\beta = \operatorname{vec}\left(\bar{c} + c_{\perp}\Phi\right) = h + H\phi, \tag{17}$$

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)

$$\operatorname{vec}[\hat{\alpha}_n(\beta) : \hat{\Psi}_n(\beta)] = \left(\sum_{t=1}^{n}\left[Z_t(\beta)Z_t(\beta)' \otimes \Sigma_t^{-1}\right]\right)^{-1} \operatorname{vec}\left(\sum_{t=1}^{n}\Sigma_t^{-1}\Delta X_t Z_t(\beta)'\right), \tag{18}$$

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

$$\begin{aligned}\operatorname{vec}\hat{\beta}_n(\alpha,\Psi) &= h + H\left(H'\sum_{t=1}^{n}\left[\alpha'\Sigma_t^{-1}\alpha \otimes X_{t-1}X'_{t-1}\right]H\right)^{-1} H' \\ &\quad \times \sum_{t=1}^{n}\left\{\operatorname{vec}\left(X_{t-1}(\Delta X_t - \Psi W_t)'\Sigma_t^{-1}\alpha\right) - \left[\alpha'\Sigma_t^{-1}\alpha \otimes X_{t-1}X'_{t-1}\right]h\right\}.\end{aligned} \tag{19}$$

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

$$\hat{\varepsilon}_t = \Delta X_t - [\hat{\alpha}_n : \hat{\Psi}_n]Z_t(\hat{\beta}_n) = \Delta X_t - \hat{\alpha}_n\hat{\beta}'_n X_{t-1} - \hat{\Psi}_n W_t, \qquad t = 1,\ldots,n. \tag{20}$$

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

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

$$\operatorname{vec}[\tilde{\Pi}_n : \tilde{\Psi}_n] = \left(\sum_{t=1}^{n}\left[Z_t Z'_t \otimes \Sigma_t^{-1}\right]\right)^{-1}\operatorname{vec}\left(\sum_{t=1}^{n}\Sigma_t^{-1}\Delta X_t Z'_t\right), \tag{21}$$

yielding the unrestricted residuals

$$\tilde{\varepsilon}_t = \Delta X_t - [\tilde{\Pi}_n : \tilde{\Psi}_n]Z_t = \Delta X_t - \tilde{\Pi}_n X_{t-1} - \tilde{\Psi}_n W_t, \qquad t = 1,\ldots,n. \tag{22}$$

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

$$LR_n(r) = -2\left[\ell_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\Psi}_n) - \ell_n(\tilde{\Pi}_n, I_{p+1}, \tilde{\Psi}_n)\right] = \sum_{t=1}^{n}\left(\hat{\varepsilon}'_t\Sigma_t^{-1}\hat{\varepsilon}_t - \tilde{\varepsilon}'_t\Sigma_t^{-1}\tilde{\varepsilon}_t\right). \tag{23}$$

The limiting behaviour of $LR_n(r)$ is characterised in Theorem 1. Define

$$Y(s) = \begin{pmatrix} Y_1(s) \\ Y_2(s)\end{pmatrix} = \begin{pmatrix}\alpha'_{\perp} \\ \alpha'\end{pmatrix}\left(\sigma(s)'^{-1} \otimes K(s)\right), \tag{24}$$

and

$$Z(s) = Y_1(s) - \int_0^1 Y_1(u)Y_2(u)'du\left[\int_0^1 Y_2(u)Y_2(u)'du\right]^{-1} Y_2(s). \tag{25}$$

Theorem 1 *In the model (3), under Assumptions 1–3 and under $\mathcal{H}_n(r, r_1)$, the LR statistic (23) satisfies, as $n \to \infty$,*

$$LR_n(r) \xrightarrow{\mathcal{L}} \left(\int_0^1 Z(s)[dW(s) + Z(s)' \operatorname{vec}(A')ds]\right)' \left(\int_0^1 Z(s)Z(s)'ds\right)^{-1} \times \left(\int_0^1 Z(s)[dW(s) + Z(s)' \operatorname{vec}(A')ds]\right). \tag{26}$$

We observe that the limiting distribution under the null hypothesis $\mathcal{H}(r)$, such that $A = 0$, depends on (the process generating) $\sigma(s)$, and on α (and hence $\alpha_\perp$). Therefore, no uniformly applicable tables of critical values can be constructed. Because $W(\cdot)$ is independent of $\sigma(\cdot)$ by Assumption 3, quantiles and p-values of the limiting distribution conditional on $\sigma(\cdot)$ can be obtained by Monte Carlo simulation of the limiting expression in (26), replacing α by $\hat{\alpha}_n$. Consistency of $\hat{\alpha}_n$ (which follows from the proof of Theorem 1) guarantees the asymptotic validity of such p-values, as both the sample size and the number of Monte Carlo replications tend to infinity.

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 (26), with $Z(s) = \sigma(s)'^{-1} \otimes K(s)$ and

$$dK(s) = \sigma(s)[dW(s) + \sigma(s)^{-1}AK(s)ds] = \sigma(s)[dW(s) + Z(s)' \operatorname{vec}(A')ds]. \tag{27}$$

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

$$LR_n(0) \xrightarrow{\mathcal{L}} \int_0^1 dK(s)' \left[\Sigma(s)^{-1} \otimes K(s)'\right] \left(\int_0^1 [\Sigma(s)^{-1} \otimes K(s)K(s)']ds\right)^{-1} \times \int_0^1 \left[\Sigma(s)^{-1} \otimes K(s)\right] dK(s). \tag{28}$$

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, 1995, Chapters 5–6), this is accomplished by considering the following two extensions of (3). To allow for a constant mean, the model becomes

$$\begin{aligned} \Delta X_t &= \alpha(\beta' X_{t-1} + \rho_0) + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t \\ &= \alpha\beta^{*\prime} X^*_{t-1} + \Psi W_t + \varepsilon_t, \end{aligned} \tag{29}$$

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

$$\begin{aligned} \Delta X_t &= \mu + \alpha(\beta' X_{t-1} + \rho_1 t) + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t \\ &= \alpha\beta^* X^*_{t-1} + \Psi^* W^*_t + \varepsilon_t, \end{aligned} \tag{30}$$

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 (16), with parameters and regressors replaced by their starred counterparts.

Adjusting the identification restrictions (17) accordingly, such that $\operatorname{vec}\beta^* = \operatorname{vec}(\bar{c}^* + c^*_\perp \Phi^*)$ with c^* of dimensions $(p+1)\times r$ and hence $c^*_\perp$ and Φ^* of dimensions $(p+1)\times(p+1-r)$ and $(p+1-r)\times r$, respectively, the switching algorithm based on (18)–(19) remains the same, with all parameters and vectors replaced by their starred counterparts. Without proof, we state the limiting distribution of the resulting LR test statistic in the following corollary.

Corollary 2 *In the models (29)–(30), under Assumptions 1–3 and under $\mathcal{H}_n(r, r_1)$, the LR statistic (23) satisfies, as $n \to \infty$,*

$$\begin{aligned} LR_n(r) \quad \xrightarrow{\mathcal{L}} \quad & \left(\int_0^1 Z^*(s)[dW(s) + Z^*(s)' \operatorname{vec}(A^{*\prime})ds]\right)' \left(\int_0^1 Z^*(s)Z^*(s)'ds\right)^{-1} \\ & \times \left(\int_0^1 Z^*(s)[dW(s) + Z^*(s)' \operatorname{vec}(A^{*\prime})ds]\right), \end{aligned} \tag{31}$$

where $A^* = [A : 0]$ and $Z^*(s)$ is defined analogously to (24)–(25), with $K(s)$ replaced by $K^*(s) = (K(s)', 1)'$ in (29), and with $K(s)$ replaced by by $K^*(s) = (K(s)', s)'$ and $Y(s)$ replaced by

$$Y^*(s) = \begin{pmatrix} \alpha'_\perp \\ \alpha' \end{pmatrix} \left\{ \left(\sigma(s)'^{-1} \otimes K^*(s)\right) - \int_0^1 \left(\Sigma(u)^{-1} \otimes K^*(u)\right) du \left[\int_0^1 \Sigma(u)^{-1} du\right]^{-1} \sigma(s)'^{-1} \right\} \tag{32}$$

in (30).

4 Estimation of the volatility process

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 $\{X_t, t = -k+1, \ldots, 0, 1, \ldots, n\}$, 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. 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. Alternatively, a two-step approach may be used, where the volatility matrix is estimated based on the residuals from least-squares estimation of the unrestricted VAR model, and the resulting

estimator $\hat{\Sigma}_t$ is then substituted for Σ_t in the expressions for the MLE and LR statistic given in the previous section.

The obvious disadvantage of such a parametric approach is that it relies on the assumption of correct specification of the volatility process. Therefore, in this paper we propose to estimate σ_t by a nonparametric kernel estimator, generalizing the approach of Boswijk (2005), which in turn is based on Hansen (1995). It should be noted, however, that as analysed by Nelson and Foster (1994) and Nelson (1996a), univariate and multivariate GARCH models (with deterministic parameter sequences instead of estimated parameters) may also be interpreted as nonparametric filters of continuous-time univariate or 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 3, 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 conditional variance matrix estimator at time t based on *leads and lags* of the outer product of the residual vector. As such, the approach is close to that of Foster and Nelson (1996), and hence may be referred to as a volatility *smoother* instead of a filter, see also Nelson (1996b). As discussed in these articles, using leads is not needed when the true volatility follows a GARCH specification, and hence depends on lags of observed variables only. However, when the true volatility process is stochastic, with its own latent shocks, then the use of leads will increase efficiency of the estimator.

Let $\{e_t\}_{t=1}^n$ denote the least-squares residual vector of the model (15) (or of the extended models (29) or (29)), or equivalently the residual vector based on the unrestricted ML estimator (21)–(22) with $\Sigma_t = I_n$. The one-sided kernel estimator proposed by Hansen (1995) is defined in the multivariate context follows. Let $k : [0,1] \to [0,1]$ be a kernel function with $\int_0^1 k(x)dx > 0$, let $N \in \mathbb{N}$ be a window width, and let

$$\hat{\Sigma}_t = \frac{\sum_{j=1}^N k(j/N) e_{t-j} e'_{t-j}}{\sum_{j=1}^N k(j/N)}, \qquad t = N+1, \ldots, n, \tag{33}$$

and $\hat{\Sigma}_t = \hat{\Sigma}_{N+1}$ for $1 \leq t \leq N$. For $t > N$, $\hat{\Sigma}_t$ is a weighted average of N lags of $e_t e'_t$. To avoid boundary effects, the first N observations of $\hat{\Sigma}_t$ (where less than N lags are available) are set equal to $\hat{\Sigma}_{N+1}$. Note that $N > p$ is a necessary condition for positive definiteness of $\hat{\Sigma}_t$ for all t; we shall see that a necessary condition for consistency of the kernel estimator is that the window width N increases with the sample size at a suitable rate.

Alternatively, a double-sided kernel estimator may be defined as

$$\tilde{\Sigma}_t = \frac{\sum_{j=-N}^N k(j/N) 1_{\{1 \leq t-j \leq n\}} e_{t-j} e'_{t-j}}{\sum_{j=-N}^N k(j/N) 1_{\{1 \leq t-j \leq n\}}}, \qquad t = 1, \ldots, n, \tag{34}$$

where $k : [-1,1] \to [0,1]$ now satisfies $\int_{-1}^1 k(x)dx > 0$. For all t, $\tilde{\Sigma}_t$ is a weighted average of leads and lags of $e_t e'_t$, with weights summing to one. For $t \leq N$, the estimator is determined by leads more

than by lags, and for $t > n - N$, the relative weight of the lags increases. This changing shape of the weight function is similar in spirit to the asymmetric beta kernel which was adapted for this purpose and analysed by Kristensen (2006).

An easily interpretable example of the one-sided (33) and the double-sided (34) kernel estimator is obtained for $k(x) = e^{-cx}1_{[0,1]}(x)$ and $k(x) = e^{-c|x|}1_{[-1,1]}(x)$, the truncated exponential and Laplace kernel, where c is chosen such that $k(1) = k(-1) \approx 0$, i.e., such that the truncation has a very small effect. Then the estimators approximately correspond to (single-sided or double-sided) exponential smoothing of $e_t e_t'$, with exponential smoothing parameter $\lambda_N = e^{-c/N}$. In the examples below we use this kernel with $c = 5$, such that $k(1) = k(-1) < 0.01$.

The estimators (33) and (34) correspond to estimators $\hat{\Sigma}_n(s) = \hat{\Sigma}_{\lfloor sn \rfloor}$ and $\tilde{\Sigma}_n(s) = \tilde{\Sigma}_{\lfloor sn \rfloor}$ of $\Sigma(s)$; corresponding estimators of $\sigma(s)$ may be defined as any (not necessarily symmetric) square root of $\hat{\Sigma}_n(s)$ and $\tilde{\Sigma}_n(s)$, but this choice will not affect the results to follow. Theorem 2 establishes consistency of $\hat{\Sigma}_n(s)$ and $\tilde{\Sigma}_n(s)$; a necessary condition is

Assumption 4 *For some $r > 1$, and $i, j = 1, \ldots, p$, $E(|\eta_{i1}\eta_{j1}|^r) < \infty$.*

Clearly this assumption is satisfied if $\eta_1 \sim N(0, I_p)$, but it is sufficient for η_1 to have finite $2r$th moments for some $r > 1$.

Theorem 2 *Consider the model (3) under Assumptions 1–4. If $N = an^b$ for some a and b satisfying $0 < a < \infty$ and $b \in (2/r, 1)$, then*

$$\sup_{s \in [0,1]} \|\hat{\Sigma}_n(s) - \Sigma(s)\| \xrightarrow{P} 0, \tag{35}$$

$$\sup_{s \in [0,1]} \|\tilde{\Sigma}_n(s) - \Sigma(s)\| \xrightarrow{P} 0, \tag{36}$$

where $\|\cdot\|$ is the Euclidean matrix norm.

The consistent estimators $\hat{\Sigma}_n(s)$ and $\tilde{\Sigma}_n(s)$ may be used for an approximate likelihood ratio test as follows. First, we may replace Σ_t by $\hat{\Sigma}_t$ or $\tilde{\Sigma}_t$ in the definition of the restricted and unrestricted estimators (18)–(19) and (21), the corresponding residuals $\hat{\varepsilon}_t$ and $\tilde{\varepsilon}_t$ in (20) and (22), and hence the likelihood ratio statistic (23). Denoting the resulting statistics by $\widehat{LR}_n(r)$ and $\widetilde{LR}_n(r)$, the next theorem establishes that the volatility estimation error has an asymptotically negligible effect.

Theorem 3 *Under the conditions of Theorem 2, $\widehat{LR}_n(r)$ and $\widetilde{LR}_n(r)$ have the same limiting distribution as $LR_n(r)$ under the null hypothesis and local alternatives, as given in Theorem 1.*

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

A second use of the estimators $\hat{\Sigma}_n(s)$ and $\tilde{\Sigma}_n(s)$ is to obtain asymptotic conditional p-values for the test statistics, based on Monte Carlo simulation of the limiting null distributions given in Theorem 1 and

Corollaries 1 and 2, with $\Sigma(s)$ replaced by one of the estimators. The consistency result of Theorem 2 implies that the resulting p-values are consistent as both the sample size and the number of Monte Carlo replications tend to infinity.

We conclude this section with some remarks about the selection of the window width N. Theorem 2 requires N to increase with the sample size at a certain rate, but does not guide us in selecting a window width for a particular sample. A cross-validation technique may be implemented as follows. The one-sided filter $\hat{\Sigma}_t$, which for $t > N$ only depends on $e_{t-j}e'_{t-j}, j > 0$, may be compared with the realisation $e_t e'_t$, leading to a distance function

$$\hat{C}_n(N) = \sum_{t=n_l}^{n} \|\hat{\Sigma}_t(N) - e_t e'_t\|^2, \tag{37}$$

which may be minimized over a suitable range $N \in [N_l, N_u]$, with $N_l > p$ and $N_u < n$, and where $n_l > N_u$ The equivalence between minimizing the sum of squared prediction errors (for exponential smoothing) and cross-validation (for kernel estimation, using a one-sided exponential kernel) was discussed by Gijbels *et al.* (1999). For the two-sided filter, we propose a leave-one-out cross-validation approach, which amounts to minimizing

$$\tilde{C}_n(N) = \sum_{t=n_l}^{n_u} \|\tilde{\Sigma}_t^{-t}(N) - e_t e'_t\|^2, \tag{38}$$

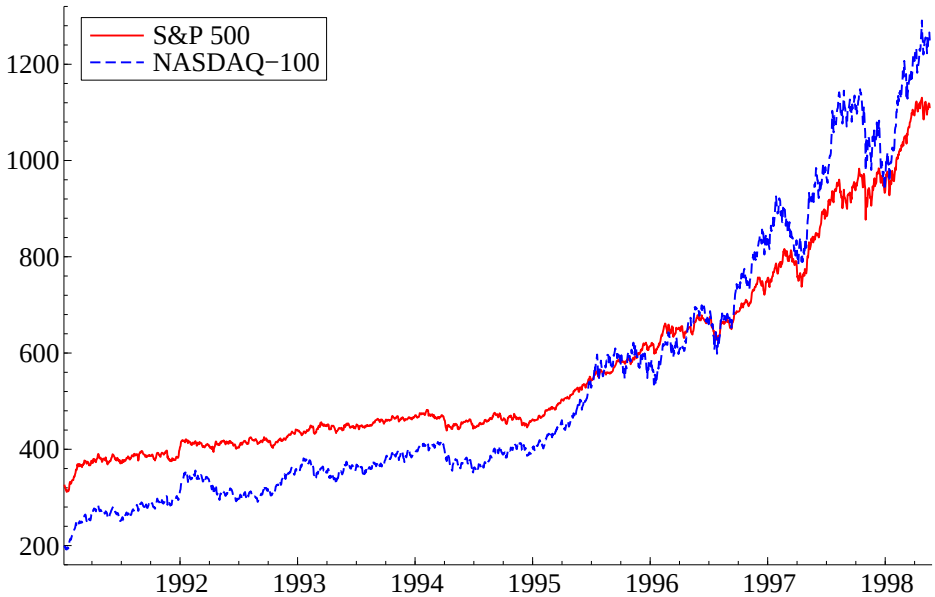
where $\tilde{\Sigma}_t^{-t}(N)$ is given by (34), but with $k(0)$ replaced by 0, such that $e_t e'_t$ does not enter the expression for $\tilde{\Sigma}_t^{-t}(N)$. Here it is advisable to choose $n_l > N_u$ and $n_u < n - N_u$, to avoid boundary effects on the cross-validation procedure.

5 Empirical application: stock market indices in the 1990s

In this section 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 and the test proposed in this paper. More recent evidence, including the stock-market run up and subsequent decline around the turn of the millennium reveals that any linear cointegrating relation between such stock market indices is likely to break down eventually, but 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

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



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.

As indicated in Figure 2, however, the residuals do not appear to be Gaussian white noise: there is evidence of volatility clustering and leptokurtosis, both most prominently in the residuals e_{1t} of the S&P 500 equation.

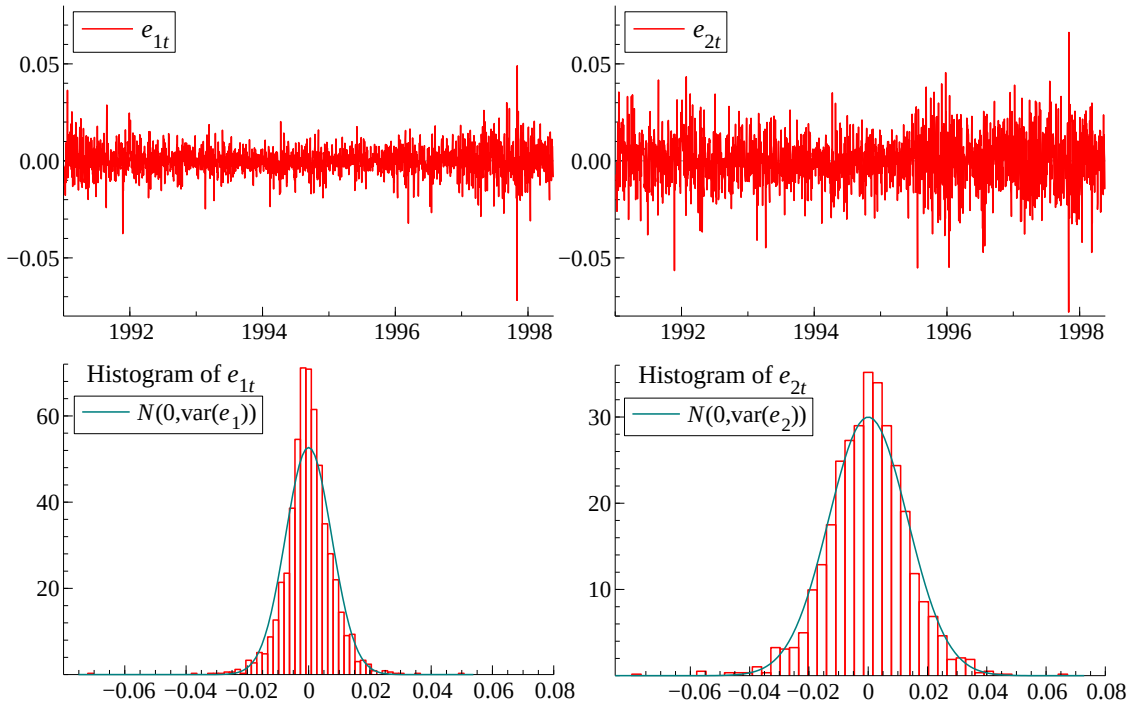
Ignoring these results, and using the Johansen trace test as a quasi-likelihood ratio test for $\mathcal{H}(0)$ leads to a test statistic of 24.37, with a (simulated) p-value of 0.075. Therefore, we reject the null hypothesis of no cointegration only if we are prepared to use a significance level of 10% or higher. The estimated cointegrating relation is (standard errors in parentheses)

$$\hat{\beta}' X_t^* = X_{1t} + \underset{(0.677)}{1.976} \times 10^{-4} t - \underset{(0.079)}{0.922} X_{2t}, \tag{39}$$

which includes a linear trend coefficient that is significantly different from zero at the usual significance level. The estimated error correction coefficients are $\hat{\alpha}_1 = -0.001$ (0.004) and $\hat{\alpha}_2 = 0.020$ (0.007), suggesting that most of the error correction is done by the NASDAQ.

Figure 3 shows the estimated volatilities and the covariance and correlation based on $\{\hat{\Sigma}_t\}$ and $\{\tilde{\Sigma}_t\}$, the nonparametrically estimated conditional covariance matrix given in (33) and (34). These figures are based on the truncated exponential / Laplace kernel $k(x) = e^{-5|x|} 1_{[-1,1]}(x)$, with N chosen to minimize the cross-validation criterion functions (37) and (38), where we have taken $N_l = 20$, $N_u = 250$ (corresponding to a range of approximately $\lambda_N \in [0.8, 0.98]$ for the exponential smoothing

Figure 2: Least-squares residuals from VAR(1) model with trend, and their histograms.



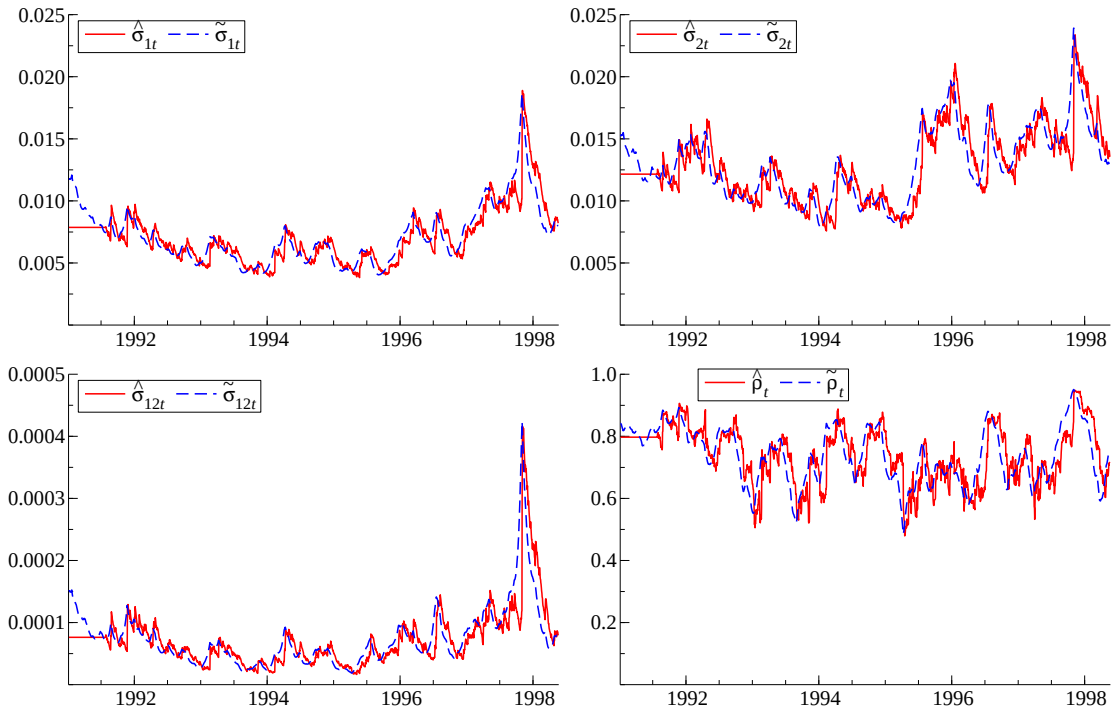
parameter), and $n_l = N_u + 1$ and $n_u = n - N_u$ The chosen window widths are given by $\hat{N} = \text{argmin}_{N\in[N_l,N_u]} \hat{C}_n(N) = 140$ (such that $\lambda_{\hat{N}} = 0.965$) and $\tilde{N} = \text{argmin}_{N\in[N_l,N_u]} \tilde{C}_n(N) = 80$ (corresponding to $\lambda_{\tilde{N}} = 0.939$).

From Figure 3 we see that both methods lead to similar estimates. As expected, the double-sided kernel estimates have smoother sample paths than the one-sided filters, with the latter lagging behind the former in terms of up- or downswings. Also, we note the familiar time-irreversible patterns around the peaks in the one-sided estimates, with sudden changes before the peak and gradual mean-reversion after the peak, whereas the double-sided estimates, by construction, lead to more symmetric patterns around the peaks. 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 much lower mean-reversion. Therefore it seems reasonable to apply the type of asymptotics implied by Assumption 3.

Before the adaptive test defined in the previous section is applied, we first simulate the asymptotic null distribution of the trace statistic based on (14), with $\Sigma(\cdot)$ replaced by $\hat{\Sigma}_n(\cdot)$ and $\tilde{\Sigma}_n(\cdot)$. The resulting p-values are 0.087 and 0.122, as opposed to the original p-value of 0.075 that is valid under the assumption of stationary volatility. This illustrates that the type of nonstationary volatility observed in financial data may have a serious consequence for the size of conventional tests.

The adaptive likelihood ratio statistic for no cointegration based on $\{\hat{\Sigma}_t\}$ has a value of 29.60, with an asymptotic p-value of 0.011; when using $\{\tilde{\Sigma}_t\}$, the test statistic and p-value become 26.66 [0.035]. Therefore, we see that the evidence in favour of cointegration has increased substantially by taking the time-varying volatility matrix into account. The resulting cointegrating relation, based on $\{\tilde{\Sigma}_t\}$ now

Figure 3: Estimated volatilities, covariance and correlation of e_{1t} and e_{2t}.



becomes

$$\hat{\beta}' X_t^* = X_{1t} + \underset{(0.672)}{1.773} \times 10^{-4} t - \underset{(0.082)}{0.903} X_{2t}, \tag{40}$$

which is quite similar to the relation found from the Gaussian constant-volatility quasi-likelihood (as well as the estimates based on $\{\hat{\Sigma}_t\}$). Similarly, the adjustment coefficients $\hat{\alpha}_1 = -0.001$ (0.004) and $\hat{\alpha}_2 = 0.018$ (0.007) are hardly affected. Note, however, that the standard errors reported in (39) and for the corresponding estimates of α are invalid in the presence of nonstationary volatility.

In summary, the empirical example in this section has illustrated 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 Discussion

Although we have used a stock market example to illustrate the approach proposed in this paper, one may argue that there is little theory to support the idea that two stock market indices should be cointegrated, and indeed such relations typically break down over longer sample periods. However, the scope for possible applications of the proposed test is much wider, including interest rate (term structure) models and exchange rate models (spot-forward relations, purchasing power parity). In principle the approach could also be used for 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 is probably advisable

to adopt a parametric model for the changing volatilities and correlations.

The assumptions made in the paper appear to exclude some, potentially relevant, cases. First, the assumed continuity of the volatility matrix $\sigma(\cdot)$ excludes level shifts in volatility or correlation. In principle, such discontinuous level shifts can be approximated arbitrarily well by continuous smooth transition functions with a very steep transition. However, the effectiveness of the nonparametric volatility filter around such events needs to be investigated further.

Another limitation is the assumed independence between the volatility process and the Brownian motion corresponding to the partial sum of standardized innovations. This independence is needed for the type of conditional inference considered here, but it excludes applications in stock markets where volatility shocks and stock price shocks are negatively correlated, corresponding to the well-known leverage effect. It will be hard to avoid this condition without formulating a fully parametric model to make the dependence explicit.

Finally, we have considered here tests based on a conditionally Gaussian likelihood function. In practice one often observes that financial returns standardized by GARCH-type volatility estimators display excess kurtosis. Note, however, that the asymptotic results in this paper are robust to this type of nonnormality, as long as the conditions of Assumption 3 and 4 are satisfied. Clearly, in such cases the proposed test loses its efficiency claims, and more powerful tests could be derived from, e.g., a Student's t likelihood. Note also that the deviations from normality of standardized residuals may be less pronounced when based on a double-sided filter, analogous to the corresponding result for residuals based on realized volatility, see Andersen *et al*. (2003).

Appendix

Proof of Lemma 1. Consider first (10), and note that (4), (6) and (7) together imply

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

Using a multivariate version of Hansen (1992)'s Theorem 2.1, it follows that $\int_0^s \sigma_n(u)dW_n(u) \xrightarrow{\mathcal{L}} \int_0^s \sigma(u)dW(u)$ in $D[0,1]^p$, jointly with (8), because $\{(\sigma_{t+1},\eta_t)\}_{t\geq 1}$ is adapted to $\{\mathcal{F}_t\}_{t\geq 1}$, and $\{\eta_t\}_{t\geq 1}$ is a martingale difference sequence with respect to $\{\mathcal{F}_t\}_{t\geq 0}$, with $\sup_n n^{-1}\sum_{t=1}^n E(\|\eta_t\eta_t'\|) < \infty$.

The proof of (11) 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 (1995), and the corresponding solution to Exercise 14.1 given in Hansen and Johansen (1998). Consider the model in companion form, for the stacked vector $\vec{X}_t = (X_t', \ldots, X_{t-k+1}')'$:

$$\vec{X}_t = \left(I_{kp} + \vec{\alpha}\vec{\beta}' + n^{-1}\vec{\alpha}_1\vec{\beta}_1'\right)\vec{X}_{t-1} + \vec{\varepsilon}_t, \tag{A.2}$$

where

$$\vec{\alpha} = \begin{bmatrix} \alpha & \Gamma_1 & \cdots & \Gamma_{k-1} \\ 0 & I_p & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I_p \end{bmatrix}, \qquad \vec{\beta} = \begin{bmatrix} \beta & I_p & 0 & 0 \\ 0 & -I_p & \ddots & 0 \\ \vdots & \ddots & \ddots & I_p \\ 0 & \cdots & 0 & -I_p \end{bmatrix}, \tag{A.3}$$

and where $\vec{\alpha}_1 = (\alpha_1', 0_{(k-1)p}')'$, $\vec{\beta}_1 = (\beta_1', 0_{(k-1)p}')'$, and $\vec{\varepsilon}_t = (\varepsilon_t', 0_{(k-1)p}')'$. Note that $\vec{\alpha}_\perp' = \alpha_\perp'[I_p : -\Gamma_1 : \cdots : -\Gamma_{k-1}]$ and $\vec{\beta}_\perp' = \beta_\perp'[I_p : I_p : \cdots : I_p]$, and hence

$$\vec{\alpha}_\perp'\vec{\alpha}_1 = \alpha_\perp'\alpha_1, \qquad \vec{\beta}_1'\vec{\beta}_\perp = \beta_1'\beta_\perp, \qquad \vec{\alpha}_\perp'\vec{\beta}_\perp = \alpha_\perp'\Gamma\beta_\perp. \tag{A.4}$$

Let $\vec{A}_n = I_{kp} + \vec{\alpha}\vec{\beta}' + n^{-1}\vec{\alpha}_1\vec{\beta}_1'$. Backward substitution in (A.2) gives the solution

$$\vec{X}_t = \vec{A}_n^t\vec{X}_0 + \sum_{j=1}^t \vec{A}_n^{t-j}\vec{\varepsilon}_j. \tag{A.5}$$

It will be convenient to work with the decomposition $\vec{X}_t = \vec{\alpha}(\vec{\beta}'\vec{\alpha})^{-1}\vec{\beta}'\vec{X}_t + \vec{\beta}_\perp(\vec{\alpha}_\perp'\vec{\beta}_\perp)^{-1}\vec{\alpha}_\perp'\vec{X}_t$; we consider the behaviour of $\vec{\beta}'\vec{X}_t$ (the stationary linear combinations) and $\vec{\alpha}_\perp'\vec{X}_t$ (the non-stationary linear combinations) separately. Assumption 1 requires that all eigenvalues of $I_{r+(k-1)p} + \vec{\beta}'\vec{\alpha}$ are less than one in absolute value, and this implies for the stationary linear combinations, by Theorem 14.1 of Johansen (1995),

$$\vec{\beta}'\vec{X}_t = (I_{r+(k-1)p} + \vec{\beta}'\vec{\alpha})^t\vec{\beta}'\vec{X}_0 + \sum_{j=1}^t (I_{r+(k-1)p} + \vec{\beta}'\vec{\alpha})^{t-j}\vec{\beta}'\vec{\varepsilon}_j + R_{\vec{\beta}t}, \tag{A.6}$$

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

$$\vec{\alpha}_\perp' \vec{X}_t = \vec{\alpha}_\perp' \vec{A}_n^t \vec{X}_0 + \sum_{j=1}^{t} \vec{\alpha}_\perp' \vec{A}_n^{t-j} \vec{\varepsilon}_j. \tag{A.7}$$

Writing $\vec{\varepsilon}_t = E\varepsilon_t = E\sigma_t\eta_t$, with $E = [I_p : 0_{p\times(k-1)p}]'$, and defining $F_n(s) = \vec{A}_n^{\lfloor sn \rfloor}$, we therefore find

$$n^{-1/2}\vec{\alpha}_\perp' \vec{X}_{\lfloor sn \rfloor} = n^{-1/2}\vec{\alpha}_\perp' F_n(s)\vec{X}_0 + \vec{\alpha}_\perp' F_n(s) \int_0^s F_n(-u)E\sigma_n(u)dW_n(u). \tag{A.8}$$

Theorem A.14 of Johansen (1995) implies $F_n(s) \to \vec{\beta}_\perp^{-1}(\vec{\alpha}_\perp'\vec{\beta}_\perp)^{-1}\exp(sA)\vec{\alpha}_\perp'$, and hence

$$F_n(-s)E\sigma_n(s) \xrightarrow{\mathcal{L}} \vec{\beta}_\perp(\vec{\alpha}_\perp'\vec{\beta}_\perp)^{-1}\exp(-sA)\sigma_K(s), \qquad s \in [0,1], \tag{A.9}$$

jointly with $W_n(\cdot) \xrightarrow{\mathcal{L}} W(\cdot)$. Because η_t is a martingale difference sequence with constant variance, Hansen (1992)'s Theorem 2.1 again implies

$$n^{-1/2}\vec{\alpha}_\perp' \vec{X}_{\lfloor sn \rfloor} \xrightarrow{\mathcal{L}} \int_0^s \exp\left((s-u)A\right)\sigma_K(u)dW(u) = K(s), \qquad s \in [0,1]. \tag{A.10}$$

The stochastic differential equation (13) follows from Itô's formula, writing $K(s) = \exp(sA)M(s) = f(s, M(s))$, with $dM(s) = \exp(-sA)\sigma_K(s)dW(s)$. Finally,

$$\begin{aligned} n^{-1/2}X_{\lfloor sn \rfloor} &= n^{-1/2}E'\vec{\beta}_\perp(\vec{\alpha}_\perp'\vec{\beta}_\perp)^{-1}\vec{\alpha}_\perp'\vec{X}_{\lfloor sn \rfloor} + o_P(1) \\ &= \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}n^{-1/2}\vec{\alpha}_\perp'\vec{X}_{\lfloor sn \rfloor} \\ &\xrightarrow{\mathcal{L}} \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}K(s), \qquad s \in [0,1], \end{aligned} \tag{A.11}$$

which concludes the proof of (11). □

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.,

$$\Lambda_n(\theta) = -2\left[\ell_n(\Pi, I_{p+1}, \Psi) - \ell_n(\tilde{\Pi}_n, I_{p+1}, \tilde{\Psi}_n)\right] = \sum_{t=1}^{n}\left(\varepsilon_t(\theta)'\Sigma_t^{-1}\varepsilon_t(\theta) - \tilde{\varepsilon}_t'\Sigma_t^{-1}\tilde{\varepsilon}_t\right). \tag{A.12}$$

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

$$\Lambda_n(\theta) = (\tilde{\theta}_n - \theta)' \sum_{t=1}^{n}[Z_t Z_t' \otimes \Sigma_t^{-1}](\tilde{\theta}_n - \theta). \tag{A.13}$$

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

$$\Theta_r = \{\theta \in \mathbb{R}^{kp^2} : \Pi = \alpha\beta'; \quad (\alpha, \beta, \Psi) \in \mathbb{R}^{p\times r} \times \mathbb{R}^{p\times r} \times \mathbb{R}^{p\times(k-1)p}\}. \tag{A.14}$$

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 measure $P_{\theta_0}^n$ 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 $D_n^{-1}(\tilde{\theta}_n - \theta_0)$ converge in distribution, and a choice of D_n that satisfies this requirement is

$$D_n = \begin{bmatrix} n^{-1}\Gamma'\alpha_\perp & n^{-1/2}\beta & 0 \\ 0 & 0 & n^{-1/2} I_{k(p-1)} \end{bmatrix} \otimes I_p, \tag{A.15}$$

such that

$$D_n'[Z_t \otimes I_p] = n^{-1/2} \begin{pmatrix} n^{-1/2}\alpha_\perp' \Gamma X_{t-1} \\ Z_t(\beta) \end{pmatrix} \otimes I_p. \tag{A.16}$$

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 $\operatorname{rank}(\alpha_\perp' \Gamma \beta_\perp) = p - r$ (Assumption 1 (c)) implies that the matrix $[\beta : \Gamma'\alpha_\perp]$ has full rank. A projection of β_1 on $\operatorname{sp}(\beta : \Gamma'\alpha_\perp)$ 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 $\operatorname{sp}(\beta)$ 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

$$\theta_n = \operatorname{vec}[\Pi_n : \Psi_0] = \operatorname{vec}[\Pi_0 : \Psi_0] + \operatorname{vec}\left[n^{-1}\alpha_1\beta_1' : 0\right] = \theta_0 + D_n \check{\tau}, \tag{A.17}$$

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

Lemma 1 implies that both under the null and under local alternatives, $n^{-1/2}\alpha_\perp'\Gamma X_{\lfloor sn \rfloor} \xrightarrow{\mathcal{L}} K(s)$, $s \in [0,1]$. 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 (1995) 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,

$$\begin{aligned} D_n' \sum_{t=1}^n [Z_t Z_t' \otimes \Sigma_t^{-1}] D_n &\xrightarrow{\mathcal{L}} \begin{bmatrix} \int_0^1 \left[K(s)K(s)' \otimes \Sigma(s)^{-1}\right] ds & 0 \\ 0 & \int_0^1 \left[\Omega(s) \otimes \Sigma(s)^{-1}\right] ds \end{bmatrix} \\ &=: \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} = J, \end{aligned} \tag{A.18}$$

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

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

$$D_n^{-1}(\tilde{\theta}_n - \theta_0) = \left(D_n' \sum_{t=1}^n [Z_t Z_t' \otimes \Sigma_t^{-1}] D_n \right)^{-1} D_n' \sum_{t=1}^n [Z_t \otimes \Sigma_t^{-1}] \varepsilon_t(\theta_0), \tag{A.19}$$

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 generalizing the results of Hansen (1995) and Phillips and Xu

(2006),

$$D_n' \sum_{t=1}^{n} [Z_t \otimes \Sigma_t^{-1}] \varepsilon_t(\theta_0) \xrightarrow{\mathcal{L}} \begin{pmatrix} \int_0^1 \left[K(s) \otimes \Sigma(s)^{-1}\right] (dU(s) + [K(s)' ds \otimes I_p] \check{\tau}_1) \\ \sum_{j=1}^{\infty} \int_0^1 [C_j \sigma(s) \otimes \sigma(s)'^{-1}] dB_j(s) \end{pmatrix}$$
$$=: \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = S, \tag{A.20}$$

where $\check{\tau}_1 = \mathrm{vec}(\alpha_1 b')$, a $p(p-r)$-vector consisting of the first rows of the value of $\check{\tau}$ in (A.17), 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 $\mathrm{vec}(\eta_t \eta_{t-j}')$. If $\sigma(\cdot)$ is non-stochastic, then the bottom right-hand side expression of (A.20) has a normal distribution with mean zero and variance $\int_0^1 \left[\Omega(s) \otimes \Sigma(s)^{-1}\right] ds$, and if $\sigma(\cdot)$ is stochastic but independent of $\{B_j(\cdot)\}_{j=1}^{\infty}$, then the same normal distribution holds conditionally on $\sigma(\cdot)$. Note that under the null, $\check{\tau}_1 = 0$ and $K(s) = \alpha_\perp' U(s)$ in (A.18) and (A.20).

The representation (A.19), together with the limit results (A.18) and (A.20), implies that $D_n^{-1}(\tilde{\theta}_n - \theta_0) = \mathrm{O}_P(1)$ under both the null and local alternatives, such that $\tilde{\theta}_n$ is consistent under both hypotheses. Generalizing the argument explained fully in Theorem A1 of Johansen (1997), this implies that the restricted MLE $\hat{\theta}_n$ is also consistent, and $D_n^{-1}(\hat{\theta}_n - \theta_0) = \mathrm{O}_P(1)$. This implies that for the derivation of the limiting distribution of the LR statistic $LR_n = \min_{\theta \in \Theta_r} \Lambda_n(\theta)$, 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 $LR_n = \min_{\tau \in \mathcal{T}_{r,n}} \Lambda_n(\theta_0 + D_n \tau)$. We will show that $\Lambda_n(\theta_0 + D_n \tau) \xrightarrow{\mathcal{L}} \Lambda(\tau)$ uniformly on compact sets, and that the restricted parameter space $\mathcal{T}_{r,n}$ converges to a limit $\mathcal{T}_r$. Because $\hat{\tau}_n = D_n^{-1}(\hat{\theta}_n - \theta_0)$ is $\mathrm{O}_P(1)$, this will then imply, by the argmax theorem (Van der Vaart, 1998, Corollary 5.58), $D_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \arg\min_{\tau \in \mathcal{T}_r} \Lambda(\tau)$ and $LR_n \xrightarrow{\mathcal{L}} \min_{\tau \in \mathcal{T}_r} \Lambda(\tau)$.

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

$$\Lambda_n(\theta_0 + D_n \tau) = \left(D_n^{-1}(\tilde{\theta}_n - \theta_0) - \tau\right)' D_n' \sum_{t=1}^{n} [Z_t Z_t' \otimes \Sigma_t^{-1}] D_n \left(D_n^{-1}(\tilde{\theta}_n - \theta_0) - \tau\right), \tag{A.21}$$

which by (A.18) and (A.19) 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

$$\begin{aligned} \mathrm{vec}(\Pi_n - \Pi_0) &= \mathrm{vec}\left(\alpha_n[\bar{c}' + \Phi_n' c_\perp'] - \alpha[\bar{c}' + \Phi' c_\perp']\right) \\ &= \mathrm{vec}\left([\alpha_n - \alpha]\beta' + \alpha[\Phi_n - \Phi]' c_\perp' + [\alpha_n - \alpha][\Phi_n - \Phi]' c_\perp'\right) \\ &= [\beta \otimes I_p] \,\mathrm{vec}(\alpha_n - \alpha) + [c_\perp \otimes I_p][I_{p-r} \otimes \alpha] \,\mathrm{vec}(\Phi_n' - \Phi') + R_n, \end{aligned} \tag{A.22}$$

where the remainder term R_n is $\mathrm{O}(\|\alpha_n - \alpha\| \, \|\Phi_n - \Phi\|)$, 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 $\mathrm{vec}(\Psi_n - \Psi_0) = n^{-1} \kappa_\psi$,

$\mathrm{vec}(\alpha_n - \alpha) = n^{-1/2}\kappa_\alpha$ and $\mathrm{vec}(\Phi_n' - \Phi') = n^{-1}\kappa_\phi$, then in the restricted parameter space Θ_r, we have

$$\theta_n = \theta_0 + D_n \begin{pmatrix} [I_q \otimes \alpha]\kappa_\phi \\ \kappa_\alpha \\ \kappa_\psi \end{pmatrix} + \begin{pmatrix} R_n \\ 0 \end{pmatrix}, \tag{A.23}$$

such that

$$\tau = D_n^{-1}(\theta_n - \theta_0) = \begin{bmatrix} I_q \otimes \alpha & 0 & 0 \\ 0 & I_r \otimes I_p & 0 \\ 0 & 0 & I_{(k-1)p} \otimes I_p \end{bmatrix} \begin{pmatrix} \kappa_\phi \\ \kappa_\alpha \\ \kappa_\psi \end{pmatrix} + \mathrm{o}(1) = G\kappa + \mathrm{o}(1). \tag{A.24}$$

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

$$\begin{aligned} LR_n \xrightarrow{\mathcal{L}} \min_{\tau \in \mathcal{T}_r} \Lambda(\tau) &= S'J^{-1}S - S'G(G'JG)^{-1}G'S \\ &= S'J^{-1}G_\perp(G_\perp' J^{-1} G_\perp)^{-1} G_\perp' J^{-1} S, \end{aligned} \tag{A.25}$$

where

$$G_\perp = \begin{bmatrix} I_q \otimes \alpha_\perp \\ 0 \end{bmatrix}. \tag{A.26}$$

It follows that $LR_n \xrightarrow{\mathcal{L}} S_1' J_1^{-1} G_{\perp 1}(G_{\perp 1}' J_1^{-1} G_{\perp 1})^{-1} G_{\perp 1}' J_1^{-1} S_1$, 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

$$G_{\perp 1} = I_q \otimes \alpha_\perp = K_{qp}(\alpha_\perp \otimes I_q)K_{qq} = K_{qp}\left([I_q : 0] \otimes I_q\right)\left([\alpha_\perp : \bar{\alpha}] \otimes I_q\right) K_{qq}. \tag{A.27}$$

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

$$\begin{aligned} G_{\perp 1}' J_1^{-1} G_{\perp 1} &= K_{qp}\left([I_q : 0] \otimes I_q\right)\left(\int_0^1 Y(s)Y(s)'ds\right)^{-1}\left([I_q : 0]' \otimes I_q\right) K_{pq} \\ &= K_{qp}\left(\int_0^1 Z(s)Z(s)'ds\right)^{-1} K_{pq}. \end{aligned} \tag{A.28}$$

Similarly, using

$$\begin{aligned} S_1 &= K_{qp}\int_0^1 \left[\sigma(s)'^{-1} \otimes K(s)\right]\left(dW(s) + [\sigma(s)^{-1} \otimes K(s)']ds K_{pq}\check{\tau}_1\right) \\ &= K_{qp}\left([\alpha_\perp : \bar{\alpha}] \otimes I_q\right)\int_0^1 Y(s)\left(dW(s) + Y(s)'ds\left([\alpha_\perp : \bar{\alpha}]' \otimes I_q\right) K_{pq}\check{\tau}_1\right) \\ &= K_{qp}\left([\alpha_\perp : \bar{\alpha}] \otimes I_q\right)\int_0^1 Y(s)\left(dW(s) + Y(s)'ds\,\mathrm{vec}(A' : b\alpha_1'\bar{\alpha})\right), \end{aligned} \tag{A.29}$$

we obtain

$$\begin{aligned} G'_{\perp 1} J_1^{-1} S_1 &= K_{qp} \left([I_q : 0] \otimes I_q\right) \left(\int_0^1 Y(s)Y(s)'ds\right)^{-1} \\ &\quad \times \int_0^1 Y(s)[dW(s) + Y(s)'ds \operatorname{vec}(A' : b\alpha_1'\bar{\alpha})] \\ &= K_{qp} \left(\int_0^1 Z(s)Z(s)'ds\right)^{-1} \int_0^1 Z(s)[dW(s) + Z(s)' \operatorname{vec}(A')ds]. \end{aligned} \tag{A.30}$$

This leads to the required result. □

Proof of Theorem 2. The proof is based on Theorem 2 in Hansen (1995), which is generalized and modified in three directions: (i) a multivariate instead of univariate volatility estimator, (ii) residuals from a regression with $I(1)$ instead of $I(0)$ regressors, and (iii) a two-sided instead of one-sided kernel. We first prove consistency of the one-sided kernel estimator $\hat{\Sigma}_n(s)$, following the steps of the proof of Hansen (1995)'s Theorem 2, and next we discuss the modifications needed to prove consistency of the two-sided kernel estimator $\tilde{\Sigma}_n(s)$.

Because of the definition of $\hat{\Sigma}_n(s)$, the required result (35) is equivalent to $\max_{1\leq t\leq n} \|\hat{\Sigma}_t - \Sigma_t\| \xrightarrow{P} 0$. We first consider the difference for $t = N+1, \ldots, n$, where $\hat{\Sigma}_t = \sum_{j=1}^N w_{jN} e_{t-j} e'_{t-j}$, with weights $w_{jN} = k(j/N)/\left(\sum_{j=1}^N k(j/N)\right)$ summing to 1 by construction. It is convenient to decompose the difference $\hat{\Sigma}_t - \Sigma_t$ into four terms, as follows:

$$\hat{\Sigma}_t - \Sigma_t = R_t^a + \sigma_t R_t^b \sigma_t' + R_t^c + R_t^d, \tag{A.31}$$

where

$$R_t^a = \sum_{j=1}^N w_{jN}(\Sigma_{t-j} - \Sigma_t), \tag{A.32}$$

$$R_t^b = \sum_{j=1}^N w_{jN}(\eta_{t-j}\eta'_{t-j} - I_p), \tag{A.33}$$

$$R_t^c = \sum_{j=1}^N w_{jN} \left[\sigma_{t-j}(\eta_{t-j}\eta'_{t-j})\sigma'_{t-j} - \sigma_t(\eta_{t-j}\eta'_{t-j})\sigma'_t\right] \tag{A.34}$$

$$R_t^d = \sum_{j=1}^N w_{jN}(e_{t-j}e'_{t-j} - \varepsilon_{t-j}\varepsilon'_{t-j}). \tag{A.35}$$

This decomposition is a multivariate generalization of the one given in Hansen (1995), equation (36). The asymptotic negligibility of R_t^a and R_t^c relates to continuity of Σ_t, and R_t^d refers to the estimation error in the residuals. The essential part is R_t^b, because its asymptotic negligibility shows that local averaging of $\varepsilon_{t-j}\varepsilon'_{t-j}$ leads to a consistent estimator of the corresponding weighted average of Σ_{t-j}, and hence (via continuity) of Σ_t. We will show that the maximum over $t = N+1, \ldots, n$ of each of these terms will converge in probability (in Euclidean norm) to zero, such that $\max_{N<t\leq n} \|\hat{\Sigma}_t - \Sigma_t\| \xrightarrow{P} 0$ (using the triangle inequality).

For the first term, the proof of Hansen (1995) is directly generalized. The property $\sum_{j=1}^{N} w_{jN} = 1$ implies that

$$\max_{N<t\leq n} \|R_t^a\| \leq \max_{N<t\leq n, 1\leq j\leq N} \|\Sigma_{t-j} - \Sigma_t\|, \tag{A.36}$$

and the right-hand side converges in probability to zero because $N = \mathrm{o}(n)$, and Assumption 3 implies $\Sigma_n(s) \xrightarrow{\mathcal{L}} \Sigma(s), s \in [0,1]$ (which in turn implies tightness of the sequence $\{\Sigma_n\}$), and continuity the sample paths of $\Sigma(s)$; see Hansen (1995), Lemma A.1.

For the second term, note that $\{\eta_t\eta_t' - I_p\}$ is an i.i.d. sequence with mean zero and finite rth moment. Again following the argument in Hansen (1995) this implies that $\max_{N<t<n} \|R_t^b\| = \mathrm{o}_P(1)$. And because $\max_{N<t\leq n} \|\sigma_t\| = \mathrm{O}_P(1)$, this in turn implies

$$\max_{N<t\leq n} \|\sigma_t R_t^b \sigma_t'\| \leq \max_{N<t\leq n} \|\sigma_t\|^2 \times \max_{N<t\leq n} \|R_t^b\| = \mathrm{o}_P(1). \tag{A.37}$$

For the third term, letting $H_{jt} = \eta_{t-j}\eta_{t-j}'$, we use

$$R_t^c = \sum_{j=1}^{N} w_{jN} \left[(\sigma_{t-j} - \sigma_t)H_{jt}\sigma_t' + \sigma_t H_{jt}(\sigma_{t-j} - \sigma_t)' + (\sigma_{t-j} - \sigma_t)H_{jt}(\sigma_{t-j} - \sigma_t)'\right], \tag{A.38}$$

such that

$$\begin{aligned} \max_{N<t\leq n} \|R_t^c\| \quad \leq \quad & 2 \max_{N<t\leq n} \|\sigma_t\| \times \max_{N<t\leq n, 1\leq j\leq N} \|\sigma_{t-j} - \sigma_t\| \times \max_{N<t\leq n} \sum_{j=1}^{N} w_{jN} \|H_{jt}\| \\ & + \max_{N<t\leq n, 1\leq j\leq N} \|\sigma_{t-j} - \sigma_t\|^2 \times \max_{N<t\leq n} \sum_{j=1}^{N} w_{jN} \|H_{jt}\|. \end{aligned} \tag{A.39}$$

Because $\{\|H_{jt}\| - p = \eta_{t-j}'\eta_{t-j} - p\}$ is an i.i.d. sequence with mean 0 and finite rth moment,

$$\max_{N<t\leq n} \sum_{j=1}^{N} w_{jN} \|H_{jt}\| = p + \max_{N<t\leq n} \sum_{j=1}^{N} w_{jN}(\eta_{t-j}'\eta_{t-j} - p) = p + \mathrm{o}_P(1) = \mathrm{O}_P(1), \tag{A.40}$$

where the $\mathrm{o}_P(1)$ term is analogous to $\max_{N<t\leq n} \|R_t^b\|$. Next, $\max_{N<t\leq n} \|\sigma_t\| = \mathrm{O}_P(1)$ because of tightness of $\{\Sigma_n\}$, and $\max_{N<t\leq n, 1\leq j\leq N} \|\sigma_{t-j} - \sigma_t\| = \mathrm{o}_P(1)$ analogous to $\max_{N<t\leq n} \|R_t^a\|$. Therefore, $\max_{N<t\leq n} \|R_t^c\| = \mathrm{o}_P(1)$.

For the fourth term, we write the unrestricted model as $\Delta X_t = BZ_t + \varepsilon_t$, where in the notation of the proof of Theorem 1, $\theta = \mathrm{vec}\, B$. 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.15) 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^d:

$$\begin{aligned} R_t^d \quad = \quad & -\left(\sum_{j=1}^{N} w_{jN}\varepsilon_{t-j}Z_{t-j}'\bar{D}_n\right)\bar{D}_n^{-1}(\bar{B}_n - B)' - (\bar{B}_n - B)\bar{D}_n'^{-1}\left(\sum_{j=1}^{N} w_{jN}\bar{D}_n' Z_{t-j}\varepsilon_{t-j}'\right) \\ & + (\bar{B}_n - B)\bar{D}_n'^{-1}\left(\sum_{j=1}^{N} w_{jN}\bar{D}_n' Z_{t-j}Z_{t-j}'\bar{D}_n\right)\bar{D}_n^{-1}(\bar{B}_n - B)'. \end{aligned} \tag{A.41}$$

Because of the convergence of the sample moments mentioned above, $\sum_{j=1}^{N} w_{jN}\bar{D}_n' Z_{t-j}\varepsilon_{t-j}' = \mathrm{o}_P(1)$ and $\sum_{j=1}^{N} w_{jN}\bar{D}_n' Z_{t-j}Z_{t-j}'\bar{D}_n = O_P(N/n) = \mathrm{o}_P(1)$. This in turn implies that $\max_{N<t\leq n}\left\|R_t^d\right\| = \mathrm{o}_P(1)$.

The steps above together imply $\max_{N<t\leq n}\|\hat{\Sigma}_t - \Sigma_t\| \xrightarrow{P} 0$. For $\{\hat{\Sigma}_t\}_{t=1}^{N}$, we use

$$\begin{aligned}\max_{1\leq t\leq N}\|\hat{\Sigma}_t - \Sigma_t\| &= \max_{1\leq t\leq N}\|\hat{\Sigma}_{N+1} - \Sigma_t\| \\ &\leq \|\hat{\Sigma}_{N+1} - \Sigma_{N+1}\| + \max_{1\leq t\leq N}\|\Sigma_t - \Sigma_{N+1}\| \xrightarrow{P} 0, \end{aligned} \tag{A.42}$$

using (A.36) and consistency of $\hat{\Sigma}_{N+1}$ proved earlier.

Now consider the double-sided kernel estimator $\tilde{\Sigma}_t$. For $N < t \leq n - N$, $\tilde{\Sigma}_t - \Sigma_t$ may be decomposed into four terms entirely analogous to (A.31), but with $\sum_{j=1}^{N} w_{jN}$ replaced by $\sum_{j=-N}^{N}\tilde{w}_{jN}$, where $\tilde{w}_{jN} = k(j/N)/\left(\sum_{j=-N}^{N} k(j/N)\right)$ again sum to 1. Asymptotic negligibility of the maximum over $t = N+1,\ldots,n-N$ of each of the four terms follows from the same arguments, after making the necessary changes in notation. For the observations $t \leq N$ and $t > n - N$, these arguments remain the same, but in such cases $\tilde{w}_{jN}$ should be replaced by

$$\tilde{w}_{jNt} = \frac{k(j/N)1_{\{1\leq t-j\leq n\}}}{\sum_{j=-N}^{N} k(j/N)1_{\{1\leq t-j\leq n\}}}. \tag{A.43}$$

Therefore, $\max_{1\leq t\leq N}\|\tilde{\Sigma}_t - \Sigma_t\| \xrightarrow{P} 0$ as required. □

Proof of Theorem 3. We will show that the convergence results (A.18) and (A.20) still apply when Σ_t is replaced by either $\hat{\Sigma}_t$ or $\tilde{\Sigma}_t$. Combining these results in the unrestricted log-likelihood ratio (A.13), using (A.19), will imply that $\hat{\Lambda}_n(\theta_0 + D_n\tau)$ and $\tilde{\Lambda}_n(\theta_0 + D_n\tau)$ (based on $\hat{\Sigma}_t$ and $\tilde{\Sigma}_t$, respectively) both converge 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.

We will first consider the results for the one-sided filter $\hat{\Sigma}_t$, and then discuss how the results are modified for the two-sided estimator $\tilde{\Sigma}_t$. Define

$$V_{t-1} := \alpha_\perp'\Gamma X_{t-1} \otimes \sigma_t'^{-1}, \tag{A.44}$$

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

$$n^{-1/2}V_{\lfloor sn\rfloor} \xrightarrow{\mathcal{L}} V(s) := K(s)\otimes\sigma(s)'^{-1}, \quad s\in[0,1], \tag{A.45}$$

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

$$D_{1n}'\sum_{t=1}^{n}[Z_tZ_t'\otimes\hat{\Sigma}_t^{-1}]D_{1n} = \frac{1}{n^2}\sum_{t=1}^{n}\hat{V}_{t-1}\hat{V}_{t-1}' \xrightarrow{\mathcal{L}} \int_0^1 V(s)V(s)'ds = J_1. \tag{A.46}$$

Furthermore, the stochastic orders of the second diagonal block and the off-diagonal block of (A.18), together with $\hat{\Sigma}_n(s) = \mathrm{O}_P(1)$, imply

$$D_{2n}'\sum_{t=1}^{n}[Z_tZ_t'\otimes\hat{\Sigma}_t^{-1}]D_{2n} = \mathrm{O}_P(1), \qquad D_{1n}'\sum_{t=1}^{n}[Z_tZ_t'\otimes\hat{\Sigma}_t^{-1}]D_{2n} = \mathrm{o}_P(1). \tag{A.47}$$

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

For the result corresponding to (A.20), the main task is to derive the limit of

$$D'_{1n}\sum_{t=1}^{n}[Z_t \otimes \hat{\Sigma}_t^{-1}]\varepsilon_t = n^{-1}\sum_{t=1}^{n}\hat{V}_{t-1}^{0}\eta_t, \tag{A.48}$$

where $\hat{V}_{t-1}^{0} = [\alpha'_{\perp}\Gamma X_{t-1} \otimes \hat{\Sigma}_t^{-1}\sigma_t] = \hat{V}_{t-1}[I_q \otimes \hat{\sigma}_t^{-1}\sigma_t]$, such that the continuous mapping theorem again implies $n^{-1/2}\hat{V}_{\lfloor sn \rfloor}^{0} \xrightarrow{\mathcal{L}} V(s), s \in [0,1]$. To show that (A.48) converges weakly to the stochastic integral $\int_0^1 V(s)dW(s)$, we cannot apply Hansen (1992)'s Theorem 2.1 for martingale differences, because $\{\hat{V}_t^0\}$ is not adapted to the filtration $\{\mathcal{F}_t\}$, with respect to which $\{\eta_t\}$ is a martingale difference sequence. This is caused by the fact that $\hat{\Sigma}_t$ depends, via the least-squares residuals e_t and hence the least-squares estimates $\bar{\theta}_n$, on the full sample. For the same reason, we cannot decompose η_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 $V_n(s) = n^{-1/2}\hat{V}_{\lfloor sn \rfloor}^{0}$ and note that $(V_n, W_n) \xrightarrow{\mathcal{L}} (V, W)$ in $D[0,1]^{qp\times p} \times D[0,1]^p$ 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 $D[0,1]^{qp\times p} \times D[0,1]^p$. This implies that, given $\epsilon > 0$, there exists an event $\Omega_\epsilon \subset \Omega$ such that $P(\Omega_\epsilon) \geq 1 - \epsilon$ and

$$\sup_{\omega\in\Omega_\epsilon} d\left((V^n, W^n)(\omega) - (V, W)(\omega)\right) = \delta_n \to 0, \tag{A.49}$$

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

$$0 = t_0 = \frac{n_1}{n} < t_1 = \frac{n_1}{n} < t_2 = \frac{n_2}{n} < \ldots < t_{k_n} = \frac{n_{k_n}}{n} = 1, \tag{A.50}$$

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

$$\begin{aligned}
\int_0^1 V_n(s)dW_n(s) &= n^{-1}\sum_{t=1}^{n}\hat{V}_{t-1}^{0}\eta_t \\
&= \sum_{j=1}^{k_n} V_n\left(t_{j-1}\right)\left[W_n(t_j) - W_n(t_{j-1})\right] + \frac{1}{n}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}+2}^{n_j}(\hat{V}_{t-1}^{0} - \hat{V}_{n_{j-1}}^{0})\eta_t \\
&=: G_n + Q_n.
\end{aligned} \tag{A.51}$$

Analogous to the arguments in Chan and Wei (1988) and Davidson (1994), $G_n \xrightarrow{\mathcal{L}} \int_0^1 V(s)dW(s)$ follows from $G^n \xrightarrow{P} \int_0^1 V(s)dW(s)$, where $G^n = \sum_{j=1}^{k_n} V^n\left(t_{j-1}\right)\left[W^n(t_j) - W^n(t_{j-1})\right]$. For the

remainder term Q_n, we note that

$$\begin{aligned} Q_n = \frac{1}{n}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}+2}^{n_j}(\hat{V}^0_{t-1}-\hat{V}^0_{n_{j-1}})\eta_t &= \frac{1}{n}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}+2}^{n_j}(V_{t-1}-V_{n_{j-1}})\eta_t \\ &\quad +\frac{1}{n}\sum_{j=1}^{k_n}\sum_{t=n_{j-1}+2}^{n_j}\left[(\hat{V}^0_{t-1}-V_{t-1})-(\hat{V}^0_{n_{j-1}}-V_{n_{j-1}})\right]\eta_t \\ &=: Q_{1n}+Q_{2n}. \end{aligned} \tag{A.52}$$

It follows from the convergence to the stochastic integral with known volatility that $Q_{1n} = \mathrm{o}_P(1)$. Since Q_{2n} is the same sequence, but with $\sigma_t'^{-1}$ replaced by $(\hat{\Sigma}_t^{-1}-\Sigma_t^{-1})\sigma_t$, and the latter is of lower order in probability than the former, we find $Q_{2n} = \mathrm{o}_P(1)$ and hence $Q_n = \mathrm{o}_P(1)$, such that

$$D_{1n}'\sum_{t=1}^{n}[Z_t\otimes\hat{\Sigma}_t^{-1}]\varepsilon_t = \int_0^1 V_n(s)dW_n(s) \xrightarrow{\mathcal{L}} \int_0^1 V(s)dW(s) = \int_0^1 \left[K(s)\otimes\Sigma(s)^{-1}\right]dU(s). \tag{A.53}$$

The corresponding result for $D_{1n}'\sum_{t=1}^{n}[Z_t\otimes\hat{\Sigma}_t^{-1}]\varepsilon_t(\theta_0)$ under local alternatives follows from combining (A.46) with (A.53).

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

$$\hat{\Lambda}_n(\theta_0+D_n\tau) \xrightarrow{\mathcal{L}} (S_1-J_1\tau_1)'J_1^{-1}(S_1-J_1\tau)+\Lambda_2(\tau_2), \tag{A.54}$$

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

Finally, the proof of the corresponding result for $\widetilde{LR}_n$ is entirely analogous; since the construction of the proof does not require $\tilde{\Sigma}_t$ or $\tilde{V}_{t-1} = (\alpha_\perp'\Gamma X_{t-1}\otimes\tilde{\sigma}_t'^{-1})$ to be adapted, the remainder term Q_{2n} is again of lower order than Q_{1n}, and hence asymptotically negligible. □

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, P. Labys (2003), "Modeling and Forecasting Realized Volatility", *Econometrica*, 71, 579–625.

Beare, B. K. (2004), "Robustifying Unit Root Tests to Permanent Changes in Innovation Variance", Working paper, Yale University.

Boswijk, H. P. (1995), "Identifiability of Cointegrated Systems", Tinbergen Institute Discussion Paper # 95-78, `http://www.ase.uva.nl/pp/bin/258fulltext.pdf`.

Boswijk, H. P. (2001), "Testing for a Unit Root with Near-Integrated Volatility", Tinbergen Institute Discussion Paper # 01-077/4, `http://www.tinbergen.nl/discussionpapers/01077.pdf`.

Boswijk, H. P. (2005), "Adaptive Testing for a Unit Root with Nonstationary Volatility", UvA-Econometrics Discussion Paper 2005/07, `http://www.ase.uva.nl/pp/bin/265fulltext.pdf`.

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, 917–947.

Cavaliere, G. and A. M. R. Taylor (2008), "Time-Transformed Unit Root Tests for Models with Non-Stationary Volatility", *Journal of Time Series Analysis*, forthcoming.

Cavaliere, G., A. Rahbek and A. M. R. Taylor (2007), "Testing for Co-integration in Vector Autoregressions with Non-Stationary Volatility", Granger Centre Discussion Paper Series 07/02, University of Nottingham, `http://www.nottingham.ac.uk/economics/grangercentre/papers/07-02.pdf`.

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.

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.

Engle, R. F., D. F. Hendry and J.-F. Richard (1983), "Exogeneity", *Econometrica*, 51, 277–304.

Foster, D. P. and D. B. Nelson (1996), "Continuous Record Asymptotics for Rolling Sample Variance Estimators", *Econometrica*, 64, 139–174.

Gijbels, I., A. Pope and M. P. Wand (1999), "Understanding Exponential Smoothing via Kernel Regression", *Journal of the Royal Statistical Society Series B*, 61, 39–50.

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. (2002), "Generalized Reduced Rank Regression", Brown University Working Paper No. 2002-02, `http://ssrn.com/abstract=302859`.

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. (1995), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. 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*, 365–387.

Kristensen, D. (2006), "Nonparametric Filtering of the Realised Spot Volatility: A Kernel-based Approach", Working paper, Columbia University, `http://www.columbia.edu/~dk2313/workingpapers/wp_files/SV-filter-22-05-06.pdf`.

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. (1990), "ARCH Models as Diffusion Approximations", *Journal of Econometrics*, 45, 7–38.

Nelson, D. B. (1996a), "Asymptotic Filtering Theory for Multivariate ARCH Models", *Journal of Econometrics*, 71, 1–47.

Nelson, D. B. (1996b), "Asymptotically Optimal Smoothing with ARCH Models", *Econometrica*, 64, 561–573.

Nelson, D. B. and D. P. Foster (1994), "Asymptotic Filtering Theory for Univariate ARCH Models", *Econometrica*, 62, 1–41.

Oberhofer W. and J. Kmenta (1974), "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models", *Econometrica*, 42, 579–590.

Phillips, P. C. B. and K.-L. Xu (2005), "Inference in Autoregression under Heteroskedasticity", *Journal of Time Series Analysis*, 27, 289–308.

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.