Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Testing for co-integration in vector autoregressions with non-stationary volatility

by

Giuseppe Cavaliere, Anders Rahbek and A. M. Robert Taylor

Granger Centre Discussion Paper No. 07/02

The University of Nottingham

Testing for Co-integration in Vector Autoregressions with Non-Stationary Volatility

Giuseppe Cavaliere[a,*], Anders Rahbek[b] and A.M.Robert Taylor[c]

[a] Department of Statistical Sciences, University of Bologna

[b] Department of Mathematical Sciences, University of Copenhagen

[c] School of Economics and Granger Centre for Time Series Econometrics, University of Nottingham

August 2007

Abstract

Many key macro-economic and financial variables are characterised by permanent changes in unconditional volatility. In this paper we analyse vector autoregressions with non-stationary (unconditional) volatility of a very general form, which includes single and multiple volatility breaks as special cases. We show that the conventional rank statistics of Johansen (1988,1991) are potentially unreliable. In particular, their large sample distributions depend on the integrated covariation of the underlying multivariate volatility process which impacts on both the size and power of the associated co-integration tests, as we demonstrate numerically. A solution to the identified inference problem is provided by considering wild bootstrap-based implementations of the rank tests. These do not require the practitioner to specify a parametric model for volatility, nor to assume that the pattern of volatility is common to, or independent across, the vector of series under analysis. The bootstrap is shown to perform remarkably well in practice.

Keywords: Cointegration; non-stationary volatility; trace and maximum eigenvalue tests; wild bootstrap.

J.E.L. Classifications: C30, C32.

1 Introduction

A number of recent applied studies have suggested time-varying behaviour, in particular a general decline, in unconditional volatility in the shocks driving macro-economic and financial time-series over the past twenty years or so is a relatively common phenomenon; see, *inter alia*, Busetti and Taylor (2003), Kim and Nelson (1999), Koop and Potter (2000), McConnell and Perez Quiros (2000), van Dijk *et al.* (2002), Sensier and van Dijk (2004) and reference therein. For example, Sensier and van Dijk (2004) report that over 80% of the real and price variables

*Corresponding author. Department of Statistical Sciences, Via Belle Arti 41, I-40126 Bologna, Italy. Tel: +39 051 2098230; fax: +39 051 232153.
E-mail addresses: giuseppe.cavaliere@unibo.it (G. Cavaliere), rahbek@math.ku.dk (A.Rahbek), robert.taylor@nottingham.ac.uk (A.M.R. Taylor)

in the Stock and Watson (1999) data-set reject the null of constant innovation variance against the alternative of a one-off change in variance. Similarly, Loretan and Phillips (1994) report evidence against the constancy of unconditional variances in stock market returns and exchange-rate data, while Hansen (1995) notes that empirical applications of autoregressive stochastic volatility models to financial data generally estimate the dominant root in the stochastic volatility process to be close to the non-stationarity boundary at unity. van Dijk *et al.* (2002) find evidence that volatility changes smoothly over time, while Watson (1999) argues that multiple changes in volatility are commonly observed Cavaliere and Taylor (2007) report evidence of multiple volatility breaks and trending volatility in the monthly producer price inflation series from the well-known Stock and Watson (1999) database.

These findings have helped stimulate an interest amongst econometricians into analysing the effects of non-constant volatility on univariate unit root and stationarity tests; see, *inter alia*, Hamori and Tokihisa (1997), Kim, Leybourne and Newbold (2002), Busetti and Taylor (2003), Cavaliere (2004), and Cavaliare and Taylor (2005,2006,2007,2008a). These authors show that standard unit root and stationarity tests based on the assumption of constant volatility can display significant size distortions in the presence of non-constant volatility. Cavaliere and Taylor (2008a) develop wild-bootstrap-based implementations of standard unit root tests which are shown to yield pivotal inference in the presence of non-stationary volatility. The impact of non-constant volatility on stable autoregressions has also been analysed by Hansen (1995), Phillips and Xu (2006) and Xu and Phillips (2007), *inter alia*, who show that non-constant volatility can again have a large impact on the behaviour of standard estimation and testing procedures.

Given that non-constant volatility has been found to be a common occurrence in univariate macroeconomic and financial time series, and to have a large impact on univariate time series procedures, it is clearly important and practically relevant to investigate the impact that such behaviour has on multivariate non-stationary time series methods. Indeed, using U.S. data Hansen (1992a) has shown that the regression error in four published co-integrating relations (namely, real per capita consumption upon real per capita disposable income; aggregate nondurables and services consumption upon disposable income; real stock prices upon real dividends, short term upon long term interest rates) are all affected by non-stationary variances. In a recent paper, Cavaliere and Taylor (2006) consider the impact of non-constant volatility on residual-based tests for the null hypothesis of co-integration of, *inter alia*, Shin (1994).

In this paper we analyse the impact of non-stationary volatility in the (vector) innovation process driving a co-integrated vector autoregressive (VAR) model. We allow for innovation processes whose variances evolve over time according to a quite general mechanism which allows, for example, single and multiple abrupt variance breaks, smooth transition variance breaks, and trending variances. We analyse the impact this has on the conventional trace and maximum eigenvalue statistics of Johansen (1988,1991), demonstrating that the asymptotic null distributions of these statistics depend upon the (asymptotic) integrated covariation of the underlying volatility process. Numerical simulation results for the case of a one time change in volatility suggests that this can cause a large impact on both the size and power properties of the associated tests.

In order to solve the identified inference problem, at least within the class of volatility processes considered, we extend the univariate wild bootstrap-based unit root tests of Cavaliere and Taylor (2008a) to the multivariate context by developing wild bootstrap-based implementations of Johansen's maximum eigenvalue and trace tests. Our proposed wild bootstrap procedure is set up in such a way that the practitioner is not required to specify any parametric model for volatility, nor to assume that the pattern of volatility is common to, or independent across, the vector of series under analysis.

In a recent paper, Boswijk and Zu (2007) discuss maximum likelihood estimation of VAR models when the (possibly non-stationary) spot volatility changes smoothly over time and can be estimated consistently. In such a case, their approach represents an important complement to the wild bootstrap method proposed in this paper. However, it is important to note that we adopt a different assumption from Boswijk and Zu (2007) regarding the class of non-stationary volatility processes allowed. In particular, while we allow for processes which display abrupt volatility shifts, Boswijk and Zu (2007) require the volatility process to be continuous. Moreover, our analysis does not require the existence of a consistent estimator of the underlying spot volatility. Other related work is considered in Hansen (2003) who considers estimation and testing in a co-integrated VAR model that allows for a finite number of deterministic breaks in the slope and covariance matrix of the system. In contrast to the wild bootstrap approach outlined in this paper, Hansen (2003) adopts a parametric approach to structural change, requiring that the location of the breaks in the parameters of the covariance matrix and the number of co-integrating relations present in the system are known. A further difference is that the innovations in Hansen (2003) are assumed to be homoskedastic within each regime, such that the moving average representation of the system within each regime is identical to that given in Johansen (1996). In particular, this entails that both the cointegrating relations and the common trends are homoskedastic within each regime.

The remainder of the paper is organized as follows. Section 2 outlines our heteroskedastic co-integrated VAR model, giving both error correction and common trends representations for the model. Here we also discuss the form of the co-integrating relationships in the context of this model. In section 3 the impact of non-stationary volatility on the large sample properties of Johansen's maximal eigenvalue and trace statistics is detailed. Here we also demonstrate the important result that the MLE of the parameters from our co-integrated VAR model remain consistent. Our wild bootstrap-based approach, which also incorporates a sieve procedure using the (consistently) estimated coefficient matrices from the co-integrated VAR model, is outlined in Section 4 and it is shown that this solves the inference problem caused by non-stationary volatility, yielding asymptotically pivotal co-integration tests. Monte Carlo experiments illustrating the effects of one time variance shifts on both standard and bootstrap co-integration tests are presented in section 5. Here it is shown that the proposed bootstrap tests perform very well in finite samples. Section 6 provides an empirical application to U.S. government bond yields. Section 7 concludes. All proofs are contained in the Appendix.

In the following '$\xrightarrow{w}$' denotes weak convergence, '$\xrightarrow{p}$' convergence in probability, and '$\xrightarrow{w}_p$' weak convergence in probability (Giné and Zinn, 1990; Hansen, 1996); $\mathbb{I}(\cdot)$ denotes the indicator function and '$x := y$' ('$x =: y$') indicates that x is defined by y (y is defined by x); $\lfloor \cdot \rfloor$ denotes the integer part of its argument. The notation $\mathcal{C}_{\mathbb{R}^{m\times n}}[0,1]$ is used to denote the space of $m \times n$ matrices of continuous functions on $[0,1]$; $\mathcal{D}_{\mathbb{R}^{m\times n}}[0,1]$ denotes the space of $m \times n$ matrices of càdlàg functions on $[0,1]$. The space spanned by the columns of any $m \times n$ matrix A is denoted as $\mathsf{col}(A)$; if A is of full column rank $n < m$, then $A_\perp$ denotes an $m \times (m-n)$ matrix of full column rank satisfying $A_\perp' A = 0$. For any square matrix, A, $|A|$ is used to denote the determinant of A, $\|A\|$ the norm $\|A\|^2 := \mathsf{tr}\{A'A\}$, and $\rho(A)$ its spectral radius (that is, the maximal modulus of the eigenvalues of A). For any vector, x, $\|x\|$ denotes the usual Euclidean norm, $\|x\| := (x'x)^{1/2}$.

2 The Heteroskedastic Co-integration Model

We consider the following VAR(k) model in error correction format:

$$\begin{aligned} \Delta X_t &= \alpha\beta' X_{t-1} + \Psi U_t + \mu D_t + \varepsilon_t,\ t = 1, ..., T & (1) \\ \varepsilon_t &= \sigma_t z_t & (2) \end{aligned}$$

where: X_t and ε_t are $p \times 1$, σ_t is $p \times p$, $U_t := \left(\Delta X'_{t-1}, ..., \Delta X'_{t-k+1}\right)'$ is $p(k-1) \times 1$ and $\Psi := (\Gamma_1, ..., \Gamma_{k-1})$, where $\{\Gamma_i\}_{i=1}^{k-1}$ are $p \times p$ lag coefficient matrices, D_t is a vector of deterministic terms, z_t is p-variate i.i.d., $z_t \sim (0, I_p)$, where I_p denotes the $p \times p$ identity matrix, and α and β are full column $p \times r$ matrices, $r \leq p$. The initial values $\mathbb{X}_0 := \left(X'_0, ..., X'_{-k+1}\right)'$ are assumed to be fixed. Observe that because z_t is i.i.d., conditionally on σ_t the term ε_t has mean vector zero and time-varying covariance matrix $\Sigma_t := \sigma_t\sigma'_t$, the latter assumed to be positive definite for all t.

Throughout the paper we assume that the process in (1) satisfies the following set of three conditions, which we label collectively as Assumption 1:

Assumption 1: (a) *all the characteristic roots associated with (1); that is of* $A(z) := I_p - \alpha\beta' z - \Gamma_1 z(1-z) - \cdots - \Gamma_{k-1} z^{k-1}(1-z) = 0$, *are outside the unit circle or equal to* 1; (b) $\det(\alpha'_\perp \Gamma \beta_\perp) \neq 0$, *with* $\Gamma := I_p - \Gamma_1 - \cdots - \Gamma_{k-1}$.

For unknown parameters α, β, Ψ, μ, and for a given sequence $\{\Sigma_t\}$, when α and β are $p \times r$ matrices not necessarily of full rank (1)-(2) denotes our heteroskedastic co-integrated VAR model, which we denote as $H(r)$. We assume that the deterministic part can be partitioned into $D_t := (D'_{1t} : D'_{2t})'$ and $\mu := (\mu_1 : \mu_2)$ where $\mu_1 = \alpha\rho'_1$ is the part of the coefficient of the deterministic terms that is constrained to be in $\mathsf{col}(\alpha)$. The model may then be written in the compact form

$$Z_{0t} = \alpha\beta^{*\prime} Z_{1t} + \mu_2 Z_{2t} + \varepsilon_t \tag{3}$$

with $Z_{0t} := \Delta X_t$, $Z_{1t} := (X'_{t-1} : D'_{1t})'$, $Z_{2t} := (U'_t : D'_{2t})'$, $\beta^* := (\beta' : \rho'_1)'$. If D_{it} is set equal to 0, it is understood that D_{it} is to be dropped from the definition of Z_{it}, $i = 1, 2$.

Through the paper the following assumption will be taken to hold on the sequence of $p \times p$ volatility matrices $\{\sigma_t\}$ of (2).

Assumption 2: *The volatility matrix* σ_t *is non-stochastic and satisfies* $\sigma_{\lfloor Tu \rfloor} := \sigma(u)$ *for all* $u \in [0, 1]$, *where* $\sigma(\cdot) \in \mathcal{D}_{\mathbb{R}^{p\times p}}[0, 1]$. *Moreover it is assumed that* $\Sigma(u) := \sigma(u)\sigma(u)'$ *is positive definite for all* $u \in [0, 1]$.

Remark 2.1. Assumption 2 generalises the corresponding scalar assumption of Cavaliere (2004) and Cavaliere and Taylor (2007) to the multivariate case, and is the key condition of the present paper. This assumption allows us to cast the dynamics of the innovation variance in a very general framework. It requires the elements of the innovation covariance matrix Σ_t only to be bounded and to display a countable number of jumps and therefore allows for an extremely wide class of potential models for the behaviour of the covariance matrix of ε_t. To see this fact, let $\Sigma(\cdot) := \sigma(\cdot)\sigma(\cdot)'$ denote the limiting spot covariance process. Models of single or multiple variance or covariance shifts, as are considered in Hansen (2003), satisfy Assumption 2 with $\Sigma(\cdot)$ piecewise constant. For instance, the case of a single break at time $\lfloor \tau_{ij} T \rfloor$ in the covariance $E(\varepsilon_{it}\varepsilon_{jt})$ obtains for $\Sigma_{ij}(u) := \Sigma_{ij}^0 + (\Sigma_{ij}^1 - \Sigma_{ij}^0)\mathbb{I}(u > \tau_{ij})$. If $\Sigma_{ij}(\cdot)$ is an affine function, then $\Sigma_{t,ij}$ displays a linear trend. Piecewise affine functions are also permitted, thereby allowing for variances which follow a broken trend, as are smooth transition variance shifts. Finally, observe that the case of constant unconditional volatility, where $\sigma_t = \sigma$, for all t, clearly also satisfies Assumption 2 with $\sigma(u) = \sigma$.

Remark 2.2. It is not strictly necessary to require that the volatility function $\sigma(\cdot)$ is non-stochastic, but this assumption allows for a considerable simplification of the theoretical set-up; see the discussion in Cavaliere and Taylor (2007). This can be weakened to allow for cases where the innovations $\{e_t\}$ and σ_t are stochastically independent at all leads and lags. In such a case, if the (exogenous) volatility process $\sigma(\cdot)$ has sample paths satisfying Assumption 2, the results presented should then be read as *conditional* on a given realization of $\sigma(\cdot)$. The conditioning argument used here in the context of the volatility function serves the same purpose as the exogeneity assumption used by Perron (1989,pp.1387-8) to permit stochastic changes in the trend function. Moreover, we conjecture that the results given in the paper will continue to hold if the condition $\sigma_{\lfloor T\cdot \rfloor} = \sigma(\cdot)$ in Assumption 2 is replaced by the weaker requirement that $\sigma_{\lfloor T\cdot \rfloor} \xrightarrow{w} \sigma(\cdot)$ in the space $\mathcal{D}_{\mathbb{R}^{p\times p}}[0,1]$ equipped with the Skorohod topology, with $\sigma(\cdot)$ being possibly stochastic and independent of $\{z_t\}$

Remark 2.3. A special case of the volatility model considered here arises by setting $\sigma_t := \Lambda V_t$, V_t a (full rank) time-varying diagonal matrix, initialized at $V_0 = I_p$, and Λ a constant $p \times p$ nonsingular matrix. In this case, $\sigma(\cdot)$ of Assumption 2 has the form $\sigma(\cdot) = \Lambda V(\cdot)$, $V(\cdot)$ now depending on a vector of càdlàg processes. A similar factorization has been employed recently by, *inter alia*, Van der Weide (2002) and Lanne and Saikkonen (2007) in the context of multivariate conditionally heteroskedastic models, and by Lanne and Lütkepohl (2005) to model non-normality in VAR processes. Finally, notice that the special case where the errors share common volatility shocks obtains, for example, by setting $V_t := v_t I_p$, where v_t is a scalar process satisfying Assumption 2.

Remark 2.4. Since the variance σ_t depends on T, a time series generated according to (1)-(2) with σ_t satisfying Assumption 2 formally constitutes a triangular array of the type $\{X_{T,t} : 0 \leq t \leq T, T \geq 1\}$, where $X_{T,t}$ is recursively defined as $\Delta X_{T,t} = \alpha\beta' X_{T,t-1} + \Psi U_{T,t} + \mu D_t + \sigma_{T,t} z_t$, $\sigma_{T,[Tu]} := \sigma(u)$. However, since the triangular array notation is not essential, for simplicity the subscript T is suppressed in the sequel. □

We now discuss some of the long-run and short-run features of the heteroskedastic co-integrated model relating these back to the corresponding standard homoskedastic co-integrated model where appropriate.

2.1 Representation

Under Assumptions 1 and 2 the model (1)-(2) admits the following representation which generalizes the well-known MA representation of a co-integrated $I(1)$ VAR model with homoskedastic innovations; cf. Johansen (1996) and the references therein.

Lemma 1 *Under Assumptions 1 and 2,*

$$X_t = C \sum_{i=1}^{t} (\sigma_i z_i + \mu D_i) + S_t + C_0. \tag{4}$$

Here $C := \beta_{\perp} \left(\alpha_{\perp}' \Gamma \beta_{\perp}\right)^{-1} \alpha_{\perp}'$ *and* $C_0 := C(I_p, -\Psi)\mathbb{X}_0$ *is a constant which depends on the initial values. The p-dimensional process* $S_t := (\alpha, \Psi) Q \mathbb{X}_{\beta t}$*, where*

$$\mathbb{X}_{\beta t} := \left\{ \begin{array}{c} \beta' X_t \textit{ for } k = 1 \\ \left(X_t'\beta, \Delta X_t', ..., \Delta X_{t-k+1}'\right)' \textit{ otherwise} \end{array} \right.$$

is a $(r + p(k-1))$*-dimensional heteroskedastic autoregressive process satisfying,*

$$\mathbb{X}_{\beta t} = \Phi \mathbb{X}_{\beta t-1} + \nu D_t + \eta_t, \; \eta_t := (\beta, I_p, 0, ..., 0)' \sigma_t z_t \tag{5}$$

where $\nu := (\beta, I_p, 0, ..., 0)' \mu$*. In particular, we say that* $\mathbb{X}_{\beta t}$ *is 'stable' as the spectral radius of* Φ *is smaller than one; that is,* $\rho(\Phi) < 1$*. The matrix* Q *is non-singular and* $(r + p(k-1)) \times (r + p(k-1))$ *dimensional.*

The result given in Lemma 1 differs from the standard case in two main respects. First, the cumulated shocks appearing on the right hand side of (4) display non-stationary volatility (unless $\sigma(\cdot)$ of Assumption 2 is constant) and, hence, do not form a standard random walk as in the constant volatility case. Second, the component S_t, although stable, is non-stationary due to the fact that its volatility changes over time. In Lemma A.1 in the Appendix we show that multivariate stable processes with heteroskedastic innovations, such as the example in (5), satisfy a law of large numbers (LLN), irrespectively of initial values. This lemma complements similar results in Phillips and Xu (2006) who consider heteroskedastic, univariate infinite-order moving average processes.

The implications of these two features of the model are discussed further below.

2.2 Common Trends

By analogy to the homoskedastic case we may define the common trends as the $(p - r) \times 1$ vector

$$P_t := \alpha_{\perp}' \sum_{i=1}^{t} \varepsilon_i = \alpha_{\perp}' \sum_{i=1}^{t} \sigma_i z_i. \tag{6}$$

Due to the time-variation in σ_t, P_t is in general not $I(1)$ in the conventional sense; rather, P_t is a $(p-r)$-dimensional process driven by heteroskedastic innovations satisfying Assumption 2. However, similarly to the homoskedastic case, the common trend component P_t is of order $T^{1/2}$ and satisfies a functional central limit theorem (FCLT), although the limiting process involved is no longer a (multivariate) standard Brownian motion. To see this fact, consider the following lemma, which holds under Assumption 2.

Lemma 2 *Let* $z_t \sim iid(0, I)$*,* $\varepsilon_t = \sigma_t z_t$ *and let* $B(\cdot)$ *denote a p-variate standard Brownian motion. Then, under Assumption 2, as* $T \to \infty$*,*

$$\left(\frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor T \cdot \rfloor} \varepsilon_t, \frac{1}{T^{3/2}} \sum_{t=1}^{T} \left(\sum_{i=1}^{t-1} \varepsilon_i \right) \varepsilon_t' \right) \xrightarrow{w} \left(M(\cdot), \int_0^1 M(s) \left(dM(s)\right)' \right),$$

where $M(\cdot) := \int_0^{\cdot} \sigma(s) \, dB(s)$ *is a p-variate continuous martingale.*

This result generalizes the well known FCLT and convergence to stochastic integrals results for partial sums of homoskedastic random walks to the case of general volatility dynamics satisfying Assumption 2; standard convergence results discussed e.g. in Hansen (1992b), Johansen (1996) follow as special cases when $\sigma(\cdot)$ is constant. It follows immediately from Lemma 2 and an application of the continuous mapping theorem that

$$P_T(\cdot) := \frac{1}{T^{1/2}} P_{\lfloor T\cdot \rfloor} \xrightarrow{w} \alpha_\perp' M(\cdot).$$

That is, the scaled common trends component does not converge in the limit to a vector Brownian motion; rather, it converges to a process with increments which although still independent are no longer identically distributed through time. More specifically, the limiting process $M(\cdot)$ is a continuous martingale with spot volatility $\sigma(\cdot)$ and integrated covariation equal to $\int_0^\cdot \Sigma(s)\, ds$; cf. Shephard (2005,p.9).

Remark 2.5. Although ε_t is heteroskedastic, it is possible for the common trends to be standard (homoskedastic) random walks. In particular, this will occur if $\alpha_\perp$ annihilates the variability of σ_t; that is, if $\alpha_\perp'\sigma_t = \alpha_\perp'\sigma_r$, all $t, r = 1, ..., T$. In such a case, P_t is a standard random walk and $T^{-1/2}P_{\lfloor T\cdot \rfloor}$ converges to a multivariate standard Brownian motion, regardless of any heteroskedasticity in the innovations ε_t.

Remark 2.6. It is interesting to analyze the form of the limiting 'common trend' process $M(\cdot)$ in the special case considered in Remark 2.3. In this case, $M(\cdot)$ is a vector variance-transformed Brownian motion; see Davidson (1994). To see this, observe that $M(\cdot)$ may be written as

$$M(\cdot) := \int_0^\cdot \sigma(s)\, dB(s) = \Lambda \int_0^\cdot V(s)\, dB(s) = \Lambda^* \begin{bmatrix} \bar{V}_1^{-1} \int_0^\cdot V_1(s)\, dB_1(s) \\ \vdots \\ \bar{V}_p^{-1} \int_0^\cdot V_p(s)\, dB_p(s) \end{bmatrix} =: \Lambda^* \begin{bmatrix} B_{\eta 1}(s) \\ \vdots \\ B_{\eta p}(s) \end{bmatrix}$$

where $\bar{V}_i := (\int_0^1 V_i(s)^2\, ds)^{1/2}$ $(i = 1, ..., p)$, $\Lambda^* := \Lambda\bar{V}$, with $\bar{V} := \mathsf{diag}(\bar{V}_i, ..., \bar{V}_p)$. Each of the $B_{\eta i}(\cdot) := \bar{V}_i^{-1} \int_0^\cdot \bar{V}_i(s)\, dB_i(s)$ is a variance-transformed (or time-change) Brownian motion with directing process $\eta_i(\cdot) := \bar{V}_i^{-2} \int_0^\cdot V_i(s)^2\, ds$; cf. Davidson (1994, p.486). Hence, $M(\cdot) = \Lambda^* B_\eta(\cdot)$, where $B_\eta(\cdot) := (B_{\eta 1}(\cdot), ..., B_{\eta p}(\cdot))'$ is a vector of independent variance-transformed Brownian motions. This implies, see Davidson (1994, p.492), that $M(\cdot)$ is a vector variance transformed Brownian motion on $[0, 1]$, defined by the covariance matrix $\Lambda^*\Lambda^{*\prime}$ and the homeomorphism , $\eta(\cdot) := (\eta_1(\cdot), ..., \eta_p(\cdot))'$. $\square$

2.3 Co-integrating Relations

Let us briefly turn to a consideration of the linear combination $\beta' X_t$. In the homoskedastic case, it is well known that under Assumption 1 $\beta' X_t$ can be given an initial distribution such that it is stationary. In the heteroskedastic case, however, stationarity does not hold in general, due to the time-variation in σ_t. Nonetheless, $\beta' X_t$ is stable, in the sense that it is free of stochastic trends.

To see this fact, taking the case of no deterministics to illustrate, recall that representation (6) implies that, apart from the contribution of the initial values, the linear combination $\beta' X_t$

depends on a linear combination of a stable process; that is,

$$\beta' X_t = (I_p, 0, ..., 0)\, \mathbb{X}_{\beta t},$$

where $\mathbb{X}_{\beta t}$ is a first-order vector autoregressive process with stable roots only. However, in contrast to the homoskedastic case, $\beta' X_t$ cannot be made stationary by an appropriate choice of the initial values since the innovations to $\mathbb{X}_{\beta t}$ have non-stationary volatility. Hence, the model considered in this paper generates co-integrating relations which are generally non-stationary due to heteroskedasticity but are stable.

A key feature of the stability of $\beta' X_t$ is, as already noticed, that the law of large numbers applies to the sample moments of $\beta' X_t$. For instance, and taking the case of $\mu = 0$ to illustrate, from Lemma A.1 in the appendix it follows that

$$\frac{1}{T}\sum_{t=1}^{T} \beta' X_t \left(\beta' X_t\right)' \xrightarrow{p} \bar{\Sigma}_{\beta\beta}$$

with $\bar{\Sigma}_{\beta\beta}$ a well-defined, full rank covariance matrix; see Section A.2 of the Appendix.

Remark 2.7. Even where ε_t is heteroskedastic, it is still possible for the co-integrating relations to be stationary. Specifically, in the $k = 1$ case, if β annihilates the variability of σ_t, in the sense that $\beta'\sigma_t = \beta'\sigma_r$ (all t, r), then $\beta' X_t$ can be made stationary by choosing the initial values appropriately. This feature shows that although in the 'heteroskedastic' VAR(1) model with time-varying volatility $\beta' X_t$ is in general heteroskedastic, strict stationarity may in fact hold if the additional restriction that $\beta'\sigma_t$ is constant over time holds. □

3 The Impact of Non-Stationary Volatility on Standard Gaussian Co-integration Analysis

In this section we focus on the properties of standard Gaussian-based Maximum Likelihood (ML) estimators and associated co-integration rank tests based on the assumption of independent, identically distributed Gaussian shocks when volatility is time-varying satisfying Assumption 2. In this case the Likelihood Ratio (LR) approach of Johansen (1991) is based on a mis-specified model and hence should be considered a pseudo Maximum Likelihood (PML) method. Two results are given in this section. First, we show that the pseudo LR (PLR) co-integration rank tests have non-pivotal asymptotic null distributions; that is, standard critical values cannot be employed in general. Second, we show that even though based on a mis-specified model PML still delivers consistent estimation of the (identified) co-integrating vector β and of the parameters α, Ψ, μ.

As is standard, let $M_{ij} := T^{-1}\sum_{t=1}^{T} Z_{it} Z'_{jt}$, $i, j = 0, 1, 2$, with Z_{it} defined as in (3), and let $S_{ij} := M_{ij.2} := M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $i, j = 0, 1$. Under the auxiliary assumption of Gaussian disturbances and constant volatility – that is, $\Sigma_t := \sigma_t\sigma_t' = \Sigma$ – the pseudo Gaussian likelihood function depends on the vector $\theta^{PML} := (\alpha, \beta, \Psi, \mu, \Sigma)$. We denote the corresponding PML estimator as $\hat{\theta}^{PML} := (\hat{\alpha}, \hat{\beta}, \hat{\Psi}, \hat{\mu}, \hat{\Sigma})$. Write the maximized (pseudo) log-likelihood under $H(r)$, say $\ell(r)$, as

$$\ell(r) = -\frac{T}{2}\log|S_{00}| - \frac{T}{2}\sum_{i=1}^{p}\log\left(1 - \hat{\lambda}_i\right)$$

where $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p$, solve the eigenvalue problem

$$\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0. \tag{7}$$

The PLR test statistic for $H(r)$ vs $H(p)$ is given by

$$Q_r := -2\left(\ell\left(r\right) - \ell\left(p\right)\right) = -T\sum_{i=r+1}^{p}\log\left(1-\hat{\lambda}_i\right) \tag{8}$$

and it is well know that under the null hypothesis the asymptotic distribution of Q_r is given by $\mathsf{tr}(\int(dB(s))F(s)'\left(\int F(s)F(s)'\right)^{-1}\int F(s)(dB(s))')$, where $B(\cdot)$ is a $(p-r)$-variate standard Brownian motion and $F(\cdot)$ depends on $B(\cdot)$ and on the deterministic term; see Johansen (1991) for further details. This result, however, no longer holds under non-stationary volatility of the form considered in Assumption 2.

More specifically, the following result holds under the null hypothesis, where to keep the presentation simple we assume, for the present, no deterministics in the model and in the estimation.

Theorem 1 *Let $\{X_t\}$ be generated as in (1)-(2) under Assumptions 1 and 2 with $\mu = 0$, and assume that z_t is symmetrically distributed with finite fourth order moment κ. Then, under the hypothesis $H(r)$, as $T \to \infty$, $Q_r := -2\left(\ell\left(r\right) - \ell\left(p\right)\right)$ has asymptotic distribution*

$$Q_r \xrightarrow{w} Q_{r,\bar{\Sigma}}^{\infty} := \mathsf{tr}\left(\int_0^1 (d\tilde{M}(s))\tilde{M}(s)'\left(\int_0^1 \tilde{M}(s)\tilde{M}(s)'ds\right)^{-1}\int_0^1 \tilde{M}(s)(d\tilde{M}(s))'\right) \tag{9}$$

where $\tilde{M}(\cdot)$ is a $(p-r)$-variate Gaussian process with independent increments and I_{p-r} integrated covariation at unity. More specifically, $\tilde{M}(\cdot)$ is the $(p-r)$-variate stochastic volatility process

$$\tilde{M}\left(u\right) := \left(\alpha_\perp'\bar{\Sigma}\alpha_\perp'\right)^{-1/2}\alpha_\perp'\int_0^u \sigma(s)dB(s)$$

where $\bar{\Sigma} := \int_0^1 \Sigma\left(s\right)ds$ and $B(\cdot)$ is a p-variate standard Brownian motion.

Remark 3.1. The asymptotic null distribution of the Q_r statistic is not a functional of a standard Brownian motion as in the homoskedastic case considered in Johansen (1991). Although, like a standard Brownian motion, this process is Gaussian, continuous and has independent increments, these increments are, however, not necessarily stationary. As is clear from Lemma 2, the asymptotic null distribution of Q_r will in general depend on the integrated covariation, $\int_0^{\cdot}\Sigma\left(s\right)ds$ of $M(\cdot)$. Consequently, inference using the standard trace statistics will not in general be pivotal if p-values are retrieved on the basis of the tabulated distributions.

Remark 3.2. As in Johansen (1991), under $H(r)$, the r largest eigenvalues solving (7), $\hat{\lambda}_1, \ldots, \hat{\lambda}_r$, converge in probability to positive numbers, while $T\hat{\lambda}_{r+1}, \ldots, T\hat{\lambda}_p$ are of $O_p(1)$. Consequently, and as in the case of constant volatility, under Assumption 2 the test based on Q_r will be consistent at rate $O_p(T)$ if the true co-integration rank is, say, $r_0 > r$. Despite this, the result in Theorem 1 states that under non-stationary volatility the sequential approach to determining the co-integration rank[1] outlined in Johansen (1992) will not in general lead to the selection of the correct co-integrating rank with probability $(1-\xi)$ in large samples, as it does in the constant volatility case. The impact of a one time break in volatility on the sequential procedure is explored numerically in section 5.

Remark 3.3. Under the special case considered in Remark 2.6, $\tilde{M}(\cdot)$ obtains as a nonsingular linear combination of a vector variance transformed Brownian motion.

Remark 3.4. From the definition of $\tilde{M}(\cdot)$ of (9) it can be seen that this process is built up from $\alpha_\perp' M(\cdot) = \alpha_\perp' \int_0^{\cdot} \sigma(s) dB(s)$. As a consequence, if $\alpha_\perp$ annihilates the variation in $\sigma(\cdot)$; that is, if $\tilde{\sigma}(s) := \alpha_\perp' \sigma(s)$ is constant over time (cf. Remark 2.5), then so the limiting distribution in (9) will reduce to the so-called multivariate Dickey-Fuller distribution, with $p-r$ degrees of freedom, $\mathsf{tr}(\int (dB(s))B(s)'(\int B(s)B(s)'ds)^{-1} \int B(s)(dB(s))')$. This result has a very important implication: provided the non-stationary volatility appears in the stable direction of the system only, then the trace test will continue to have the same asymptotic distribution as reported in Johansen (1991).

Remark 3.5. The discussion outlined in this section extends to the so-called maximum eigenvalue test; that is, a PLR test based for $H(r)$ vs $H(r+1)$. As is known, this test leads to the statistic is given by

$$Q_{r,\max} := -2\left(\ell(r) - \ell(r+1)\right) = -T \log\left(1 - \hat{\lambda}_{r+1}\right) ,$$

see, for example, Equation (2.14) of Johansen (1991). It then follows from the proof of Theorem 1 that the null asymptotic distribution of $Q_{r,\max}$ corresponds to the distribution of the maximum eigenvalue of the real symmetric random matrix $\int (d\tilde{M}(s))\tilde{M}(s)'(\int \tilde{M}(s)\tilde{M}(s)'ds)^{-1}$ $\int \tilde{M}(s)(d\tilde{M}(s))'$ appearing in (9). Hence, as for the trace test, inference based on the maximal eigenvalue statistics will not in general be pivotal if p-values based on a homoskedasticity assumption are used.

Remark 3.6. The results given in this section can be generalized to the case of deterministic time trends of the well form analyzed, for example, in Sections 5.7 and 6.1 of Johansen (1996). These are straightforward generalisations of the representations given in Johansen (1996), replacing the standard Brownian motion $B(\cdot)$ with the limiting process $\tilde{M}(\cdot)$ from Lemma 2 throughout. □

We conclude this section by demonstrating that the PML estimator $\hat{\theta}^{PML}$ is consistent.

Theorem 2 *Under the conditions of Theorem 1,* $T^{1/2}(\hat{\beta}-\beta) \xrightarrow{p} 0$. *Moreover,* $\hat{\alpha} \xrightarrow{p} \alpha$, $\hat{\Psi} \xrightarrow{p} \Psi$ *and* $\hat{\Sigma} \xrightarrow{p} \bar{\Sigma}$ *as* $T \to \infty$.

Remark 3.7. Theorem 2 shows that in the presence of time-varying volatility of the form specified in Assumption 2, the PLR estimators of α, β and Ψ remain consistent. Moreover, the estimator of the pseudo parameter Σ converges in probability to the (asymptotic) average innovation variance, $\int_0^1 \Sigma(s)\, ds$. These are key properties which will be needed to establish the usefulness of the bootstrap co-integration rank test which will be presented in the next section. □

[1]This procedure starts with $r=0$ and sequentially raises r by one until for $r=\hat{r}$ the test statistic $Q_{\hat{r}}$ does not exceed the ξ level critical value for the test.

4 Bootstrapping the PLR Test

As demonstrated in Theorem 1, in the presence of volatility of the form specified in Assumption 2 the asymptotic null distributions of the PLR tests on the co-integration rank will, in general, depend on the asymptotic integrated covariation, $\int_0^{\cdot} \sigma(r)\,\sigma(r)'\,dr$, implying that inference based on the standard homoskedastic critical values will not be pivotal; cf. Remark 3.1. In this section we show that because, as was shown in Theorem 2, we can still consistently estimate α, β and Ψ, (asymptotically) pivotal p-values can be obtained in the presence of time-varying heteroskedasticity of the form considered in Assumption 2 using re-sampling methods.

Our proposed re-sampling algorithm draws on the wild bootstrap literature (see, *inter alia*, Wu, 1986; Liu, 1988; Mammen, 1993) and allows us to construct bootstrap unit root tests which are asymptotically robust to non-stationary volatility. In the context of the present problem, the wild bootstrap scheme is required, rather than the standard residual re-sampling schemes used for other bootstrap co-integration tests proposed in the literature; see, e.g., Swensen (2006) and, in the univariate case, Inoue and Kilian (2002), Paparoditis and Politis (2003), Park (2003). This is because, unlike these other schemes, the wild bootstrap replicates the pattern of heteroskedasticity present in the original shocks; cf. Remark 4.1 below.

4.1 The Bootstrap Algorithm

Let us start by considering the problem of testing the null hypothesis $H(r)$ against $H(p)$, $r < p$. Swensen (2006, section 2) discusses at length a way of implementing a bootstrap version of the well known trace test in this case. Here we extend his approach by modifying his resampling scheme in order to account the presence of time-varying volatility using the wild bootstrap. Implementation of the wild bootstrap requires us only to estimate the VAR(k) model under $H(p)$ (i.e., the unrestricted VAR) and under $H(r)$.

As in section 3, let $\hat{\Psi} := (\hat{\Gamma}_1, ..., \hat{\Gamma}_{k-1})$ and $\hat{\mu}$ denote the PML estimates of Ψ and μ, respectively, from the model under $H(p)$; the corresponding unrestricted residuals are denoted by $\hat{\varepsilon}_t$, $t = 1, ..., T$. In addition, let $\hat{\alpha}, \hat{\beta}$ denote the PML estimates of α,β under the null hypothesis $H(r)$. The bootstrap algorithm we consider in this section requires that the roots of the equation $|\hat{A}^*(z)| = 0$ are either one or outside the unit circle, where

$$\hat{A}(z) := (1-z)\,I_p - \hat{\alpha}\hat{\beta}' z - \hat{\Gamma}_1 (1-z)\,z - ... - \hat{\Gamma}_{k-1}(1-z)\,z^{k-1} \ ;$$

moreover, we also require that $|\hat{\alpha}_\perp \hat{\Gamma} \hat{\beta}_\perp| \neq 0$, $(\hat{\Gamma} := I_p - \hat{\Gamma}_1 - ... - \hat{\Gamma}_{k-1})$. While the latter condition is always satisfied in practice, if the former condition is not met, then the bootstrap algorithm cannot be implemented, because the bootstrap samples may become explosive; cf. Swensen (2006, Remark 1). However, in such cases any estimated roots which have modulus greater than unity may simply be shrunk to have modulus strictly less than unity; cf. Burridge and Taylor (2001,p.73).

The following steps constitute our wild bootstrap algorithm:

Algorithm 1 (Wild Bootstrap Co-integration Test)

Step 1: *Generate T bootstrap residuals ε_t^b, $t = 1, ..., T$, according to the device*

$$\varepsilon_t^b := \hat{\varepsilon}_t w_t$$

where $\{w_t\}_{t=1}^T$ denotes an independent $N(0,1)$ scalar sequence;

Step 2: *Construct the bootstrap sample recursively from*

$$\Delta X_t^b := \hat{\alpha}\hat{\beta}' X_{t-1}^b + \hat{\Gamma}_1 \Delta X_{t-1}^b + ... + \hat{\Gamma}_{k-1} \Delta X_{t-k+1}^b + \varepsilon_t^b, \; t = 1, ..., T,$$

with initial values, $X_{-k+1}^b, ..., X_0^b$;

Step 3: *Using the bootstrap sample, $\{X_t^b\}$, obtain, setting the bootstrap lag length, k^b, equal to k, the bootstrap test statistic, $Q_r^b := -2\left(\ell^b(r) - \ell^b(p)\right)$, where $\ell^b(r)$ and $\ell^b(p)$ denote the bootstrap analogues of $\ell(r)$ and $\ell(p)$, respectively;*

Step 4: *Bootstrap p-values are then computed as, $p_T^b := 1 - G_T^b(Q_r)$, where $G_T^b(\cdot)$ denotes the cumulative distribution function (cdf) of Q_r^b.* $\square$

Remark 4.1. Notice that the bootstrap shocks, ε_t^b, replicate the pattern of heteroskedasticity present in the original shocks since, conditionally on $\hat{\varepsilon}_t$, ε_t^b is independent over time with zero mean and variance $\hat{\varepsilon}_t \hat{\varepsilon}_t'$. Specifically, notice that, conditionally on the data,

$$T^{-1/2} \sum_{i=1}^{\lfloor T \cdot \rfloor} \varepsilon_t^b = T^{-1/2} \sum_{i=1}^{\lfloor T \cdot \rfloor} \hat{\varepsilon}_t w_t \sim N\left(0, \frac{1}{T} \sum_{t=1}^{\lfloor T \cdot \rfloor} \hat{\varepsilon}_t \hat{\varepsilon}_t'\right).$$

Since, as it will be shown below, $T^{-1} \sum_{t=1}^{\lfloor T \cdot \rfloor} \hat{\varepsilon}_t \hat{\varepsilon}_t' \approx T^{-1} \sum_{t=1}^{\lfloor T \cdot \rfloor} \sigma_t \sigma_t' \to \int_0^{\cdot} \Sigma(s)\, ds$, the cumulated bootstrap shocks display the same (asymptotic) integrated covariation as the original shocks. This will turn out to be a key property for establishing that the wild bootstrap statistic Q_r^b has the same first-order asymptotic null distribution as the standard Q_r statistic; cf. section 4.2 below.

Remark 4.2. As is standard, the bootstrap samples are generated by imposing the null co-integration rank on the re-sampling scheme, thereby avoiding the difficulties with the use of unrestricted estimates of the impact matrix Π; see Basawa *et al.* (1991) in the univariate case and Swensen (2006) in the multivariate case.

Remark 4.3. As is well known in the wild bootstrap literature (see Davidson and Flachaire, 2001, for a review) in certain cases improved accuracy can be obtained by replacing the Gaussian distribution used for generating the pseudo-data by an asymmetric distribution with $E(w_t) = 0$, $E(w_t^2) = 1$ and $E(w_t^3) = 1$ (Liu, 1988). A well known example is Mammen's (1993) two-point distribution: $P(w_t = -0.5(\sqrt{5}-1) = 0.5(\sqrt{5}+1)/\sqrt{5} = p$, $P(w_t = 0.5(\sqrt{5}+1)) = 1 - p$. Davidson and Flachaire (2001) also consider the Rademacher distribution: $P(w_t = 1) = 1/2 = P(w_t = -1)$. We found no discernible differences between the finite sample properties of the bootstrap unit root tests based on the Gaussian or the Mammen or Rademacher distributions. This finding is consistent with evidence reported in Table 5 of

Gonçalves and Kilian (2004,p.105) in the context of hypothesis testing in stationary univariate autoregressive models.

Remark 4.4. In practice, the cdf $G_T^b(\cdot)$ required in Step 4 of Algorithm 1 will not be known, but can be approximated in the usual way through numerical simulation; cf. Hansen (1996) and Andrews and Buchinsky (2000). This is achieved by generating N (conditionally) independent bootstrap statistics, $Q_{n:r}^b$, $n = 1, ..., N$, computed as above but recursively from

$$\Delta X_{n:t}^b := \hat{\alpha}\hat{\beta}' X_{n:t-1}^b + \hat{\Gamma}_1 \Delta X_{n:t-1}^b + ... + \hat{\Gamma}_{k-1} \Delta X_{n:t-k+1}^b + \varepsilon_{n:t}^b,\ t = 1, ..., T,$$

for some initial values $X_{n:-k+1}^b, ..., X_{n:0}^b$ and with $\{\{w_{n:t}\}_{t=1}^T\}_{n=1}^N$ a doubly independent $N(0,1)$ sequence. The simulated bootstrap p-value is then computed as $\tilde{p}_T^b := N^{-1}\sum_{n=1}^N \mathbb{I}\left(Q_{n:r}^b > Q_r\right)$, and is such that $\tilde{p}_T^b \overset{a.s.}{\to} p_T^b$ as $N \to \infty$. Note that an asymptotic standard error for $\tilde{p}_T^b$ is given by $\tilde{p}_T^b(1-\tilde{p}_T^b)/N)^{1/2}$; cf. Hansen (1996, p.419).

Remark 4.5. The maximum eigenvalue statistic, $Q_{r,\max}$ for $H(r)$ vs $H(r+1)$ can be bootstrapped in the same way as outlined for Q_r above, replacing Q_r^b with $Q_{r,\max}^b := -2\left(\ell^b(r) - \ell^b(r+1)\right)$ in Steps 3 and 4 of Algorithm 1, and similarly in Remark 4.4.

4.2 Asymptotic Theory

The asymptotic validity of the wild bootstrap method outlined in Algorithm 1 is now established in Theorem 3. The proof of Theorem 3 is a modification of the proof of Proposition 1 in Swensen (2006) to the case of wild bootstrap and heteroskedastic innovations. In order to keep our presentation simple, we again demonstrate our result for the case of no deterministic variables. The equivalence of the first-order limiting null distributions of the Q_r^b and Q_r statistics can also be shown to hold for the deterministic time trends models discussed in Remark 3.6. Again this is straightforward to show and is omitted in the interests of brevity.

Theorem 3 *Let the conditions of Theorem 1 hold. Then, under the null hypothesis $H(r)$, $Q_r^b \overset{w}{\to}_p Q_{r,\bar{\Sigma}}^\infty$ as $T \to \infty$. Moreover, $p_T^b \overset{w}{\to} U[0,1]$.*

Remark 4.6. A comparison of the result for Q_r^b in Theorem 3 with that given for Q_r in Theorem 1 demonstrates the usefulness of the wild bootstrap: as the number of observations diverges, the bootstrapped statistics have the same first-order null distribution as the original test statistics. Consequently, the bootstrap p-values are (asymptotically) uniformly distributed under the null hypothesis, leading to tests with (asymptotically) correct size even in the presence of non-stationary volatility of the form given in Assumption 2.

Remark 4.7. Because Step 2 of the bootstrap procedure outlined in Algorithm 2 imposes the null hypothesis $H(r)$ on the re-sampling procedure, the bootstrap eigenvalues, $\hat{\lambda}_1^b > \ldots > \hat{\lambda}_p^b$, say, which solve the bootstrap analogue of (7) will, regardless of the true co-integrating rank, be such that $\hat{\lambda}_1^b, \ldots, \hat{\lambda}_r^b$ will converge in probability to positive numbers, while $T\hat{\lambda}_{r+1}^b, \ldots, T\hat{\lambda}_p^b$ will be of $O_p(1)$. An immediate consequence of this is that the bootstrap Q_r^b statistic will remain of $O_p(1)$ when the true co-integrating rank r_0 exceeds r, which obviously implies from Step 4 of Algorithm 2 that our bootstrap procedure will be consistent at rate $O_p(T)$, due to the divergence of the standard Q_r statistic; cf. Remark 3.2. This, coupled with the asymptotically correct size of our proposed bootstrap tests under the true co-integrating rank, means that the sequential procedure of Johansen (1992), as outlined in footnote 1, applied to the bootstrap Q_r test will, unlike the corresponding procedure for the standard Q_r test (cf. Remark 3.2), correctly select the true co-integrating rank with probability $(1-\xi)$ in large samples even in the presence of non-stationary volatility satisfying Assumption 2.

Remark 4.8. Notice that Theorem 3 does not show that the wild bootstrap is able to provide an asymptotic refinement, but simply that it is able to *retrieve* the true asymptotic distribution of the reference test statistic under the null hypothesis. Indeed, one would not expect to be able to achieve any asymptotic refinement in this case because the asymptotic distribution of Q_r is non-pivotal under Assumption 2 (cf. Theorem 1). For similar results in the univariate (unit root) case see Cavaliere and Taylor (2008a).

Remark 4.9. Given the results in Theorem 3, it follows straightforwardly that the limiting null distribution of the bootstrap maximum eigenvalue statistic, $Q^b_{r,\max}$, coincides with that given in Remark 3.4, so that again our wild bootstrap procedure will deliver (asymptotically) correctly sized maximum eigenvalue co-integration tests under Assumption 2.

5 Finite Sample Simulations

In this section we use Monte Carlo simulation methods to compare the finite sample size and power properties of the PLR co-integration rank test of Johansen (1991, 1996) with its wild bootstrap version proposed in Section 4. We also compare the properties of the sequential approach of Johansen (1992) when applied using the PLR test and its bootstrap analogue.

As in Johansen (2002) and Swensen (2006), we consider as our simulation DGP an $I(1)$ $VAR(1)$ process where we set the dimension of the VAR process to $p=5$, and consider both the case of no co-integration ($r=0$) and of a single co-integrating vector ($r=1$). In the $r=1$ case, the DGP we use is of the form

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t,\ \varepsilon_t := \sigma_t z_t$$

where α and β are $p\times 1$ vectors and $z_t := (z_{1,t},...,z_{p,t})'$ is a p-dimensional Gaussian process with mean zero and covariance matrix I_p. As in Remark 2.3, we make the following assumption on the volatility term: $\sigma_t = V_t$, with $V_t := \mathsf{diag}(V_{1,t},...,V_{p,t})$ a time-varying diagonal matrix initialized at $V_0 := I_p$. Moreover, as in Johansen (2002) and Swensen (2006) we consider DGPs with $\beta := (1,0,...,0)'$ and $\alpha := (a_1, a_2, 0, ..., 0)'$. This leads to the model

$$\begin{aligned}
\Delta X_{1,t} &= a_1 X_{1,t-1} + V_{1,t} z_{1,t} \\
\Delta X_{2,t} &= a_2 X_{1,t-1} + V_{2,t} z_{2,t} \\
\Delta X_{i,t} &= \qquad\qquad V_{i,t} z_{i,t}, \qquad i = 3,...,p
\end{aligned}$$

with $|1+a_j|<1$, $j=1,2$. In the $r=0$ case, the model reduces to the multivariate random walk with serially uncorrelated but heteroskedastic innovations,

$$\Delta X_t = \varepsilon_t,\ \varepsilon_t := V_t z_t.$$

The simulation model considered above therefore generalises that used by previous authors in that we are allowing the volatility V_t to vary over time rather than being constant. In particular, in what follows we will consider the case where the volatility of each element of $\varepsilon_t := (\varepsilon_{1t}, ..., \varepsilon_{pt})'$ may display a one time change. Corresponding results for other non-stationary volatility models such as those considered in Cavaliere (2004), Cavaliere and Taylor (2007) did not yield qualitatively different results from those presented here for the one time change case and are consequently omitted in the interests of brevity.

We consider five different cases, according to the how many of the shocks ε_{it} display a one time change in volatility. In the j-th heteroskedastic model, $j = 1, ..., p$, we assume that

$$\begin{aligned} V_{i,t} &= v_t \text{ for } i = 1, ..., j \\ V_{i,t} &= 1 \text{ for } i = j+1, ..., p. \end{aligned}$$

Hence, we are implicitly assuming that the heteroskedastic shocks display a common volatility process, v_t. As regards v_t, we consider the case where volatility displays either a positive (v_t switches from 1 to $\delta > 1$) or a negative (v_t switches from 1 to $\delta < 1$) shift at time $T^* = \lfloor \tau T \rfloor$. In order to limit the number of experiments, we vary the break fraction among $\tau \in \{0, 1/3, 2/3\}$ and the break magnitude among $\delta \in \{1/3, 3\}$. The case $\tau = 0$ indicates that the volatility is constant over time. The values of $\tau = 1/3$ and $\tau = 2/3$ therefore allow for either an early or a late volatility shift, while $\delta = 3$ and $\delta = 1/3$ allow the magnitude of the volatility shift to be either positive (of size 3 standard deviations) or negative (of size 1/3 standard deviation). Notice that the chosen values of δ are empirically relevant; for example, the size of the late negative volatility shift in U.S. real GDP growth reported in McConnell and Perez-Quiros (2000, Table III) was found to be $\delta = 1/2.5$, while Kahn *et al.* (2002, Table 1) show that the variability (standard deviation) of the core CPI in the 69:1–83:4 period was three times larger ($\delta = 1/3$) than in the 84:1–00:4 period. Although not reported here we also experimented with other values of δ. As might be expected, values of δ further from (closer to) unity give rise to larger (smaller) size distortions in the standard co-integration tests than those reported.

The reported simulations were programed using the **rndKMn** function of Gauss 7. All experiments were conducted using $10,000$ replications. The sample sizes were chosen within the set $\{100, 200, 400\}$ and the number of replications used in the wild bootstrap algorithm was set to 399. All tests were conducted at the nominal 0.05 significance level. No deterministics were included in the estimation. For the standard PLR tests we employed asymptotic critical values as reported in Table 15.1 of Johansen (1996).

5.1 The Non-Co-Integrated Model ($r = 0$)

Table 1 reports the finite sample (empirical) size properties of both the standard PLR Q_0 test and its bootstrap analogue test for $H(0) : r = 0$ against $H(5) : r = 5$, in the presence of a one time change in volatility occurring in either none of the series, one of the series, and so on through to the case where all five series display a break in volatility. Table 2 reports the corresponding properties of the sequential procedures of Johansen (1992) using the standard and bootstrap Q_r ($r = 0, ..., 4$) tests, as described in footnote 1 with $\xi = 0.05$.

Tables 1 – 2 about here

Consider first the results in Table 1. Even in the absence of a break in volatility it can be seen from the first panel of Table 1 that the bootstrap Q_0 test displays finite sample size closer to the nominal level than the standard test based on asymptotic critical values; for example, while the standard test has size of 8.4% for $T = 100$, the bootstrap test has size of 5.4%. It is, however, where shifts in volatility occur that the benefits of the bootstrap test become most apparent. The results in the second panel of Table 1 show that under a one time change in volatility the standard Q_0 test displays very unreliable size properties. The size distortions seen in the standard test are worse, other things being equal, for early negative and late positive *vis-à-vis* late negative and early positive changes, and worsen as the number of elements of ε_t that display a break in volatility increases. For example, while an early positive break in ε_{1t} only effects only a modest size inflation to 6.8%, an early negative break in all of the elements of ε_t yields a massive inflation of size to 63.8%, in each case for $T = 400$. For a given break, notice also that the size distortions in the standard Q_0 test do not change very much as the sample size increases, suggesting that the asymptotic distribution theory given in Theorem 1 provides a useful predictor for the finite sample behaviour of Q_0 under a break in volatility. In contrast, the size properties of our bootstrap Q_0 test seem largely satisfactory throughout. A small degree of finite sample oversize is seen with the bootstrap tests for early negative and late positive breaks in all of the elements of ε_t. However, this has all but gone by $T = 400$ and even for the smaller sample sizes considered still represents an enormous improvement on the size properties of the standard test.

Notice also from the second panel of Table 1 that the results for a late positive break are quite similar throughout to those for an early negative break. Similarly, the results for an early positive break are similar to those for a late negative break throughout. These similarities were also observed in the results reported in Tables 2-4 and so to avoid unnecessary duplication in what follows we only report results for early negative and late negative shifts.

Consider next the results in Table 2. Since all of the tests were run at the 5% significance level, in the constant volatility case both the standard and bootstrap sequential procedures should select $r = 0$ with probability 95% and $r > 0$ with probability 5%. Consistent with the results in Table 1, we see from the first panel of Table 2 that under constant volatility the procedure based on the bootstrap PLR tests gets considerably closer to these proportions in small samples than the procedure based on the standard PLR tests. Where volatility is non-constant, the procedure based on the standard PLR test performs very poorly indeed, as can be seen from the results in the second panel of Table 2. For example, in the presence of an early volatility shift in all of the elements of ε_t the standard procedure only selects the correct co-integrating rank 36.2% if time even for $T = 400$; indeed, 17.6% of time it will indicate conclude that the true co-integrating rank is two. In contrast, the procedure based on the bootstrap PLR tests appears to perform very well in practice, with its empirical probability of selecting the true co-integrating rank of zero converging rapidly towards 95 % throughout; cf. Remark 4.7. In the same example as above, the bootstrap-based procedure selects the true co-integrating rank 93.5% of the time.

5.2 The Co-Integrated Model ($r = 1$)

In the $r = 1$ case, observe that the matrix $\alpha'_{\perp}$ is given by

$$\alpha_{\perp}^{'} = \begin{pmatrix} -a_2/a_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

which clearly implies that, in the case where $a_2 = 0$, the common trends depend on the vector $(\varepsilon_{2t}, ...\varepsilon_4)'$ but not on ε_{1t}. Consequently, it is to be expected that under $a_2 = 0$ variance shifts in ε_{1t} will not affect the rejection frequency of the rank test, at least in samples of sufficiently large size. Conversely, when $a_2 \neq 0$, the common trends depend on the whole vector ε_t. Also, notice that since $\beta' X_t$ depends on ε_{1t} only, the co-integrating relation $\beta' X_t$ is stationary even in the presence stationary volatility shifts in $(\varepsilon_{2t}, ...\varepsilon_{5t})'$.

Tables 3 – 5 about here

Consider first the results in Table 3 for the standard Q_1 test and its bootstrap analogue. The results in the first panel for the case where volatility is constant show that the bootstrap Q_1 test displays good size properties regardless of the value of a_2. The standard Q_1 test is a little oversized for $a_2 = -0.4$ when $T = 100$ but otherwise displays good size. The results in the second panel of Table 3 show that, as with the corresponding results in Table 1, the standard PLR test does not display anything like adequate size control in the presence of breaks in volatility. The one exception occurs, as predicted, where $a_2 = 0$ and there is a shift in ε_{1t} only, and here the standard PLR test is size controlled. However, where breaks occur in the other elements of ε_t the size of the PLR test exceeds the nominal level, with these distortions worsening with the number of elements of ε_t display a break. These distortions are larger, other things equal, for $a_2 = -.04$, as might be expected given the fact that a break in ε_{1t} has no impact on the tests in this case. In contrast to the standard test, the bootstrap Q_1 test displays excellent size control throughout, with only two cases occurring in the whole table (both for $T = 100$) were the size exceeds 7%.

Tables 4 and 5 report corresponding results for the sequential procedure of Johansen (1992) for $a_2 = 0$ and $a_2 = -0.4$, respectively. Since now the co-integrating rank is one, in the constant volatility case both the standard and bootstrap procedures should select $r = 0$ with probability 0%, $r = 1$ with probability 95% and $r > 1$ with probability 5%. Again this is pretty much the case as the results in the first panel of both Tables 4 and 5 show. While these proportions are largely maintained by the bootstrap-based procedure in the second panel of Tables 4 and 5, the same cannot be said for the procedure based on the standard PLR tests, which as with the corresponding results in Table 2 has a strong tendency to over-estimate the co-integrating rank, even in large samples. For example, under an early volatility shift affecting each element of ε_t, with $a_2 = -0.4$ and $T = 400$, the standard procedure selects the true co-integrating rank of one 53% of the time, a rank of two 36.2% of the time and a rank of three 9.6% of the time. In contrast, the bootstrap procedure picks the true rank 95% of the time, a rank of two 4.4% of the time and a rank of three 0.5% of the time. It is also interesting to also note that in small samples the standard procedure displays a lesser tendency to under-estimate the true co-integration rank than the bootstrap procedure - for example, for $T = 100$, $a_2 = 0$, and an early negative shift in the first four elements of ε_t, the standard procedure selects a co-integrating rank equal to zero only 12.7% of the time, while the bootstrap procedure does so 61.7% of the time. This result is of course an artefact of the uncontrolled size of the standard Q_0 test, this test in fact having size of 45.5% in this case; cf. Table 1.

6 Empirical Illustration

In this section we illustrate the methods discussed in this paper with a short application to the term structure of interest rates; see Campbell and Shiller (1987) for an early reference. According to traditional theory, aside from a constant or stationary risk premium, long-term interest rates are an average of current and expected future short term rates over the life of the investment. Hence, provided interest rates are well described as $I(1)$ variables, bond rates at different maturities should be driven by a single common stochastic trend, with the spreads between rates at different maturities being stationary. Although early studies tend to corroborate this view, see, for example, Hall *et al.* (1992), more recent research, based on broader sets of maturities, suggests that yields are better characterised by *more* than one common trend, reflecting possible non-stationarities in the risk premia and additional risk factors, such as the slope and curvature of the yield curve; see, for example, the discussion in Giese (2006).

We consider monthly data of U.S. treasury zero-coupon yields, say $R_t^{(n)}$ where n denotes the maturity, with $n = 1$ (one-month), 3 (three-months), 12 (one year), 24 (two years) and 60 (5 years). The sample data cover the period 1970:1–2000:12, thereby considerably extending the 1970:1–1988:12 sample used by Hall *et al.* (1992); see Giese (2006) for further details on the data. Yield levels are displayed in the upper panel of Figure 1, with the corresponding first differences displayed in the middle panel of the figure.

Figure 1 and Table 6 about here

Let $X_t := (R_t^{(1)}, R_t^{(3)}, R_t^{(12)}, R_t^{(24)}, R_t^{(60)})'$. As is standard, we fit a VAR model for X_t with restricted intercept; that is, $D_{2t} = 0$ and $D_{1t} = 1$ in (3). The VAR is estimated using Gaussian maximum likelihood under the assumption of constant volatility; cf. Section 3. The number of lags was set to $k = 4$.

The final panel of Figure 1 reports estimates of the *variance profiles*[2] of the five unrestricted residual series from the estimated VAR(4) model. That is,

$$\hat{\eta}_i(u) := \frac{\sum_{t=1}^{\lfloor Tu \rfloor} \hat{\varepsilon}_{it}^2}{\sum_{t=1}^{T} \hat{\varepsilon}_{it}^2},\ u \in [0,1]$$

for each $i = 1, ..., 5$; see Cavaliere and Taylor (2007) for further details. Should ε_{it} have constant volatility, so the corresponding estimate of the variance profile should approximately

[2]For a given univariate time series e_t satisfying Assumption 2; that is, $e_t = \sigma_t z_t$ with $\sigma_{\lfloor Tu \rfloor} := \sigma(u) \in \mathcal{D}_{\mathbb{R}}[0,1]$ and z_t iid $(0,1)$, Cavaliere and Taylor (2007) define the variance profile as $\eta(u) := (\int_0^1 \sigma(s)^2 ds)^{-1} \int_0^u \sigma(s)^2 ds$, $u \in [0,1]$. The variance profile satisfies $\eta(u) = u$, all u, under homoskedasticity while it deviates from u in the presence of non-stationary volatility.

follow the 45° line. A quick inspection of Figure 1, however, suggests that this does not seem to be the case here, with the estimated variance profiles reflecting the increased variability in month-to-month yield changes observed in the late 1970s and early 1980s. Furthermore, according to the tests proposed in Cavaliere and Taylor (2008b), the deviations of the estimated variance profiles from the 45° line are all statistically significant indicating the presence of non-stationary volatility effects in the data. Notably, the shape of the estimated variance profiles are also consistent with the findings of Hansen (2003) who argues for the presence of two shifts (the first in September 1979 and the second in October 1982) in the covariance matrix of a system containing monthly U.S. treasury zero-coupon yields with maturities of 1,3,6,9,12,60 and 84 months measured over the period 1970:1-1995:12. Notice also that the five estimated variance profiles display similar dynamics; as documented in the Monte Carlo study in section 5, the presence of common volatility dynamics among the errors of the VAR process is likely to inflate evidence in favour of co-integration.

Table 6 reports the results of the standard and bootstrap co-integration rank tests. For the standard tests p-values were computed as suggested in MacKinnon, Haug and Michelis (1999). The standard sequential procedure detects three co-integrating relation at any conventional significance level, with a fourth co-integration relation being significant at the 5% level with a p-value of about 3%. Although this result seems to corroborate the traditional view of the expectation hypothesis of the term structure, it may in fact hide the presence of additional risk factors.

To shed further light on this issue, and given that the presence of non-stationary volatility in the data is likely to inflate the evidence in favour of co-integration, we make use of the wild bootstrap approach proposed in Section 4 in order to obtain heteroskedasticity-robust p-values for the trace test statistics. Using the wild bootstrap algorithm with $k^b = k = 4$ and 399 bootstrap replications, we obtain the p-values reported in the last column of Table 6.

The wild bootstrap p-values indicate that the evidence of a single common trend underlying the term structure is much weaker than is the case when using standard p-values. Using a 5% significance level there is now a clear indication of two common trends underlying the dynamics of the five yields considered. This result, which is in line with the recent findings of Giese (2006), consequently provides further support in favour of recent multifactor theories of the term structure; see, for example, Diebold, Ji and Li (2007).

7 Conclusions

In this paper we have shown that non-stationary behaviour in the unconditional volatility of the innovations has potentially serious implications for the reliability of tests for co-integration based on the trace and maximum eigenvalue statistics of Johansen (1988,1991). We have shown that in such cases the limiting null distributions of these statistics depend on the asymptotic integrated covariation of the underlying volatility process, thereby effecting tests whose true size can be significantly in excess of the nominal significance level when using conventional critical values. In order to rectify this problem, we have proposed a wild bootstrap-based approach to testing for co-integration rank. Our proposed bootstrap co-integration rank tests have the considerable advantage that they are not tied to a given parametric model of volatility. The proposed wild bootstrapping scheme was shown to deliver co-integration rank statistics which share the same first-order limiting null distributions as the corresponding standard co-integration statistics, confirming the asymptotic validity of our bootstrap tests within the class of non-stationary volatility considered. Monte Carlo evidence was reported for the case of a one time change in volatility which suggested that the proposed bootstrap co-integration tests perform well in finite samples avoiding the large over-size problems that can occur with the standard tests, the latter being worse, *ceteris paribus*, where the volatility shifts were common across the individual series. An empirical application to the term structure of interest rates was also reported which suggested the presence of two common trends in U.S. government bond yields over different maturities, consistent with recent multifactor theories of the term structure.

Acknowledgements

We thank Bruce Hansen and conference participants at the Tinbergen Institute Conference '20 Years of Cointegration: Theory and Practice in Prospect and Retrospect', Rotterdam, March 23–24, 2007 for useful feedback. We are also grateful to Julia Giese for having provided us the U.S. bond yields data. Giuseppe Cavaliere and Anders Rahbek thank the Danish Social Sciences Research Council, Project Number 2114-04-001, for financial support.

A Appendix

A.1 Preliminary Lemmata

Throughout we let $\sigma^{*2} := \sup_t \|\Sigma_t\|$, where $\sigma^{*2} < \infty$ due to the càdlàg assumption on the volatility process $\Sigma_t := \sigma_t \sigma_t'$.

Consider the p-dimensional heteroskedastic VAR processes:

$$\begin{aligned} Y_t &= A_1 Y_{t-1} + \ldots + A_m Y_{t-m} + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \\ X_t &= B_1 X_{t-1} + \ldots + B_n X_n + \varepsilon_t, \end{aligned} \tag{A.1}$$

with z_t i.i.d.$(0, I)$, symmetric and with finite fourth order moment. The corresponding characteristic polynomials are denoted as $A(z) = 1 - A_1 z - \ldots - A_m z^m$ and $B(z) = 1 - B_1 z - \ldots - B_n z^n$ respectively. The processes are well-defined for $t = 1, .., T$ with initial values $\mathbb{Y}_0 = (Y_0', Y_{-1}', ..., Y_{-m+1}')'$ and $\mathbb{X}_0 = (X_0', X_{-1}', ..., X_{-n+1}')'$. Then we have:

Lemma A.1 *Consider the VAR heteroskedastic processes Y_t and X_t defined in (A.1), where the roots of* $\det|A(z)| = 0$ *and* $\det|B(z)| = 0$ *are all assumed to lie outside the unit circle. Then as $T \to \infty$, for $k \geq 0$,*

$$\frac{1}{T} \sum_{t=1}^{T} Y_t X_{t+k}' \xrightarrow{p} \sum_{i=0}^{\infty} \Theta_i \bar{\Sigma} \Gamma_{i+k}' \tag{A.2}$$

where $\bar{\Sigma} := \int_0^1 \Sigma(s)\, ds$, and Θ_i and Γ_i are the coefficients obtained by inversion of the characteristic polynomials $A(z)$ and $B(z)$ respectively.

Proof. Rewrite (A.2) as

$$\frac{1}{T}\sum_{t=1}^{T} Y_t X'_{t+k} = \frac{1}{T}\sum_{t=1}^{T} EY_t X'_{t+k} + \frac{1}{T}\sum_{t=1}^{T}\left(Y_t X'_{t+k} - E\left(Y_t X'_{t+k}\right)\right). \tag{A.3}$$

We split the proof into two parts. In Part I we show that the first term converges as desired to $\sum_{i=0}^{\infty}\Theta_i\bar{\Sigma}\Gamma'_{i+k}$. Next, in Part II we show that the second term on the right hand side of (A.3) converges in probability to zero; that is,

$$\frac{1}{T}\sum_{t=1}^{T}\left(Y_t X'_{t+k} - E\left(Y_t X'_{t+k}\right)\right) \xrightarrow{p} 0. \tag{A.4}$$

Part I. With $\mathbb{Y}_t = \left(Y'_t, \dots, Y'_{t-m}\right)'$ then $\mathbb{Y}_t = \mathbb{A}\mathbb{Y}_{t-1} + e_t$ where the spectral radius of $\mathbb{A}$ is smaller than one, $\rho(\mathbb{A}) < 1$, by assumption. Likewise for the VAR(n) process X_t. Hence, we may restrict attention to the case of VAR processes of order one. Therefore, let Y_t and X_t solve the VAR(1) equations $Y_t = AY_{t-1} + \varepsilon_t$ and $X_t = BX_{t-1} + \varepsilon_t$, with solutions,

$$Y_t = \sum_{i=0}^{t-1} A^i\varepsilon_{t-i} + A^t Y_0, \qquad X_t = \sum_{i=0}^{t-1} B^i\varepsilon_{t-i} + B^t X_0. \tag{A.5}$$

Inserting these solutions gives,

$$\frac{1}{T}\sum_{t=1}^{T} EY_t X'_{t+k} = \frac{1}{T}\sum_{t=1}^{T} A^t Y_0 X'_0 B^{t+k\prime} + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=0}^{t-1} A^i\Sigma_{t-i}B^{k+i\prime}.$$

Now, $\left\|\frac{1}{T}\sum_{t=1}^{T} A^t Y_0 X'_0 B^{t+k\prime}\right\| = O\left(\frac{1}{T}\right)$ since $\rho(A)$ and $\rho(B) < 1$ implies in particular that $\left\|A^t\right\|\left\|B^t\right\| \leq c\rho^t$ for some $0 < \rho < 1$ and constant c as $t \to \infty$. Next,

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{i=0}^{t-1} A^i\Sigma_{t-i}B^{k+i\prime} = \frac{1}{T}\sum_{j=1}^{T}\sum_{t=j}^{T} A^{t-j}\Sigma_j B^{t+k-j\prime}$$

and the results holds by,

$$\begin{aligned}
&\mathsf{vec}\left(\frac{1}{T}\sum_{j=1}^{T}\sum_{t=j}^{T} A^{t-j}\Sigma_j B^{t+k-j\prime}\right) = \frac{1}{T}\sum_{j=1}^{T}\sum_{t=j}^{T}\left(A^{t-j}\otimes B^{t+k-j}\right)\mathsf{vec}\left(\Sigma_j\right) \\
&= \frac{1}{T}\sum_{j=1}^{T}\sum_{i=0}^{T-j}\left(A^i\otimes B^{k+i}\right)\mathsf{vec}\left(\Sigma_j\right) \\
&= \frac{1}{T}\sum_{j=1}^{T}\sum_{i=0}^{\infty}\left(A^i\otimes B^{k+i}\right)\mathsf{vec}\left(\Sigma_j\right) + \frac{1}{T}\sum_{j=1}^{T}\left(\sum_{i=T-j+1}^{\infty}\left(A^i\otimes B^{k+i}\right)\right)\mathsf{vec}\left(\Sigma_j\right) \\
&= \sum_{i=0}^{\infty}\left(A^i\otimes B^{k+i}\right)\mathsf{vec}\left(\bar{\Sigma}\right) + O\left(\frac{1}{T}\right) = \mathsf{vec}\left(\sum_{i=0}^{\infty} A^i\bar{\Sigma}B^{k+i\prime}\right) + O\left(\frac{1}{T}\right)
\end{aligned}$$

where for the $O\left(\frac{1}{T}\right)$ term

$$\left\| \frac{1}{T} \sum_{j=1}^{T} \left(\sum_{i=T-j+1}^{\infty} \left(A^i \otimes B^{k+i}\right) \right) \mathsf{vec}\left(\Sigma_j\right) \right\| \leq \sigma^{*2} c/T$$

as $\rho(A \otimes B) = \rho(A)\,\rho(B)$. Finally, note that $\Theta_i = (I_p, 0, \ldots, 0)\,\mathbb{A}^i\,(I_p, 0, \ldots, 0)'$ and likewise $\Gamma_i = (I_p, 0, \ldots, 0)\,\mathbb{B}^i\,(I_p, 0, \ldots, 0)'$ with

$$\mathbb{A} = \begin{pmatrix} A_1 & A_2 & \cdots & A_m \\ I_p & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & & I_p & 0 \end{pmatrix}, \qquad \mathbb{B} = \begin{pmatrix} B_1 & B_2 & \cdots & B_n \\ I_p & 0 & \cdots & 0 \\ & \ddots & & \vdots \\ 0 & & I_p & 0 \end{pmatrix}.$$

Part II. As in Part 1, we consider the VAR(1) case and without loss of generality we set $Y_t = X_t =: U_t$. Hence we establish that

$$\frac{1}{T} \sum_{t=1}^{T} \left(U_t U_{t+k}' - E\left(U_t U_{t+k}'\right)\right) \xrightarrow{p} 0 \tag{A.6}$$

with U_t a heteroskedastic VAR process of order one, $U_t = AU_{t-1} + \varepsilon_t$. Re-write (A.6) as,

$$\frac{1}{T} \sum_{t=1}^{T} \left(U_t U_{t+k} - E\left(U_t U_{t+k}\right)\right) = \frac{1}{T} \sum_{t=1}^{T} \left(U_t U_t' - E\left(U_t U_t'\right)\right) A^{k\prime} + \frac{1}{T} \sum_{t=1}^{T} \left(U_t \left(\sum_{i=0}^{k-1} A^i \varepsilon_{t+k-i} \right)' \right).$$

The last term tends to zero by the LLN for martingale differences since $\|\Sigma_t\| = \sigma^{*2} < \infty$. Next, for any $\lambda \in \mathbb{R}^p$ set $V_T \equiv \frac{1}{T} \sum_{t=1}^{T} \lambda' \left(U_t U_t' - E\left(U_t U_t'\right)\right) \lambda$,

$$P\left(|V_T| > \delta\right) \leq E\left|V_T\right|^2 / \delta^2 = \mathrm{Var}\left(V_T\right) / \delta^2 = \frac{1}{\delta} \mathrm{Var}\left(\frac{1}{T} \sum_{t=1}^{T} \left(\lambda' U_t\right)^2 \right)$$

and with $Z := \left(\lambda' U_1, ..., \lambda' U_T\right)'$, $C := \frac{1}{T} I_T$, we thus need to show that

$$\mathrm{Var}\left(Z' C Z\right) \to 0. \tag{A.7}$$

By simple recursion,

$$\lambda' U_t = \sum_{i=0}^{t-1} \lambda' A^i \varepsilon_{t-i} + \lambda' A^t U_0.$$

and, hence,

$$EZ := \xi = \lambda' \left(AU_0, ..., A^T U_0\right) \neq 0.$$

From the identity, where C is symmetric,

$$\text{Var}\left(Z'CZ\right) \equiv \text{Var}\left((Z-\xi)'C(Z-\xi)\right) + 4\text{Var}\left(\xi'CZ\right) + 2\text{Cov}\left((Z-\xi)'C(Z-\xi), 2\xi'CZ\right),$$

it follows that (A.7) holds if,

$$\text{Var}\left((Z-\xi)'C(Z-\xi)\right) \to 0 \quad \text{and} \quad \text{Var}\left(\xi'CZ\right) \to 0.$$

If Z is Gaussian, then $\text{Var}\big((Z-\xi)'C(Z-\xi)\big) = 2\,\mathsf{tr}\,(C\Omega C\Omega)$, where $\text{Var}(Z) =: \Omega$. If Z is not Gaussian but symmetric with finite fourth order moment κ, then

$$Var\left((Z-\xi)'C(Z-\xi)\right) \leq 2\max\left(1, \frac{\kappa}{3}\right)\mathsf{tr}\,(C\Omega C\Omega).$$

Next,

$$\begin{aligned}
\mathsf{tr}\,(C\Omega C\Omega) &= \frac{1}{T^2}\mathsf{tr}\left(\Omega^2\right) = \frac{1}{T^2}\mathsf{tr}\left(\text{Var}\left(Z\right)\right)^2 \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\text{Cov}\left(\lambda'U_t, \lambda'U_s\right)\right)^2 \\
&= \frac{1}{T^2}\left[\sum_{t=1}^{T}\text{Var}\left(\left(\lambda'U_t\right)^2\right) + 2\sum_{t=1}^{T}\sum_{j=1}^{T-t}\text{Cov}(\left(\lambda'U_t\right)^2, \left(\lambda'U_{t+j}\right)^2)\right] \\
&= \frac{1}{T^2}\left[\lambda'\sum_{i=0}^{t+r-1}A^i\sigma^2_{t+r-i}A^{i\prime}\lambda + 2\lambda'\sum_{i=0}^{t-1}A^i\Sigma_{t-i}A^{i\prime}\left(A^j\right)'\lambda\right] \\
&\leq \frac{c}{T^2}\left[\sum_{t=1}^{T}\sum_{s=t}^{T}\left(\sum_{i=1}^{\min(t,s)}\left\|A^{s+t-2i}\right\|\right)^2\right] = O\left(\frac{1}{T}\right).
\end{aligned}$$

Here we have used the fact that $\frac{1}{T}\sum_{t=1}^{T}\left\|A^t\right\|^2 = O(1)$ as $\left\|A^t\right\| = O\left(|\rho|^t\right)$, with $|\rho| < 1$ the spectral radius of A. We can therefore conclude that $\text{Var}\big((Z-\xi)'C(Z-\xi)\big) \to 0$. Likewise,

$$Var\left(\xi'CZ\right) = \xi'C\Omega C'\xi \leq \frac{c}{T^2}\sum_{t=1}^{T}\sum_{s=t}^{T}\sum_{i=1}^{t}\left\|A^{s+t-i}\right\|^2 = O\left(\frac{1}{T}\right).$$

This concludes Part II. □

A.2 Proof of Lemmas 1-2, Theorems 1-2 and Related Lemmas

Proof of Lemma 1: Without loss of generality, set $\mu D_t = 0$. With $\mathbb{X}_t := \left(X'_t, ..., X'_{t-k+1}\right)'$ the system can be written in companion form as,

$$\Delta\mathbb{X}_t = \mathbb{A}\mathbb{B}'\mathbb{X}_{t-1} + e_t$$

with $e_t := (\varepsilon'_t, 0, ..., 0)'$, $\mathbb{X}_0$ fixed and

$$\mathbb{A} := \begin{pmatrix} \alpha & \Gamma_1 & \Gamma_2 & ... & \Gamma_{k-1} \\ 0 & I_p & 0 & ... & 0 \\ 0 & 0 & I_p & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & I_p \end{pmatrix} \quad \mathbb{B} := \begin{pmatrix} \beta & I_p & 0 & ... & 0 \\ 0 & -I_p & I_p & ... & 0 \\ 0 & 0 & -I_p & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & ... & -I_p \end{pmatrix}. \tag{A.8}$$

Note that with $\mathbb{X}_{\beta t} := \mathbb{B}'\mathbb{X}_t$,

$$\begin{aligned}\mathbb{X}_{\beta t} &= \left(I_{r+p(k-1)} + \mathbb{B}'\mathbb{A}\right)\mathbb{X}_{\beta t-1} + \mathbb{B}'e_t \\ &= \sum_{i=0}^{t-1}\left(I_{r+p(k-1)} + \mathbb{B}'\mathbb{A}\right)^i \mathbb{B}'e_{t-i} + \left(I_{r+p(k-1)} + \mathbb{B}'\mathbb{A}\right)^t \mathbb{X}_{\beta 0}.\end{aligned} \tag{A.9}$$

Using this, standard arguments and recursions give,

$$\mathbb{X}_t = \mathbb{C}\sum_{i=1}^{t} e_i + \mathbb{S}_t + \mathbb{C}\mathbb{X}_0,$$

where $\mathbb{C} = \mathbb{B}_\perp(\mathbb{A}'_\perp\mathbb{B}_\perp)^{-1}\mathbb{A}'_\perp$, and $\mathbb{S}_t = \mathbb{A}\left(\mathbb{B}'\mathbb{A}\right)^{-1}\mathbb{X}_{\beta t}$. It follows that as $X_t = (I_p, 0, ..., 0)\,\mathbb{X}_t$, then X_t can be written as

$$X_t = C\sum_{i=1}^{t}\sigma_i z_i + S_t + C(I_p, -\Psi)\mathbb{X}_0 \tag{A.10}$$

where $S_t = P\mathbb{X}_{\beta t}$, where $\mathbb{X}_{\beta t}$ is the heteroskedastic VAR process of order one given by (A.9) with autoregressive coefficient $\left(I_{r+p(k-1)} + \mathbb{B}'\mathbb{A}\right) =: R$, for which $\rho\,(R) < 1$ by Assumption 1. The matrix P is of dimension $p \times (r + p(k-1))$ and is given by

$$P := (I_p, 0, ..., 0)'\,\mathbb{A}\left(\mathbb{B}'\mathbb{A}\right)^{-1} = (\alpha, \Psi)\,Q$$

for $Q := (\mathbb{B}'\mathbb{A})^{-1}$. Note that the singular variance of $\mathbb{B}'e_t$ equals,

$$\Omega_t := \mathrm{Var}\left(\mathbb{B}'e_t\right) = (\beta, I_p, 0, ..., 0)'\,\Sigma_t\,(\beta, I_p, 0, ..., 0)\,,$$

and in particular,

$$\|\Omega_t\| \le c\,\|\Sigma_t\| = c\sigma^{*2}\ .$$

Note also that,

$$\mathbb{A}_\perp = (I_p, -\Gamma_1, ..., -\Gamma_{k-1})'\,\alpha_\perp, \quad \mathbb{B}_\perp = (I_p, ..., I_p)'\,\beta_\perp,$$

from which the various expressions follow by simple algebraic identities. □

Proof of Lemma 2**:** The convergence to $M\,(\cdot)$ follows as in Lemma 1 of Cavaliere and Taylor (2006), while the convergence to the stochastic integral holds by Theorem 2.1 in Hansen (1992b) as $\sup_T \left\|T^{-1}\Sigma_{t=1}^T E(\varepsilon_t\varepsilon_t')\right\| = \sup_T \left\|T^{-1}\Sigma_{t=1}^T \sigma_t\sigma_t'\right\| \le \sigma^{*2} < \infty$ under Assumption 2. □

Proof of Theorem 1: To prove Theorem 1 proceed as in Hansen, Dennis and Rahbek (2002) by initially proving Lemma A.2 and Lemma A.3 below.

Lemma A.2 *Under the assumptions of Theorem 1 and Assumption 2, and with $\mu = 0$, then as $T \to \infty$,*

$$S_{00} \xrightarrow{p} \bar{\Sigma}_{00},\ \beta' S_{10} \xrightarrow{p} \bar{\Sigma}_{\beta 0} \text{ and } \beta' S_{11}\beta \xrightarrow{p} \bar{\Sigma}_{\beta\beta}\ . \tag{A.11}$$

In terms of these the following identities hold,

$$\bar{\Sigma}_{00} = \alpha\bar{\Sigma}_{\beta 0} + \bar{\Sigma},\ \ \bar{\Sigma}_{0\beta} = \alpha\bar{\Sigma}_{\beta\beta} \tag{A.12}$$

with $\bar{\Sigma} := \int_0^1 \Sigma(u)\,du$*, and finally,*

$$\bar{\Sigma}_{00}^{-1} - \bar{\Sigma}_{00}^{-1}\alpha(\alpha'\bar{\Sigma}_{00}^{-1}\alpha)^{-1}\alpha'\bar{\Sigma}_{00}^{-1} = \alpha_\perp(\alpha_\perp'\bar{\Sigma}\alpha_\perp)^{-1}\alpha_\perp'. \tag{A.13}$$

Proof: Consider $\beta' S_{10} = \beta' M_{10} - \beta' M_{12} M_{22}^{-1} M_{20}$. Using Lemma 1 and that by definition,

$$\Delta X_t = \alpha\beta' X_{t-1} + \Psi\Delta\mathbb{X}_{t-1} + \varepsilon_t, \tag{A.14}$$

the first term equals,

$$\beta' M_{10} = \frac{1}{T}\sum_{t=1}^{T}\beta' X_{t-1}\Delta X_t' = \frac{1}{T}\sum_{t=1}^{T}\beta' S_{t-1}\left((\alpha,\Psi)\,\mathbb{X}_{\beta t-1} + \varepsilon_t\right)'.$$

where $S_t = (\alpha,\Psi)\,Q\mathbb{X}_{\beta t}$. Hence Lemma A.1 implies that,

$$\beta' M_{10} \xrightarrow{p} \Omega_{\beta 0} := \beta'(\alpha,\Psi)\,Q\sum_{i=0}^{\infty}\left[\Phi^i(\beta, I_p, 0, ..., 0)'\,\bar{\Sigma}\,(\beta, I_p, 0, ..., 0)\,\Phi^{i\prime}\right](\alpha,\Psi)'.$$

Likewise the terms $\beta' M_{12}$, M_{22} and M_{20} converge in probability and we conclude that

$$\beta' S_{10} \xrightarrow{p} \bar{\Sigma}_{\beta 0} := \Omega_{\beta 0} - \Omega_{\beta 2}\Omega_{22}^{-1}\Omega_{20}.$$

Identical arguments lead to (A.11).
The identities in (A.12) follow by postmultiplying (A.14) by (the transpose of) $\beta' X_{t-1}, \Delta X_t'$ and $\Delta\mathbb{X}_{t-1}$ respectively and taking averages and using Lemma A.1 as before. To prove the identity in (A.13) use the projection identity

$$I_p = \bar{\Sigma}_{00}^{-1}\alpha(\alpha'\bar{\Sigma}_{00}^{-1}\alpha)^{-1}\alpha' + \alpha_\perp(\alpha_\perp'\bar{\Sigma}_{00}\alpha_\perp)^{-1}\alpha_\perp'\bar{\Sigma}_{00}$$

and $\alpha_\perp'\bar{\Sigma}_{00} = \alpha_\perp'\bar{\Sigma}$, see (A.12). □

Lemma A.3 *Define the* $(p-r)$*-dimensional process,*

$$G(u) := \beta_\perp' C M(u), \tag{A.15}$$

where $M(\cdot)$ *is defined in Lemma 2. Then under the conditions of Lemma 1 as* $T \to \infty$*,*

$$\frac{1}{\sqrt{T}}\beta_\perp' X_{\lfloor Tu\rfloor} \xrightarrow{w} G(u)\ , \tag{A.16}$$

$$\beta_\perp' S_{10}\alpha_\perp = \beta_\perp' S_{12}\alpha_\perp \xrightarrow{w} \int_0^1 G(s)\,dM(s)'\,\alpha_\perp\ , \tag{A.17}$$

$$\frac{1}{T}\beta_\perp' S_{11}\beta_\perp \xrightarrow{w} \int_0^1 G(s)G(s)'ds\ , \tag{A.18}$$

and furthermore,

$$\sqrt{T}\beta' S_{10}\alpha_\perp = \sqrt{T}\beta' S_{1\varepsilon}\alpha_\perp \xrightarrow{w} N_{r\times p-r}(0, \bar{\Sigma}_{\beta\beta}\otimes\alpha_\perp'\bar{\Sigma}\alpha_\perp), \tag{A.19}$$

$$\beta' S_{11}\beta_\perp \in O_p(1). \tag{A.20}$$

PROOF: Note first that (A.16) and (A.18) follow immediately from Lemma 2 using the continuous mapping theorem. To prove (A.17) note that

$$\beta_\perp' S_{1\xi} = \beta_\perp' M_{1\varepsilon} - \beta_\perp' M_{12} M_{22}^{-1} M_{21}$$

where $M_{1\varepsilon} := T^{-1}\sum_{t=1}^{T}\Delta X_t \varepsilon_t'$. Consider first $\beta_\perp' M_{1\varepsilon}$ and use the representation of X_t to see that

$$\beta_\perp' M_{1\varepsilon} = \frac{1}{T}\left(\beta_\perp' C \sum_{t=1}^{T} (\sum_{i=1}^{t-1}\varepsilon_i)\varepsilon_t' + \beta_\perp' \sum_{t=1}^{T} S_{t-1}\varepsilon_t' + \beta_\perp' C_0 \sum_{t=1}^{T}\varepsilon_t'\right),$$

which by Lemma 2, the LLN in Lemma A.1 and the fact that ε_t and ε_{t-1} are uncorrelated weakly converges to $\int_0^1 G(s)\, dM(s)'$. Next, $M_{\varepsilon 2} := T^{-1}\sum_{t=1}^{T}\varepsilon_t U_t'$ tends to zero in probability by the law of large numbers. Since $\beta_\perp' M_{12} \in O_p(1)$ and M_{22} converges in probability by the law of large numbers, we conclude that (A.17) holds.
Finally (A.19) holds by applying the central limit theorem to the martingale difference sequence $\beta' X_{t-1}\varepsilon_t'$ rewriting $S_{1\varepsilon}$ as above and using the LLN in Lemma A.1. □

Mimicking the proof of Theorem 11.1 in Johansen (1996), the results in Lemmas A.2 and A.3 imply immediately that the smallest $p-r$ solutions of the eigenvalue problem $S(\lambda) = 0$ normalized by T converge to those of the equation

$$\left|\rho \int_0^1 G(s)\, G(s)'\, ds - \int_0^1 G(s)\,(dM(s))'\,\alpha_\perp \left(\alpha_\perp' \bar{\Sigma}\alpha_\perp\right)^{-1}\alpha_\perp' \int_0^1 (dM(s))\, G(s)'\right| = 0. \tag{A.21}$$

With $\bar{\Sigma} := \int_0^1 \Sigma(s)\, ds$, define

$$\tilde{M} := \left(\alpha_\perp' \bar{\Sigma}\alpha_\perp\right)^{-1/2}\alpha_\perp' M.$$

We may then express (A.21) as,

$$\left|\rho \int_0^1 \tilde{M}(s)\,\tilde{M}(s)'\, ds - \int_0^1 \tilde{M}(s)\,(d\tilde{M}(s))' \int_0^1 (d\tilde{M}(s))\tilde{M}(s)'\right| = 0 \tag{A.22}$$

and, hence,

$$-2\log Q\left(H(r)\,|H_2(p)\right) \xrightarrow{w} \mathsf{tr}\left(\int_0^1 (d\tilde{M}(s))\tilde{M}(s)' \left(\int_0^1 \tilde{M}(s)\,\tilde{M}(s)'\, ds\right)^{-1} \int_0^1 \tilde{M}(s)\,(d\tilde{M}(s))'\right).$$

It follows that the limiting process $\tilde{M}$ is continuous on $[0,1]$, has independent (but not necessarily stationary) increments and quadratic variation at time $u \in [0,1]$ given by

$$[\tilde{M}]_u = \left(\alpha_\perp' \bar{\Sigma}\alpha_\perp\right)^{-1/2}\alpha_\perp' \left(\int_0^u \Sigma(s)\, ds\right)\alpha_\perp \left(\alpha_\perp' \bar{\Sigma}\alpha_\perp\right)^{-1/2} \;;$$

consequently, the integrated covariation on $[0,1]$ equals the identity matrix I. □

PROOF OF THEOREM 2: The result follows using Lemmas A.2 and A.3 and mimicking the proofs of Lemmas 13.1 and Theorem 13.3 in Johansen (1996).

A.3 Proof of Theorem 3 and Related Lemmas

As in Swensen (2006) we use P^* to denote the bootstrap probability and likewise E^* denotes expectation under P^*. Moreover, without loss of generality we set initial values to zero. We first introduce four lemmas which constitute the basic ingredients of the proof of Theorem 3.

Lemma A.4 *Under the conditions of Theorem 3,*

$$X_t^b = \hat{C} \sum_{i=1}^{t} \varepsilon_i^b + T^{1/2} R_t^b$$

where for all $\eta > 0$, $P^* \left(\max_{t=1,\dots,T} \left\| R_t^b \right\| > \eta \right) \to 0$ *in probability as* $T \to \infty$.

Proof : From the proof of Lemma 1 with $\mathbb{X}_t^b := \left(X_t^{b\prime}, \dots, X_{t-k+1}^{b\prime} \right)'$ and $\mathbb{X}_0^b := 0$ we find directly from (A.10) that $X_t^b = (I_p, 0, \dots, 0)\, \mathbb{X}_t^b$ has the representation,

$$X_t^b = \hat{C} \sum_{i=1}^{t} \varepsilon_i^b + T^{1/2} R_t^b$$

with $\hat{C} := (I_p, 0, \dots, 0)\, \widehat{\mathbb{B}}_\perp (\widehat{\mathbb{A}}_\perp' \widehat{\mathbb{B}}_\perp)^{-1} \widehat{\mathbb{A}}' = \hat{\beta}_\perp (\hat{\alpha}_\perp' \hat{\Gamma} \hat{\beta}_\perp)^{-1} \hat{\alpha}_\perp'$, and $R_t^b := (\hat{\alpha}, \hat{\Psi})(\widehat{\mathbb{B}}' \widehat{\mathbb{A}})^{-1} \sum_{i=0}^{t-1} (I_{pk} + \widehat{\mathbb{B}}' \widehat{\mathbb{A}})^i (T^{-1/2} \widehat{\mathbb{B}}' e_{t-i}^b)$, $\hat{\Psi} := (\hat{\Gamma}_1, \dots, \hat{\Gamma}_{k-1})$, $\widehat{\mathbb{A}}$,$\widehat{\mathbb{B}}$ being defined as $\mathbb{A}$,$\mathbb{B}$ of (A.8) with α and β replaced by the corresponding estimators $\hat{\alpha}, \hat{\beta}$, and $e_t^b := \left(\varepsilon_t^{b\prime}, 0, \dots, 0 \right)'$. Next, note that

$$\begin{aligned}
\max_{t=1,\dots,T} \left\| R_t^b \right\| &\leq \left\| (\hat{\alpha}, \hat{\Psi})(\widehat{\mathbb{B}}' \widehat{\mathbb{A}})^{-1} \sum_{i=0}^{t-1} (I_{pk} + \widehat{\mathbb{B}}' \widehat{\mathbb{A}})^i \left(T^{-1/2} \widehat{\mathbb{B}}' e_{t-i}^b \right) \right\| \\
&\leq \psi_T \max_{t=1,\dots,T} \left\| T^{-1/2} \eta_t^b \right\|
\end{aligned}$$

where $\eta_t^b = \widehat{\mathbb{B}}' e_t^b = \left(\hat{\beta}, I_p, 0, \dots, 0 \right)' \varepsilon_t^b$ and $\psi_T \xrightarrow{p} \psi$. This holds by using the established consistency of the estimators; see Theorem 2. In particular, note that for sufficiently large T we have, by continuity, that $\rho(I_{pk} + \widehat{\mathbb{B}}' \widehat{\mathbb{A}}) < 1$, which implies that $||(I_{pk} + \widehat{\mathbb{B}}' \widehat{\mathbb{A}})^i|| \leq$const.$\lambda^i$ for some $0 < \lambda < 1$. As a consequence,

$$\psi_T = \left\| (\hat{\alpha}, \hat{\Psi})(\widehat{\mathbb{B}}' \widehat{\mathbb{A}})^{-1} \right\| \left\| \sum_{i=0}^{t-1} (I_{pk} + \widehat{\mathbb{B}}' \widehat{\mathbb{A}})^i \right\| \xrightarrow{p} \psi.$$

Finally, by showing that $P^* \left(\max_{t=1,\dots,T} \left\| T^{-1/2} \eta_t^b \right\| > \eta \right)$ is of order $o_p(1)$ implies the desired, $P^* \left(\max_{t=1,\dots,T} \left\| R_t^b \right\| > \eta \right) \xrightarrow{p} 0$. This again holds if $P^* \left(T^{-1/2} \max_{t=1,\dots,T} \left\| \varepsilon_t^b \right\| > \eta \right) = o_p(1)$. But

$$\begin{aligned}
P^* \left(T^{-1/2} \max_{t=1,\dots,T} \left\| \varepsilon_t^b \right\| > \eta \right) &\leq \frac{1}{\eta^4 T^2} \sum_{t=1}^{T} E^* \left(\varepsilon_t^{b\prime} \varepsilon_t^b \right)^2 = \frac{1}{\eta^4 T^2} \sum_{t=1}^{T} \left(\hat{\varepsilon}_t' \hat{\varepsilon}_t \right)^2 E\left(z_t^4 \right) \\
&= \frac{3}{\eta^4 T} \left(\frac{1}{T} \sum_{t=1}^{T} \left(\hat{\varepsilon}_t' \hat{\varepsilon}_t \right)^2 \right) \xrightarrow{p} 0
\end{aligned}$$

as $T^{-1} \sum_{t=1}^{T} \left(\hat{\varepsilon}_t' \hat{\varepsilon}_t \right)^2 = O_p(1)$ under the assumption that ε_t has bounded fourth moment. □

Lemma A.5 *Under the conditions of Theorem 3,*

$$S_T^b(\cdot) := \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor T\cdot \rfloor} \varepsilon_t^b \xrightarrow{w}_p M(\cdot) \ .$$

PROOF: Conditionally on $\{\hat{\varepsilon}_t\}_{t=1}^T$, $S_T^b(\cdot)$ is a Gaussian process with independent increments and covariance

$$E^*\left(S_T^b(\cdot)\, S_T^b(\cdot)'\right) = \frac{1}{T} \sum_{t=1}^{\lfloor T\cdot \rfloor} \hat{\varepsilon}_t \hat{\varepsilon}_t'.$$

Consequently, the veracity of Lemma A.5 follows if $T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \hat{\varepsilon}_t\hat{\varepsilon}_t' \rightarrow \int_0^u \sigma(s)\,\sigma(s)'\,ds$ in probability, uniformly for all $u \in [0,1]$. Now, since $T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \hat{\varepsilon}_t\hat{\varepsilon}_t'$ is monotonically increasing in u and the limit function is continuous in u, it suffices to prove pointwise convergence; cf. Hansen (1999, proof of Lemma A.10). Pointwise convergence follows by noticing that

$$\frac{1}{T} \sum_{t=1}^{\lfloor Tu \rfloor} \hat{\varepsilon}_t \hat{\varepsilon}_t' = \frac{1}{T} \sum_{t=1}^{\lfloor Tu \rfloor} \varepsilon_t \varepsilon_t' + o_p(1) \ ,$$

where $T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \varepsilon_t\varepsilon_t' \rightarrow \int_0^u \sigma(s)\,\sigma(s)'\,ds$ by a simple modification of Lemma A.1 where we account for the fact that the summation is taken from 1 to $\lfloor Tu \rfloor$ with u between 0 and 1. □

Lemma A.6 *Define the $(p-r)$-dimensional process,*

$$G(u) := \beta_\perp' C M(u), \tag{A.23}$$

where $M(\cdot)$ is defined in Lemma 2. Then under the assumptions of Theorem 3,

$$\tfrac{1}{\sqrt{T}} \hat{\beta}_\perp' X_{\lfloor Tu \rfloor}^b \xrightarrow{w}_p G(u) \ , \tag{A.24}$$

$$\hat{\beta}_\perp' S_{10}^b \alpha_\perp = \hat{\beta}_\perp' S_{12}^b \alpha_\perp \xrightarrow{w}_p \int_0^1 G(s)\, dM(s)'\, \alpha_\perp \ , \tag{A.25}$$

$$\tfrac{1}{T} \hat{\beta}_\perp' S_{11}^b \hat{\beta}_\perp \xrightarrow{w}_p \int_0^1 G(s)G(s)'ds \ , \tag{A.26}$$

and furthermore,

$$\sqrt{T}\hat{\beta}' S_{10}^b \hat{\alpha}_\perp = \sqrt{T}\hat{\beta}' S_{1\varepsilon}^b \hat{\alpha}_\perp \xrightarrow{w}_p N_{r\times p-r}(0, \bar{\Sigma}_{\beta\beta} \otimes \alpha_\perp' \bar{\Sigma} \alpha_\perp), \tag{A.27}$$

$$\hat{\beta}' S_{11}^b \hat{\beta} \in O_{p^*}(1) \tag{A.28}$$

in probability as $T \rightarrow \infty$.

PROOF: Applying Lemma A.4 and Lemma A.5, the results hold by mimicking the proof of Lemma 6 in Swensen (2006). □

Lemma A.7 *Under the conditions of Theorem 3,*

$$P^*\left(\left\|S_{00}^b-\bar{\Sigma}_{00}\right\|>\eta\right) \rightarrow 0 \tag{A.29}$$
$$P^*\left(\left\|S_{01}^b\hat{\beta}-\bar{\Sigma}_{0\beta}\right\|>\eta\right) \rightarrow 0 \tag{A.30}$$
$$P^*\left(\left\|\hat{\beta}'S_{11}^b\hat{\beta}-\bar{\Sigma}_{\beta\beta}\right\|>\eta\right) \rightarrow 0 \tag{A.31}$$

in probability as $T\rightarrow\infty$*.*

Proof: We only consider the proof of (A.29), without loss of generality. Let

$$M:=\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbb{X}_t\Delta\mathbb{X}_t',\ M^b:=\frac{1}{T}\sum_{t=1}^{T}\Delta\mathbb{X}_t^b\Delta\mathbb{X}_t^{b\prime}$$

with $\mathbb{X}_t:=\left(X_t',...,X_{t-k+1}'\right)'$ and $\mathbb{X}_t^b:=\left(X_t^{b\prime},...,X_{t-k+1}^{b\prime}\right)'$. Moreover, $\bar{\Sigma}_M$ is well-defined by Lemma A.1, where $\bar{\Sigma}_M:=p\lim_{T\rightarrow\infty}M$. The stated result follows if we can show that M^b converges in probability to $\bar{\Sigma}_M$.
Initially, notice that

$$\left\|M^b-\bar{\Sigma}_M\right\|\leq\left\|M^b-\bar{\Sigma}_{M^b}\right\|+\left\|\bar{\Sigma}_{M^b}-\bar{\Sigma}_M\right\|$$

where $\bar{\Sigma}_{M^b}:=E^*\left(M^b\right)$, similar to Swensen (2006) By Lemmas 1 and A.4, it holds that $\mathbb{X}_t$ and $\mathbb{X}_t^b$ have a moving average representation of the form

$$\Delta\mathbb{X}_t=\sum_{i=0}^{t-1}\Phi^i\eta_{t-i},\ \Delta\mathbb{X}_t^b=\sum_{i=0}^{t-1}\hat{\Phi}^i\eta_{t-i}^b,$$

where $\eta_t^b=\hat{\eta}_tw_t$. As previously noted in the proof of Lemma A.4, $\left\|\Phi^i\right\|<c\lambda^i$ for some constant $c>0$ and $0<\lambda<1$, and likewise for T large enough, $\left\|\hat{\Phi}^i\right\|<c\lambda^i$.
Next, to see that $\left\|\bar{\Sigma}_{M^b}-\bar{\Sigma}_M\right\|$ tends to zero in probability, note that

$$\begin{aligned}\bar{\Sigma}_{M^b} &= E^*\left(\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\hat{\Phi}^i\eta_{t-i}^b\right)\left(\sum_{i=0}^{t-1}\hat{\Phi}^i\eta_{t-i}^b\right)'\right)\\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\hat{\Phi}^iE^*\left(\eta_{t-i}^b\eta_{t-i}^{b\prime}\right)\hat{\Phi}^{i\prime}\right)=\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\hat{\Phi}^i\hat{\eta}_{t-i}\hat{\eta}_{t-i}'\hat{\Phi}^{i\prime}\right).\end{aligned}$$

Using the arguments made in Part I of the proof of Lemma A.1, this converges in probability to $\bar{\Sigma}_M$, provided

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\eta}_t\hat{\eta}_t'-\eta_t\eta_t'\right)\xrightarrow{p}0,$$

which is implied by the consistency of the estimators, see Theorem 2.

Next consider the term $\left\|M^b - \bar{\Sigma}_{M^b}\right\|$. We have

$$
\begin{aligned}
M^b &= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\hat{\Phi}^i\eta_{t-i}^b\right)\left(\sum_{i=0}^{t-1}\hat{\Phi}^i\eta_{t-i}^b\right)' \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\hat{\Phi}^i\eta_{t-i}^b\left(\hat{\Phi}^i\eta_{t-i}^b\right)'\right) + \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i,j=0,i\neq j}^{t-1}\hat{\Phi}^i\eta_{t-i}^b\eta_{t-j}^{b\prime}\hat{\Phi}^{j\prime}\right) \\
&= M_1^b + M_2^b.
\end{aligned}
$$

First, notice that

$$
M_1^b - \bar{\Sigma}_{M^b} = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\hat{\Phi}^i\hat{\eta}_{t-i}\hat{\eta}_{t-i}'\hat{\Phi}^{i\prime}\kappa_{t-i}\right)
$$

with $\kappa_t := \left(w_t^2 - 1\right)$ being i.i.d. with mean zero and finite moments of all order. Now,

$$
\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\mathsf{vec}\left(\hat{\Phi}^i\hat{\eta}_{t-i}\hat{\eta}_{t-i}'\hat{\Phi}^{i\prime}\kappa_{t-i}\right)\right) = \frac{1}{T}\sum_{t=1}^{T}\kappa_t\left(\sum_{i=0}^{T-t}\hat{\Phi}^i\otimes\hat{\Phi}^i\right)\mathsf{vec}\left(\hat{\eta}_t\hat{\eta}_t'\right)
$$

which implies that

$$
\begin{aligned}
P^*\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=0}^{t-1}\mathsf{vec}\left(\hat{\Phi}^i\hat{\eta}_{t-i}\hat{\eta}_{t-i}'\hat{\Phi}^{i\prime}\kappa_{t-i}\right)\right)\right\| > \delta\right) &\leq \frac{1}{T^2\delta^2}\sum_{t=1}^{T}E^*\left\|\kappa_t\left(\sum_{i=0}^{T-t}\hat{\Phi}^i\otimes\hat{\Phi}^i\right)\mathsf{vec}\left(\hat{\eta}_t\hat{\eta}_t'\right)\right\|^2 \\
&\leq \frac{E\left(\kappa_t^2\right)}{T\delta^2}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\sum_{i=0}^{T-t}\hat{\Phi}^i\otimes\hat{\Phi}^i\right)\mathsf{vec}\left(\hat{\eta}_t\hat{\eta}_t'\right)\right\|^2\right).
\end{aligned}
$$

Thus, with $c_T = c + o_p(1)$,

$$
\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\sum_{i=0}^{T-t}\hat{\Phi}^i\otimes\hat{\Phi}^i\right)\mathsf{vec}\left(\hat{\eta}_t\hat{\eta}_t'\right)\right\|^2 \leq \frac{c_T}{T}\sum_{t=1}^{T}\left\|\mathsf{vec}\left(\hat{\eta}_t\hat{\eta}_t'\right)\right\|^2,
$$

which converges in probability as ε_t and, therefore, η_t have bounded fourth order moment. This establishes the result that $M_1^b - \bar{\Sigma}_{M^b} = o_p(1)$. It can similarly be shown that $M_2^b = o_p(1)$, which completes the proof. □

Proof of Theorem 3**:** The stated results can be shown to hold by following the proof of Theorem 1 and using Lemmas A.6 and A.7 above, see also Swensen (2006, proof of Proposition 1). □

References

Andrews, D.W.K., and M. Buchinsky (2000) A three-step method for choosing the number of bootstrap repetitions, *Econometrica* 68, 23–51.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991), Bootstrapping unstable first-order autoregressive processes, *Annals of Statistics* 19, 1098-1101.

Boswijk, H.P. and Y. Zu (2007), Testing for cointegration with nonstationary volatility, Working Paper, University of Amsterdam.

Burridge, P. and A.M.R. Taylor (2001), On regression-based tests for seasonal unit roots in the presence of periodic heteroscedasticity, *Journal of Econometrics* 104, 91–117.

Busetti, F. and A.M.R. Taylor (2003), Testing against stochastic trend in the presence of variance shifts, *Journal of Business and Economic Statistics* 21, 510–531.

Campbell, John Y and R.J.Shiller (1987), Cointegration and tests of present value models, *Journal of Political Economy* 95, 1062–1088.

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

Cavaliere G. and A.M.R. Taylor (2005), Stationarity tests under time-varying variances, *Econometric Theory* 21, 1112–1129.

Cavaliere G. and A.M.R. Taylor (2006), Testing the null of co-integration in the presence of variance breaks, *Journal of Time Series Analysis* 27, 619–636.

Cavaliere G. and A.M.R. Taylor (2007), Testing for unit roots in time series models with non-stationary volatility, *Journal of Econometrics*, forthcoming.

Cavaliere G. and A.M.R. Taylor (2008a), Bootstrap unit root tests for time series with non-stationary volatility, *Econometric Theory*, forthcoming.

Cavaliere G. and A.M.R. Taylor (2008b), Time-transformed unit root tests for models with non-stationary volatility, *Journal of Time Series Analysis*, forthcoming.

Davidson J. (1994), *Stochastic limit theory*, Oxford: Oxford University Press.

Davidson, R. and E. Flachaire (2001), The wild bootstrap, tamed at last, Working paper, IER#1000, Queens University.

Diebold, F.X., L. Ji and C. Li (2007), A three-factor yield curve model: non-affine structure, systematic risk sources, and generalized duration, forthcoming in L.R. Klein (ed.), *Macroeconomics, Finance and Econometrics: Essays in Memory of Albert Ando*, Cheltenham, U.K.: Edward Elgar.

Giese J. (2006), Level, slope, curvature: characterising the yield curve's derivatives in a cointegrated VAR model, Working paper, Nuffield College.

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

Gonçalves, S. and L. Kilian (2004), Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* 123, 89-120.

Hall, A.D., H. Anderson and C.W.J. Granger (1992), A cointegration analysis of treasury bill yields, *Review of Economics and Statistics* 74, 116–26.

Hamori, S. and A. Tokihisa (1997), Testing for a unit root in the presence of a variance shift, *Economics Letters* 57, 245–253.

Hansen, B.E. (1992a), Heteroskedastic cointegration, *Journal of Econometrics* 54, 139–158.

Hansen, B.E. (1992b), 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, B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 413–430.

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

Hansen, E., A. Rahbek and J. Dennis (2002), ARCH innovations and their impact on cointegration rank testing, Working paper no.22, Centre for Analytical Finance.

Hansen, P. R. (2003), Structural Changes in the Cointegrated Vector Autoregressive Model, *Journal of Econometrics* 114, 261–295.

Inoue, A. and L. Killian (2002), Bootstrapping autoregressive processes with possible unit roots, *Econometrica* 70, 377–391.

Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231–254.

Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551–1580.

Johansen, S. (1992), Determination of cointegration rank in the presence of a linear trend, *Oxford Bulletin of Economics and Statistics* 54, 383–397.

Johansen, S. (1996), *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford: Oxford University Press.

Johansen, S. (2002), A sample correction of the test for cointegrating rank in the vector autoregressive model, *Econometrica* 70, 1929–1961.

Kahn, J.A., M.M. McConnell and G. Perez-Quiroz (2002), On the causes of the increased stability of the U.S. economy, *Federal Reserve Bank of New York Economic Policy Review* 8, 183–202.

Kim, C.-J. and C.R. Nelson (1999), Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economics and Statistics* 81, 608–616.

Kim, T.H., S.J. Leybourne and P. Newbold (2002), Unit root tests with a break in innovation variance, *Journal of Econometrics* 109, 365–387.

Koop, G. and S.M. Potter (2000), Nonlinearity, structural breaks or outliers in economic time series?, in W.A. Barnett, D.F. Hendry, S. Hylleberg, T. Terasvirta, D. Tjostheim and A.H. Wurtz (eds.), *Nonlinear Econometric Modelling in Time Series Analysis*, Cambridge: Cambridge University Press, 61–78.

Lanne M. and H. Lütkepohl (2005), Structural vector autoregressions with nonnormal residuals, EUI Working Paper, ECO No. 2005/25.

Lanne M. and P. Saikkonen (2007), A multivariate generalized orthogonal factor GARCH model, *Journal of Business and Economic Statistics* 25, 61–75

Liu, R.Y. (1988), Bootstrap procedures under some non i.i.d. models, *Annals of Statistics* 16, 1696–1708.

Loretan M. and P.C.B. Phillips (1994), Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets, *Journal of Empirical Finance* 1, 211–248.

MacKinnon, J.G., A.A. Haug, and L. Michelis (1999), Numerical distribution functions of likelihood ratio tests for cointegration, *Journal of Applied Econometrics*14, 563–577.

Mammen, E. (1993), Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics* 21, 255–285.

McConnell, M.M. and G. Perez Quiros (2000), Output fluctuations in the United States: what has changed since the early 1980s?, *American Economic Review* 90, 1464–1476.

Osterwald-Lenum, M. (1992), A note with quantiles for the asymptotic distribution of the maximum likelihood cointegration rank test statistic, *Oxford Bulletin of Economics and Statistics* 53, 461–471.

Paparoditis, E.P. and D.N. Politis (2003), Residual-based block bootstrap for unit root testing, *Econometrica* 71, 813-855.

Park, J.Y. (2003), Bootstrap unit root tests, *Econometrica* 71, 1845–1895.

Perron, P. (1989), The great crash, the oil price shock, and the unit root hypothesis, *Econometrica* 57, 1361-1401.

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

Sensier, M. and D. van Dijk (2004), Testing for volatility changes in U.S. macroeconomic time series, *Review of Economics and Statistics* 86, 833-839.

Shephard, N. (2005), *Stochastic Volatility: Selected Readings*, Oxford University Press: Oxford.

Stock, J.H. and M.W. Watson (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in R.F. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger,* Oxford University Press: Oxford, 1–44.

Swensen, A.R. (2006), Bootstrap algorithms for testing and determining the cointegration rank in VAR models, *Econometrica* 74, 1699–1714.

Van der Weide, R. (2002), GO-GARCH: a multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* 17, 549–564.

van Dijk D., D.R. Osborn and M. Sensier (2002), Changes in variability of the business cycle in the G7 countries, Erasmus University Rotterdam, Econometric Institute Report EI 2002-28.

Watson M.W. (1999), Explaining the increased variability in long-term interest rates, *Federal Reserve Bank of Richmond Economic Quarterly* 85, 71–96.

Wu, C.F.J. (1986), Jackknife, bootstrap, and other resampling methods, *Annals of Statistics* 14, 1261–1295.

Xu, K.-L. and Phillips, P.C.B. (2007), Adaptive estimation of autoregressive models with time-varying variances, *Journal of Econometrics*, forthcoming.

Table 1: Size of Standard and Bootstrap PLR Tests for Rank = 0 Against Rank = 5. True Rank is 0.

T	Standard PLR test No volatility shifts	Bootstrap PLR test No volatility shifts
100	8.4	5.4
200	6.6	5.0
400	5.6	4.7

		Standard PLR test: Single volatility shift				Bootstrap PLR test: Single volatility shift			
shift in:	T	early negative	late negative	early positive	late positive	early negative	late negative	early positive	late positive
ε_{1t}	100	8.8	8.6	8.1	8.8	5.3	5.6	4.9	5.4
	200	7.1	6.7	7.4	7.8	4.8	4.7	5.2	5.6
	400	7.0	6.6	6.8	7.5	5.2	4.8	5.5	5.6
$\varepsilon_{1t}, \varepsilon_{2t}$	100	13.3	9.8	10.5	14.7	6.0	5.2	5.6	6.6
	200	12.6	9.4	9.3	13.5	5.5	5.7	5.4	5.8
	400	11.3	8.2	9.0	13.4	5.0	5.1	5.5	5.8
$\varepsilon_{1t}, .., \varepsilon_{3t}$	100	25.5	14.8	15.6	26.1	6.7	6.0	6.6	8.0
	200	24.5	13.6	13.1	24.6	7.2	5.6	5.7	6.4
	400	22.5	12.3	13.2	23.6	5.4	5.3	5.4	6.2
$\varepsilon_{1t}, .., \varepsilon_{4t}$	100	45.5	23.6	22.9	45.9	9.1	6.0	7.0	10.6
	200	42.2	20.4	21.1	43.1	7.2	5.6	5.8	7.5
	400	41.6	19.1	20.8	42.0	6.3	5.3	5.1	6.1
$\varepsilon_{1t}, .., \varepsilon_{5t}$	100	67.9	37.5	35.8	66.1	11.5	7.2	7.8	12.1
	200	65.2	32.3	34.4	63.5	7.8	5.8	6.5	8.7
	400	63.8	31.8	32.2	63.3	6.5	5.6	5.6	6.6

Table 2: Standard and Bootstrap Sequential Procedures for Selecting the Co-integration Rank. True Rank is 0.

	Standard PML test, no volatility shifts					Bootstrap PML test, no volatility shifts				
T	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$
100	91.6	7.4	0.9	0.2	0.0	94.6	4.6	0.7	0.1	0.0
200	93.4	5.8	0.7	0.0	0.0	95.0	4.5	0.4	0.0	0.0
400	94.4	5.1	0.4	0.0	0.0	95.3	4.4	0.2	0.0	0.0

Standard PML test, single volatility shift

		early negative shift					late negative shift				
shift in:	T	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$
	100	91.2	7.8	0.9	0.1	0.0	91.4	7.8	0.7	0.1	0.0
ε_{1t}	200	92.9	6.4	0.6	0.1	0.0	93.3	6.0	0.5	0.1	0.0
	400	93.0	6.4	0.5	0.0	0.0	93.4	5.8	0.6	0.1	0.0
	100	86.7	11.7	1.4	0.2	0.1	90.2	8.9	0.8	0.1	0.0
$\varepsilon_{1t}, \varepsilon_t$	200	87.4	11.4	1.2	0.0	0.0	90.6	8.5	0.9	0.1	0.0
	400	88.7	10.3	0.9	0.1	0.0	91.8	7.4	0.8	0.0	0.0
	100	74.5	21.9	3.2	0.2	0.1	85.2	12.9	1.7	0.2	0.0
$\varepsilon_{1t},..,\varepsilon_{3t}$	200	75.5	21.0	3.1	0.3	0.1	86.4	12.1	1.3	0.1	0.0
	400	77.5	19.2	3.1	0.2	0.0	87.7	10.7	1.4	0.2	0.0
	100	54.5	35.0	9.2	1.1	0.2	76.4	19.9	3.1	0.5	0.1
$\varepsilon_{1t},..,\varepsilon_{4t}$	200	57.8	32.5	8.4	1.1	0.1	79.6	17.7	2.3	0.4	0.0
	400	58.4	32.6	7.8	1.0	0.2	80.9	16.6	2.3	0.3	0.0
	100	32.1	42.0	19.9	5.5	0.5	62.5	28.9	7.1	1.2	0.2
$\varepsilon_{1t},..,\varepsilon_{5t}$	200	34.8	41.4	18.5	4.7	0.5	67.7	25.3	6.2	0.6	0.2
	400	36.2	41.4	17.6	4.3	0.6	68.2	25.7	5.3	0.7	0.1

Bootstrap PML test, single volatility shift

		early negative shift					late negative shift				
shift in:	T	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$
	100	94.7	4.5	0.6	0.1	0.0	94.4	5.1	0.4	0.1	0.0
ε_{1t}	200	95.2	4.4	0.4	0.1	0.0	95.3	4.3	0.3	0.1	0.0
	400	94.8	4.9	0.3	0.0	0.0	95.2	4.2	0.5	0.1	0.0
	100	94.0	5.3	0.5	0.1	0.0	94.8	4.7	0.5	0.0	0.0
$\varepsilon_{1t}, \varepsilon_t$	200	94.5	5.0	0.5	0.0	0.0	94.3	5.3	0.4	0.0	0.0
	400	95.0	4.5	0.5	0.0	0.0	94.9	4.7	0.4	0.0	0.0
	100	93.3	5.8	0.8	0.1	0.0	94.0	5.4	0.6	0.0	0.0
$\varepsilon_{1t},..,\varepsilon_{3t}$	200	92.8	6.3	0.7	0.0	0.1	94.4	5.2	0.4	0.0	0.0
	400	94.6	5.0	0.4	0.0	0.0	94.7	4.7	0.6	0.0	0.0
	100	90.9	8.0	1.0	0.0	0.1	94.0	5.3	0.7	0.0	0.0
$\varepsilon_{1t},..,\varepsilon_{4t}$	200	92.8	6.3	0.8	0.1	0.0	94.4	5.0	0.5	0.0	0.0
	400	93.7	5.5	0.7	0.1	0.0	94.7	4.7	0.5	0.0	0.0
	100	88.5	9.9	1.3	0.2	0.0	92.8	6.3	0.8	0.1	0.1
$\varepsilon_{1t},..,\varepsilon_{5t}$	200	92.2	6.6	1.1	0.1	0.0	94.2	5.0	0.7	0.1	0.1
	400	93.5	5.8	0.6	0.2	0.0	94.4	5.0	0.6	0.0	0.0

Table 3: Size of Standard and Bootstrap PLR Tests for Rank = 1 Against Rank = 5. True Rank is 1 ($a_1 = -0.4$; $-0.4a_2 = 0.0$).

	Standard PLR test, No volatility shifts				Bootstrap PLR test, No volatility shifts			
T	$a_2 = 0.0$		$a_2 = -0.4$		$a_2 = 0.0$		$a_2 = -0.4$	
100	4.9	0.0	7.4	0.0	3.8	0.0	5.4	0.0
200	6.0	0.0	6.7	0.0	5.3	0.0	5.2	0.0
400	5.8	0.0	5.7	0.0	4.9	0.0	5.0	0.0

		Standard PLR test, Single volatility shift				Bootstrap PLR test, Single volatility shift			
		$a_2 = 0.0$		$a_2 = -0.4$		$a_2 = 0.0$		$a_2 = -0.4$	
shift in:	T	early negative	late negative	early negative	late negative	early negative	late negative	early negative	late negative
	100	4.7	5.0	8.2	9.8	3.5	3.9	4.9	5.6
ε_{1t}	200	5.6	5.8	7.6	7.6	4.8	4.8	5.5	5.3
	400	5.9	5.6	6.2	6.3	4.9	4.7	4.8	4.7
	100	6.1	4.8	8.0	7.4	3.8	3.3	4.7	4.8
$\varepsilon_{1t}, \varepsilon_{2t}$	200	6.4	6.7	7.2	6.8	4.9	5.2	5.3	4.9
	400	6.3	6.1	6.2	6.3	4.8	5.0	4.5	4.9
	100	10.8	7.5	14.0	10.4	4.1	3.7	5.9	5.5
$\varepsilon_{1t}, .., \varepsilon_{3t}$	200	12.2	8.9	13.4	9.1	5.2	5.3	5.2	5.0
	400	12.4	8.3	11.7	8.1	5.1	5.0	5.0	5.0
	100	25.2	11.1	28.1	15.6	5.1	3.9	7.4	5.5
$\varepsilon_{1t}, .., \varepsilon_{4t}$	200	26.4	13.5	24.9	14.2	6.4	5.2	6.2	5.2
	400	26.0	12.7	25.9	12.5	5.6	5.0	5.4	5.0
	100	44.5	21.0	50.6	26.4	5.9	3.9	8.3	6.0
$\varepsilon_{1t}, .., \varepsilon_{5t}$	200	48.6	22.5	49.2	23.3	6.3	5.2	6.8	5.3
	400	47.7	22.0	47.0	21.5	5.9	5.5	5.0	4.8

Table 4: Standard and Bootstrap Sequential Procedures for Selecting the Co-integration Rank. True Rank is 1 ($a_1 = -0.4; a_2 = 0.0$).

	Standard PML test, no volatility shifts					Bootstrap PML test, no volatility shifts				
T	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$
100	43.4	51.7	4.4	0.4	0.1	53.2	42.9	3.4	0.3	0.1
200	0.7	93.4	5.4	0.5	0.1	1.2	93.6	4.8	0.4	0.1
400	0.0	94.2	5.2	0.5	0.0	0.0	95.1	4.5	0.4	0.0

		Standard PML test, single volatility shift: early negative shift					Standard PML test, single volatility shift: late negative shift					Bootstrap PML test, single volatility shift: early negative shift					Bootstrap PML test, single volatility shift: late negative shift				
shift in	T	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$	$r=0$	$r=1$	$r=2$	$r=3$	$r=4,5$
	100	41.5	53.9	4.3	0.2	0.1	41.8	53.2	4.5	0.4	0.1	51.8	44.7	3.0	0.3	0.1	53.5	42.7	3.4	0.4	0.1
ε_{1t}	200	2.1	92.3	5.2	0.4	0.0	1.0	93.3	5.2	0.5	0.1	3.3	91.9	4.4	0.4	0.0	2.0	93.2	4.3	0.4	0.1
	400	0.0	94.1	5.4	0.4	0.0	0.0	94.4	5.1	0.4	0.1	0.0	95.1	4.6	0.3	0.0	0.0	95.3	4.3	0.4	0.0
	100	34.0	59.9	5.5	0.5	0.0	40.2	55.0	4.4	0.4	0.0	55.4	40.9	3.4	0.3	0.1	56.6	40.1	3.0	0.3	0.0
$\varepsilon_{1t}, \varepsilon_{2t}$	200	1.4	92.2	5.8	0.6	0.1	1.0	92.3	6.1	0.6	0.0	5.1	90.0	4.3	0.5	0.0	2.3	92.5	4.7	0.4	0.0
	400	0.0	93.7	5.6	0.5	0.1	0.0	93.9	5.7	0.4	0.0	0.0	95.2	4.2	0.4	0.1	0.0	95.0	4.6	0.4	0.0
	100	25.0	64.3	9.8	1.0	0.1	32.7	59.7	6.7	0.8	0.1	58.8	37.1	3.7	0.3	0.1	58.0	38.3	3.2	0.4	0.1
$\varepsilon_{1t}, .., \varepsilon_{3t}$	200	0.8	87.0	11.0	1.1	0.1	0.7	90.4	8.1	0.7	0.1	8.1	86.6	4.6	0.6	0.1	3.1	91.6	4.9	0.4	0.0
	400	0.0	87.6	11.3	1.0	0.1	0.0	91.7	7.6	0.6	0.0	0.0	94.9	4.6	0.5	0.0	0.0	95.0	4.7	0.3	0.0
	100	12.7	62.1	21.5	3.5	0.2	25.6	63.3	9.7	1.2	0.2	61.7	33.3	4.3	0.6	0.1	60.9	35.2	3.4	0.3	0.1
$\varepsilon_{1t}, .., \varepsilon_{4t}$	200	0.3	73.2	23.0	3.0	0.4	0.5	86.0	12.2	1.3	0.1	12.7	81.0	5.7	0.6	0.1	5.1	89.7	4.6	0.5	0.1
	400	0.0	74.0	22.6	3.1	0.2	0.0	87.3	11.5	1.2	0.1	0.0	94.4	5.0	0.5	0.1	0.0	95.0	4.7	0.3	0.0
	100	5.6	49.9	32.9	10.4	1.2	15.6	63.4	17.7	2.9	0.4	62.9	31.2	5.1	0.6	0.2	64.0	32.1	3.3	0.4	0.1
$\varepsilon_{1t}, .., \varepsilon_{5t}$	200	0.1	51.3	36.7	10.5	1.4	0.2	77.4	19.8	2.4	0.3	19.3	74.4	5.6	0.6	0.1	7.1	87.7	4.5	0.6	0.1
	400	0.0	52.3	36.6	9.6	1.6	0.0	78.0	19.1	2.5	0.4	0.0	94.0	5.2	0.6	0.1	0.0	94.5	5.1	0.3	0.1

Table 5: Standard and Bootstrap Sequential Procedures for Selecting the Co-integration Rank. True Rank is 1 ($a_1 = -0.4; a_2 = -0.4$).

	Standard PML test no volatility shifts					Bootstrap PML test no volatility shifts				
T	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4, 5$	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4, 5$
100	3.0	89.6	6.7	0.6	0.1	5.2	89.4	4.8	0.5	0.1
200	0.0	93.3	6.3	0.3	0.1	0.0	94.8	4.9	0.3	0.1
400	0.0	94.3	5.3	0.3	0.1	0.0	95.0	4.7	0.3	0.0

		Standard PML test: single volatility shift										Bootstrap PML test: single volatility shift									
		early negative shift					late negative shift					early negative shift					late negative shift				
shift in:	T	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4, 5$	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4, 5$	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4, 5$	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4, 5$
	100	0.0	91.8	7.5	0.6	0.1	0.0	90.2	8.8	0.9	0.1	0.0	95.1	4.5	0.3	0.1	0.0	94.4	5.0	0.6	0.1
ε_{1t}	200	0.0	92.4	7.0	0.5	0.1	0.0	92.4	7.0	0.5	0.1	0.0	94.5	4.9	0.5	0.1	0.0	94.7	4.9	0.3	0.0
	400	0.0	93.8	5.8	0.4	0.0	0.0	93.7	5.9	0.4	0.1	0.0	95.2	4.5	0.3	0.0	0.0	95.3	4.3	0.3	0.0
	100	4.3	87.7	7.4	0.5	0.1	3.4	89.2	6.8	0.5	0.1	11.5	83.9	4.2	0.5	0.0	8.3	86.9	4.4	0.3	0.1
$\varepsilon_{1t}, \varepsilon_{2t}$	200	0.0	92.8	6.6	0.5	0.1	0.0	93.2	6.2	0.5	0.1	0.0	94.7	4.9	0.3	0.1	0.0	95.1	4.5	0.4	0.1
	400	0.0	93.8	5.7	0.5	0.0	0.0	93.7	5.8	0.5	0.1	0.0	95.5	4.1	0.4	0.0	0.0	95.1	4.4	0.4	0.1
	100	2.8	83.2	12.4	1.5	0.1	2.8	86.8	9.3	1.0	0.1	15.1	79.1	5.2	0.6	0.0	9.4	85.1	4.9	0.5	0.1
$\varepsilon_{1t}, .., \varepsilon_{3t}$	200	0.0	86.6	12.5	0.9	0.0	0.0	90.9	8.4	0.6	0.0	0.1	94.7	4.8	0.3	0.0	0.0	95.0	4.5	0.5	0.0
	400	0.0	88.3	10.7	1.0	0.1	0.0	91.9	7.4	0.6	0.1	0.0	95.0	4.4	0.5	0.1	0.0	95.0	4.5	0.4	0.0
	100	1.2	70.7	23.8	3.8	0.4	1.8	82.6	13.6	1.9	0.2	19.6	73.1	6.6	0.6	0.2	11.7	82.7	4.8	0.7	0.1
$\varepsilon_{1t}, .., \varepsilon_{4t}$	200	0.0	75.1	21.3	3.3	0.3	0.0	85.8	12.9	1.2	0.1	0.1	93.7	5.5	0.6	0.1	0.0	94.8	4.7	0.4	0.1
	400	0.0	74.1	22.7	2.9	0.2	0.0	87.5	11.1	1.3	0.1	0.0	94.6	4.8	0.5	0.1	0.0	95.0	4.5	0.5	0.1
	100	0.4	49.0	38.3	10.5	1.8	0.7	72.9	22.2	3.6	0.5	24.9	66.8	7.1	0.9	0.3	15.3	78.7	5.2	0.6	0.2
$\varepsilon_{1t}, .., \varepsilon_{5t}$	200	0.0	50.8	37.0	10.9	1.3	0.0	76.7	20.0	3.1	0.3	0.4	92.8	6.1	0.6	0.1	0.0	94.7	4.9	0.4	0.0
	400	0.0	53.0	36.2	9.6	1.2	0.0	78.5	18.5	2.6	0.4	0.0	95.0	4.4	0.5	0.1	0.0	95.2	4.2	0.5	0.1

Table 6: Standard and Bootstrap Co-integration Tests. Monthly U.S. Interest Rate Data, 1970:1-2000:12.

r	Eigenvalue	Q_r	Asymptotic 5% Critical Value	p-value	Wild Bootstrap p-value
0	0.202	193.66	75.74	0.000	0.000
1	0.152	110.42	53.42	0.000	0.000
2	0.074	49.66	34.80	0.008	0.030
3	0.048	21.24	19.99	0.037	0.095
4	0.009	3.25	9.13	0.544	0.644

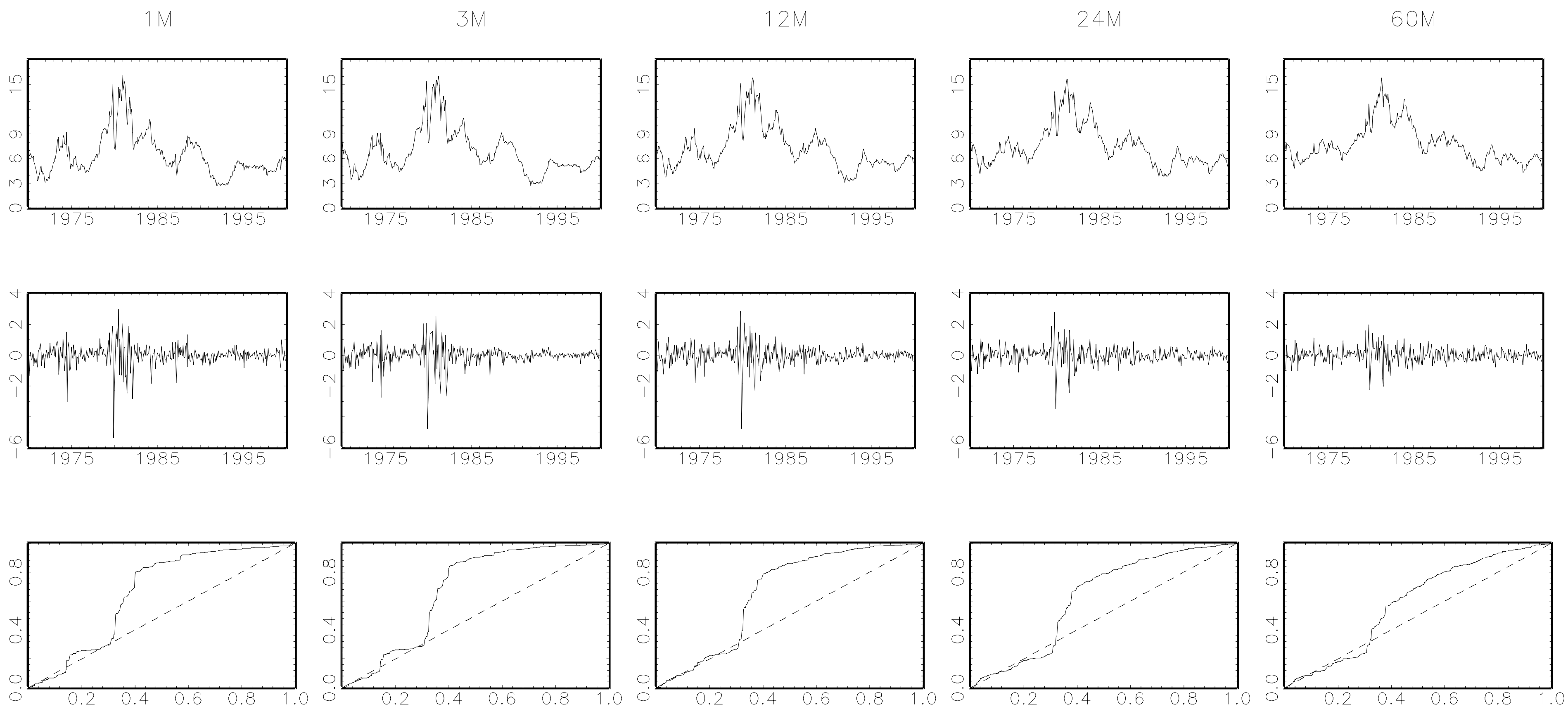

Figure 1: Levels, First Differences and Estimated Variance Profiles for Monthly data of U.S. Treasury Zero-Coupon Yields with One Month, Three Months, One Year, Two Years and Five Years Maturity, 1970–2000.