Inference on Co-integration Parameters in Heteroskedastic Vector Autoregressions*

H. Peter Boswijk[a], Giuseppe Cavaliere[b], Anders Rahbek[c] and A.M. Robert Taylor[d]

[a] Amsterdam School of Economics and Tinbergen Institute, University of Amsterdam

[b] Department of Statistical Sciences, University of Bologna

[c] Department of Economics, University of Copenhagen and CREATES

[d] Essex Business School, University of Essex

January 20, 2015

Abstract

We consider estimation and hypothesis testing on the coefficients of the co-integrating relations and the adjustment coefficients in vector autoregressions driven by shocks which display both conditional and unconditional heteroskedasticity of a quite general and unknown form. We show that the conventional results in Johansen (1996) for the maximum likelihood estimators and associated likelihood ratio tests derived under homoskedasticity do not in general hold under heteroskedasticity. As a result, standard confidence intervals and hypothesis tests on these coefficients are potentially unreliable. Solutions based on Wald tests (using a "sandwich" estimator of the variance matrix) and on the use of the wild bootstrap are discussed. These do not require the practitioner to specify a parametric model for volatility. We establish the conditions under which these methods are asymptotically valid. A Monte Carlo simulation study demonstrates that significant improvements in finite sample size can be obtained by the bootstrap over the corresponding asymptotic tests in both heteroskedastic and homoskedastic environments. An application to the term structure of interest rates in the US illustrates the difference between standard and bootstrap inferences regarding hypotheses on the co-integrating vectors and adjustment coefficients.

Keywords: Co-integration; adjustment coefficients; (un)conditional heteroskedasticity; heteroskedasticity-robust inference; wild bootstrap.

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

*We are grateful to the Editor, Peter Robinson, and Associate Editor and four referees for their helpful and constructive comments on an earlier version of this paper. Cavaliere, Rahbek and Taylor thank the Danish Council for Independent Research, Sapere Aude | DFF Advanced Grant (Grant nr: 12-124980) for financial support. Cavaliere also thanks the Italian Ministry of Education, University and Research (PRIN project "Multivariate statistical models for risk assessment") for financial support. Correspondence to: Giuseppe Cavaliere, Department of Statistical Sciences, Via Belle Arti 41, I-40126 Bologna, Italy. Tel: +39 051 2098230; fax: +39 051 232153. *E-mail addresses*: h.p.boswijk@uva.nl (H.P. Boswijk), giuseppe.cavaliere@unibo.it (G. Cavaliere), anders.rahbek@econ.ku.dk (A. Rahbek), rtaylor@essex.ac.uk (A.M.R. Taylor)

1 Introduction

In this paper we focus on the problem of conducting inference (estimation and hypothesis testing) on the coefficients of the co-integrating relations and associated adjustment parameters, based around the likelihood-based methods of Johansen (1996), in vector autoregressive time series which display time-varying behaviour in the variance of the driving shocks. We allow for both unconditional heteroskedasticity (often referred to as non-stationary volatility in the literature) and conditional heteroskedasticity in our analysis. It is well known that the assumption of conditional homoskedasticity appears inconsistent with financial and macroeconomic data; see, for example, Gonçalves and Kilian (2004). A large body of recent applied work has grown suggesting that the assumption of constant *unconditional* volatility is also at odds with what is observed in the data, with a general decline in the unconditional volatility of the shocks driving macroeconomic series in the twenty years or so leading up to the recent financial crisis, the so-called "Great Moderation", commonly observed; see, for example, *inter alia*, Kim and Nelson (1999) and McConnell and Perez Quiros (2000) and the references therein.

These empirical findings have helped stimulate research into the impact of time-varying conditional and unconditional volatility on standard time series methods. Of most relevance to this paper, Cavaliere, Rahbek and Taylor (2010b) analyse the impact this has on the conventional co-integration rank pseudo likelihood ratio (PLR) tests of Johansen (1996). They demonstrate that the asymptotic null distributions of the PLR statistics, which are constructed under the assumption that the innovations are i.i.d. and Gaussian, are non-pivotal in the presence of unconditional heteroskedasticity. Cavaliere, Rahbek and Taylor (2014) [CRT] show that wild bootstrap implementations of the PLR tests are, however, asymptotically valid.[1] Cavaliere, Rahbek and Taylor (2010a) provide a separate treatment for the case where the shocks are conditionally heteroskedastic but unconditionally homoskedastic. They show that the standard PLR tests (based on asymptotic critical values) are asymptotically valid, but that the corresponding wild bootstrap tests can deliver considerable finite sample improvements.

In this paper we make two distinct contributions to the literature. Utilising a very general set-up which combines the assumptions of Cavaliere, Rahbek and Taylor (2010a) and Cavaliere, Rahbek and Taylor (2010b) into a unified framework, our first contribution is to examine the impact of time-varying volatility on the large sample properties of the standard likelihood-based methods of estimation and hypothesis testing on the coefficients of the long run relations and the associated adjustment coefficients (β and α, respectively, in standard notation) detailed in Johansen (1996). In particular, we analyse the pseudo maximum likelihood (PML) estimates

[1] The algorithm proposed in CRT generates bootstrap samples using estimates all of which are obtained under the rank restriction imposed by the null, as is also done in Cavaliere, Rahbek and Taylor (2012), who use an i.i.d., rather than wild, re-sampling scheme. Cavaliere, Rahbek and Taylor (2010a,b) also propose an alternative algorithm, along the lines of that considered in Swensen (2006) using restricted estimates only for the long run parameters of the model. Cavaliere, Rahbek and Taylor (2012) and CRT demonstrate that the algorithms they propose are preferable to those proposed in Cavaliere, Rahbek and Taylor (2010a,b).

of these parameters and the associated PLR test for linear restrictions on these parameters, both derived under the assumption of an i.i.d. Gaussian pseudo-likelihood. We also analyse the corresponding Wald statistic, based around a PML ("sandwich") variance matrix estimator. We demonstrate that although the PML estimates are consistent, standard confidence intervals and PLR test statistics based on the PML estimates of α and β will not be reliable in general, their form depending on nuisance parameters arising from any heteroskedasticity in the shocks. Where the shocks are unconditionally homoskedastic, however, inference on β alone is shown to be asymptotically pivotal. For this to hold for the PLR tests involving α, conditional heteroskedasticity must also be absent from the shocks. We show that asymptotically robust inference can be achieved on α, regardless of any heteroskedasticity present, by using the Wald statistic. This also holds when using the Wald statistic to test hypotheses involving β, provided the shocks are unconditionally homoskedastic, but in general is not true when non-stationary volatility is present. These results complement those given in Hansen (1992a) for the case of a single equation error-correction model (as in Engle and Granger, 1987), driven by an error term whose volatility follows a first-order integrated ($I(1)$) process.

Our second contribution is to develop wild bootstrap implementations of the standard PLR and Wald tests. Extant bootstrap methods for testing hypotheses on the co-integration parameters deal with tests on β only and are at most devised for the case of independent, identically distributed shocks; see Omtzigt and Fachin (2006), Cavaliere, Nielsen and Rahbek (2015) and the references therein. In contrast, we derive the conditions under which wild bootstrap implementations of the PLR and Wald tests of hypotheses on both α and β can replicate the first order limiting null distributions of the corresponding standard test statistics. In such cases asymptotically valid bootstrap inference can be performed in the presence of time-varying volatility using the wild bootstrap versions of these tests. For the bootstrap PLR tests involving α this requires the assumption of a further moment condition and the assumption of the absence of asymmetric volatility clustering, as formally defined below after Assumption 2. For the PLR tests involving only β neither of these additional assumptions is required, while for the Wald tests, the additional assumption on the form of the volatility clustering is also not required. When testing joint hypotheses on α and β, statistical leverage effects (defined after Assumption 2) need to be ruled out for bootstrap inference based on PLR and Wald tests.

The remainder of the paper is organised as follows. Section 2 defines the heteroskedastic model, discussing in detail the type of time-varying volatility that we consider. We then characterise the asymptotic behaviour of the common trends in the process. Next, we introduce a class of hypotheses on the co-integrating vectors and error correction coefficients. Section 3 derives the asymptotic null distributions of the PLR and Wald test statistics for the class of hypotheses we consider. The wild bootstrap approach, based on a sieve-type procedure using the PML coefficient matrix estimates from the co-integrated VAR model, is outlined in Section 4. Here the conditions under which the wild bootstrap tests deliver asymptotically valid inference are also detailed. In Section 5 we use Monte Carlo simulation evidence to compare the small

sample size properties of the standard (asymptotic) tests and their bootstrap analogues for a variety of heteroskedastic co-integrated VAR models. An empirical application of the proposed methods to the term structure of interest rates in the US is presented in Section 6. Section 7 concludes. All proofs are contained in the Appendix.

In the following '$\xrightarrow{w}$' denotes weak convergence and '$\xrightarrow{p}$' convergence in probability, in each case as the sample size, T, diverges; $\mathbb{I}(\cdot)$ 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 $[0,1]$, equipped with the Skorohod metric. The space spanned by the columns of any $m \times n$ matrix A is denoted as $\mathrm{col}(A)$; 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 $\|A\|^2 := \mathrm{tr}\{A'A\}$, 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, $\|x\| := (x'x)^{1/2}$. Finally, $\otimes$ denotes the Kronecker product.

2 The Heteroskedastic VAR Model and Hypotheses

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

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \alpha\rho_1' D_{1t} + \mu_2 D_{2t} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1}$$

where X_t is a p-variate vector process, with initial values $(X_{1-k}, \ldots, X_0)$, which are known and taken to be fixed in the statistical analysis, and D_{1t} and D_{2t} are vectors of deterministic terms, such as a constant or linear trend, of dimensions d_1 and d_2, respectively. The disturbance ε_t is assumed to be a p-variate vector martingale difference sequence relative to some filtration $\mathcal{F}_t$, with finite and positive definite conditional variance matrix. Further conditions on ε_t are discussed below. The parameter matrices α and β, which are our key focus in this paper, are of dimension $p \times r$, with $0 < r < p$, and $\{\Gamma_j\}_{j=1}^{k-1}$ are $p \times p$ lag coefficient matrices. The co-integration rank, r, is assumed to be known in what follows; in practice this would first be determined using the wild bootstrap co-integration rank tests of CRT. The parameter matrices ρ_1 and μ_2 are of dimension $d_1 \times r$ and $p \times d_2$, respectively; note that D_{1t} enters the model through the error correction term $\alpha(\beta' X_{t-1} + \rho_1' D_{1t})$ only, whereas D_{2t} appears unrestrictedly. The usual cases of interest which we consider in this paper are: (i) $D_{1t} = 1$, $D_{2t} = 0$ (restricted constant), and (ii) $D_{1t} = t$, $D_{2t} = 1$ (restricted linear trend); see Johansen (1992).

We assume that the process in (1) satisfies the following condition (referred to as the '$I(1,r)$ condition' hereafter):

Assumption 1. (a) *the characteristic equation associated with (1), i.e.* $|A(z)| = 0$ *with* $A(z) := (1-z) I_p - \alpha\beta' z - \sum_{j=1}^{k-1} \Gamma_j z (1-z)$, *has* $p - r$ *roots equal to 1 and all other roots*

outside the unit circle, and (b) α *and* β *have full column rank* r*.*

An implication of Assumption 1 is that ΔX_t and $\beta' X_t$ may be written as linear processes in terms of ε_t, with exponentially decaying coefficient matrices. That is, these are "stable" processes in the sense of Cavaliere, Rahbek and Taylor (2010b), which would reduce to stationary [or $I(0)$] processes if the unconditional variance of ε_t were constant. Because we allow for time-varying behaviour in the variance matrices (both conditional and unconditional), the definitions of integrated and co-integrated processes do not formally apply in the present case, although it will be convenient still to refer to the elements of β as co-integration parameters (as in the title of this paper). The assumption on the number of unit roots excludes integration of a higher order, and is equivalent to the assumption that $|\alpha_\perp' \Gamma \beta_\perp| \neq 0$, with $\Gamma := I_p - \sum_{j=1}^{k-1} \Gamma_j$; see Johansen (1996).

As regards the sequence ε_t, we assume the following:

Assumption 2. *The process* ε_t *can be written as* $\varepsilon_t = \sigma_t z_t$*, where:*

(a) $\sigma_t = \sigma(t/T)$*, where* $\sigma(\cdot)$ *is a non-stochastic element of* $\mathcal{D}_{\mathbb{R}^{p\times p}}[0,1]$ *such that* $\Sigma(u) := \sigma(u)\sigma(u)' > 0$ *for all* $u \in [0,1]$*;*

(b) z_t *is a* p*-vector martingale difference sequence relative to a filtration* $\mathcal{F}_t$*, with conditional variance matrix* $h_t := E(z_t z_t' | \mathcal{F}_{t-1})$*, satisfying*

i. $T^{-1}\sum_{t=1}^{T} h_t \xrightarrow{p} E(z_t z_t') = I_p$,

ii. $T^{-1}\sum_{t=1}^{T}(h_t \otimes z_{t-i}) \xrightarrow{p} E(z_t z_t' \otimes z_{t-i}) = \varrho_i$ *for* $i \geq 1$*, with* $\sup_{i\geq 1} \|\varrho_i\| =: \varrho < \infty$,

iii. $T^{-1}\sum_{t=1}^{T}(h_t \otimes z_{t-i} z_{t-j}') \xrightarrow{p} E(z_t z_t' \otimes z_{t-i} z_{t-j}') = \tau_{ij}$ *for* $i, j \geq 1$*, with* $\sup_{i,j\geq 1} \|\tau_{ij}\| < \infty$,

iv. $\sup_t E\|z_t\|^{4r} < \infty$ *for some* $r > 1$*.*

Assumption 2 implies that ε_t is a vector martingale difference sequence relative to $\mathcal{F}_t$, with conditional variance matrix $\Sigma_{t|t-1} := E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = \sigma_t h_t \sigma_t'$, and time-varying unconditional variance matrix $\Sigma_t := E(\varepsilon_t \varepsilon_t') = \sigma_t \sigma_t' > 0$.[2] As such, it combines the assumptions of Cavaliere, Rahbek and Taylor (2010a) and Cavaliere, Rahbek and Taylor (2010b), who consider VAR models with stationary conditional heteroskedasticity or non-stationary unconditional volatility, respectively. These are obtained as special cases with $\sigma(\cdot) = \sigma$ (constant unconditional variance, hence only conditional heteroskedasticity) and $h_t = I_p$ (so $\Sigma_{t|t-1} = \Sigma_t = \Sigma(t/T)$, only unconditional non-stationary volatility). As discussed in Cavaliere, Rahbek and Taylor (2010b),

[2]Notice that, since σ_t is unrestricted, the assumption that $E(z_t z_t') = I_p$ made in part (b)i is without loss of generality and is made only to simplify notation. In particular, any $\varepsilon_t = \sigma_t z_t$ satisfying Assumption 2 with $E(z_t z_t') = \Omega$ can also be expressed as $\varepsilon_t = \tilde{\sigma}_t \tilde{z}_t$ with $E(\tilde{z}_t \tilde{z}_t') = I_p$ and $\tilde{\sigma}_t := \sigma_t \Omega^{-1/2}$, where both $\tilde{z}_t$ and $\tilde{\sigma}_t$ satisfy Assumption 2.

Assumption 2 (a) implies that the elements of Σ_t are only required to be bounded and to display a countable number of jumps, therefore allowing for an extremely wide class of potential models for the behaviour of the variance matrix of ε_t, including single or multiple variance or covariance shifts, variances which follow a broken trend, and smooth transition variance shifts. Assumption 2 (b), which is closely related to Assumption A of Gonçalves and Kilian (2004), allows for (possibly asymmetric) volatility clustering and statistical leverage effects. Volatility clustering, such as generalised autoregressive-conditional heteroskedasticity (GARCH), is allowed for by the assumption that z_t is a martingale difference sequence where the quantity $\tau_{ii} := E(z_t z_t' \otimes z_{t-i} z_{t-i}')$ is not necessarily equal to $E(z_t z_t') \otimes E(z_{t-i} z_{t-i}') = I_p$. In particular, Deo (2000) provides examples of stochastic volatility and GARCH processes that satisfy Assumption 2 (b), including Gaussian GARCH(1,1) processes with a finite unconditional 8th moment. Statistical leverage occurs where the quantity $\varrho_i := E(h_t \otimes z_{t-i})$ is non-zero for some $i = 1, 2, ...$ Likewise asymmetric volatility clustering, that is correlation between the conditional variance h_t and cross products of past shocks $z_{t-i} z_{t-j}'$, is allowed for by non-zero τ_{ij} for $i \neq j$. For the benefit of the reader when excluding leverage and asymmetric volatility clustering respectively, we summarise this as $\varrho = 0$ and $\tau = 0$, where $\varrho := \sup \|\varrho_i\|$ and $\tau := \sup_{i,j \geq 1, i \neq j} \|\tau_{ij}\|$; cf. Assumption 2.

It should be stressed that Assumption 2 (b) rules out non-stationary conditional heteroskedasticity such as is generated by, for example, near-integrated GARCH models and the non-stationary autoregressive volatility processes of Hansen (1995); see Cavaliere and Taylor (2009, p. 1235) for further discussion of models allowing for non-stationary conditional heteroskedasticity. The non-stationary non-linear heteroskedastic (NNH) framework of Park (2002) and Chung and Park (2007) is also not permitted under our assumptions. In our set-up the volatility function $\sigma_t = \sigma(t/T)$ is defined through a triangular array, thereby allowing for breaks in correlation and variance which are not permitted in the NNH setup where $\sigma_t = f(v_t)$ with v_t a random walk and $f(\cdot)$ a time-invariant function.

2.1 Representation

Before we introduce the class of hypotheses we will consider in this paper, we provide some preliminary results on the stable and unstable linear combinations of X_t. Let $\mu D_t = \mu_1 D_{1t} + \mu_2 D_{2t}$, where $\mu_1 = \alpha \rho_1'$, and let $\mathbb{X}_t := (X_t', \ldots, X_{t-k+1}')'$, and $\mathbb{X}_{\beta t} := \left(X_t'\beta, \Delta X_t', \ldots, \Delta X_{t-k+1}'\right)' = \mathbb{B}'\mathbb{X}_t$, where $\mathbb{B}$ is defined implicitly; if $k = 1$, then $\mathbb{B} = \beta$ and $\mathbb{X}_{\beta t} = \beta' X_t$. Lemma 1 of Cavaliere, Rahbek and Taylor (2010b) applies directly to our model under Assumptions 1 and 2, and states that

$$\mathbb{X}_{\beta t} = \Phi \mathbb{X}_{\beta, t-1} + F(\mu D_t + \varepsilon_t), \tag{2}$$

where the autoregressive matrix is defined by $\Phi := I_{r+p(k-1)} + \mathbb{B}'\mathbb{A}$, with

$$\mathbb{A} := \begin{pmatrix} \alpha & \Psi \\ 0 & I_{p(k-1)} \end{pmatrix}, \qquad \Psi := [\Gamma_1, \ldots, \Gamma_{p-1}],$$

and where $F := (\beta, I_p, 0, \ldots, 0)'$. Because $\rho(\Phi) < 1$, $\mathbb{X}_{\beta t}$ is a stable process. For X_t, Lemma 1 of Cavaliere, Rahbek and Taylor (2010b) yields the result that

$$X_t = C \sum_{i=1}^{t} (\mu D_i + \varepsilon_i) + S_t + C_0, \tag{3}$$

where $C := \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$, $S_t := (I_r, 0, \ldots, 0)\mathbb{A}(\mathbb{B}'\mathbb{A})^{-1}\mathbb{X}_{\beta t}$, and where C_0 is a constant, depending on the initial values, defined by $C_0 := C\,(I_p, -\Psi)\,\mathbb{X}_0$. These results are purely algebraic and, hence, do not depend on specific assumptions on $\varepsilon_t = \sigma_t z_t$. The result in (3) implies that the stochastic part of $\alpha'_\perp \Gamma X_t$ is given by the sum of the heteroskedastic random walk $\sum_{i=1}^{t} \alpha'_\perp \varepsilon_i$, and the stable process $\alpha'_\perp \Gamma S_t$.

The following result gives the limiting behaviour of the random walk component, $\sum_{i=1}^{t} \varepsilon_i$, and of a particular sample moment matrix. It is a direct extension of Lemma 2 of Cavaliere, Rahbek and Taylor (2010b), although in the present context we also allow for stationary conditional heteroskedasticity in z_t of the form specified in Assumption 2 (b):

Lemma 1. *Let $\varepsilon_t = \sigma_t z_t$ satisfy Assumption 2, and let $W(\cdot)$ denote a p-variate standard Brownian motion. Then,*

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

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

2.2 Hypotheses

For unknown parameters α and β (of dimension $p \times r$ but not necessarily of full column rank), ρ_1, $\{\Gamma_j\}_{j=1}^{k-1}$ and μ_2, and for a given sequence $\{\varepsilon_t\}$ satisfying Assumption 2, (1) is the unrestricted heteroskedastic co-integrated VAR model, denoted in what follows as H_r. It will be convenient to write the model in the compact form

$$Z_{0t} = \alpha \beta^{\#\prime} Z_{1t} + \Psi^{\#} Z_{2t} + \varepsilon_t, \tag{4}$$

with $Z_{0t} := \Delta X_t$, $Z_{1t} := (X'_{t-1}, D'_{1t})'$, $Z_{2t} := (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}, D'_{2t})'$, $\beta^{\#} := (\beta', \rho_1')'$ and $\Psi^{\#} := (\Psi, \mu_2)$. If D_{it} is set equal to 0, it is understood that D_{it} is to be dropped from the definition of Z_{it}, $i = 1, 2$. We also define $p^{\#} := p + d_1$, the number of rows of $\beta^{\#}$.

Within this model we wish to test linear hypotheses on the co-integration parameters $\beta^{\#}$ and the adjustment coefficients α. Because $\beta^{\#}$ is only identified up to its column space, some restrictions are needed to identify the individual components of $\beta^{\#}$, before further (over-identifying) restrictions can be considered. Therefore, we normalise β by $c'\beta = I_r$ for some known $p \times r$ matrix of full column rank, so that $\beta = \bar{c} + c_\perp \beta_2$, where $\beta_2 = \bar{c}'_\perp \beta$ is a $(p - r) \times r$ matrix of free parameters; where $\bar{c} = c(c'c)^{-1}$ and $\bar{c}_\perp = c_\perp (c'_\perp c_\perp)^{-1}$. Defining

$$\beta_2^{\#} := \begin{pmatrix} \beta_2 \\ \rho_1 \end{pmatrix}, \qquad c^{\#} := \begin{pmatrix} c \\ 0 \end{pmatrix}, \qquad c_\perp^{\#} := \begin{pmatrix} c_\perp & 0 \\ 0 & I_{d_1} \end{pmatrix},$$

and $\bar{c}^{\#} = c^{\#}(c^{\#\prime}c^{\#})^{-1}$, $\bar{c}^{\#}_{\perp} = c^{\#}_{\perp}(c^{\#\prime}_{\perp}c^{\#}_{\perp})^{-1}$, we similarly have $\beta^{\#} = \bar{c}^{\#} + c^{\#}_{\perp}\beta^{\#}_2$, with $\beta^{\#}_2 = \bar{c}^{\#\prime}_{\perp}\beta^{\#}$. Based on this normalisation, we consider the following class of hypotheses on $\beta^{\#}_2$ and α:

$$\begin{aligned} H_{0\beta} \quad &: \quad R_\beta \operatorname{vec}\beta^{\#}_2 = q_\beta, \qquad H_{0\alpha} : R_\alpha \operatorname{vec}\alpha' = q_\alpha, \\ H_{0\alpha\beta} = H_{0\beta} \cap H_{0\alpha} \quad &: \quad R\theta = q, \end{aligned} \tag{5}$$

where R_β and R_α are matrices of dimensions $r_\beta \times (p^{\#} - r)r$ and $r_\alpha \times pr$, respectively, of full row rank, and q_β and q_α are $r_\beta \times 1$ and $r_\alpha \times 1$ vectors; furthermore, $R = \operatorname{diag}(R_\beta, R_\alpha)$, $q = (q'_\beta, q'_\alpha)$, and $\theta = ((\operatorname{vec}\beta^{\#}_2)', (\operatorname{vec}\alpha')')'$. The theory developed in this paper could be extended to more general non-linear restrictions, and restrictions linking α and β, but (5) appears to offer a sufficient level of generality for most practical purposes. In what follows, we will often use the notation H_0 in a generic sense to mean either $H_{0\beta}$, $H_{0\alpha}$ or $H_{0\alpha\beta}$.

For obtaining the PML estimators under the restrictions (5), needed to compute the PLR statistic, we rewrite the joint hypothesis $H_{0\alpha\beta}$ (together with the normalisation of $\beta^{\#}$) as:

$$H'_{0\alpha\beta} : \operatorname{vec}\beta^{\#} = H\phi + h, \quad \operatorname{vec}\alpha' = G\psi + g, \qquad (\phi, \psi) \in \mathbb{R}^{l_\phi} \times \mathbb{R}^{l_\psi}, \tag{6}$$

where $H = Q_\perp$ and $h = Q(Q'Q)^{-1}((\operatorname{vec} I_r)', q'_\beta)'$, with $Q = [\ (I_r \otimes c^{\#}) \quad (I_r \otimes \bar{c}^{\#}_{\perp})R'_\beta\ ]$, where $G = (R'_\alpha)_\perp$ and $g = R'_\alpha(R_\alpha R'_\alpha)^{-1}q_a$, and where ϕ and ψ are unrestricted parameters, of dimensions $l_\phi = (p^{\#} - r)r - r_\beta$ and $l_\psi = pr - r_\alpha$. Null hypotheses of the form (6), but with $g = 0$, were considered by Boswijk (1995) (see also Boswijk and Doornik (2004)), as a generalisation of the restrictions $\beta^{\#}_i = H_i\phi_i + h_i$ on the separate vectors of $\beta^{\#}$ considered by Johansen (1995). If only restrictions on $\beta^{\#}$ are considered (i.e., for testing $H_{0\beta}$), then $G = I_{pr}$ and $g = 0$; for hypotheses $H_{0\alpha}$ on only α, take $H = (I_r \otimes c^{\#}_{\perp})$ and $h = \operatorname{vec}\bar{c}^{\#}$.

3 Asymptotic Inference

In this section we analyse asymptotic inference on the class of hypotheses in (5) defined in the previous section, in the model (1) under Assumptions 1 and 2. As in Johansen (1996), the analysis will be based on the Gaussian pseudo-likelihood, derived from the assumption that ε_t is an i.i.d. $N(0, \Sigma)$ sequence. Assumption 2 implies that in general, this likelihood is based on a misspecified model, and the purpose of this section is to assess the asymptotic consequences of this mis-specification. We analyse the PLR test as well as a Wald test based on the PML variance matrix estimate of the parameter estimators.

To simplify notation, asymptotic results will be provided only for the model with a restricted constant, where $Z_{1t} = (X'_{t-1}, 1)'$ and $Z_{2t} = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1})'$, so $\Psi^{\#} = \Psi$. Analogous results can be obtained for models with a more general specification of the deterministic components, but at the cost of more involved notation.

In what follows, for a given vector of parameters, θ say, the unrestricted PML estimator for θ will be denoted $\hat{\theta}$, while the PML estimator obtained under the restrictions in (5) will be denoted $\tilde{\theta}$.

3.1 The PLR Test

The concentrated pseudo-log-likelihood in terms of the parameters α, $\beta^{\#}$ and Σ can be expressed in terms of the sample moment matrices

$$S_{ij} := M_{ij} - M_{i2}M_{22}^{-1}M_{2j}, \quad i,j = 0,1, \tag{7}$$

with $M_{ij} := T^{-1}\sum_{t=1}^{T} Z_{it}Z_{jt}'$, $i,j = 0,1,2$. Up to a constant, the pseudo-log-likelihood is given by

$$\ell(\alpha,\beta^{\#},\Sigma) = -\frac{T}{2}\log|\Sigma| - \frac{T}{2}\operatorname{tr}\Sigma^{-1}(S_{00} - 2\alpha\beta^{\#\prime}S_{10} + \alpha\beta^{\#\prime}S_{11}\beta^{\#}\alpha'). \tag{8}$$

If α is unrestricted, the log-likelihood may be further concentrated with respect to α and Σ to yield

$$\ell(\beta^{\#}) = -\frac{T}{2}\log\left|S_{00} - S_{01}\beta^{\#}(\beta^{\#\prime}S_{11}\beta^{\#})^{-1}\beta^{\#\prime}S_{10}\right| - \frac{Tp}{2}. \tag{9}$$

The maximiser of $\ell(\beta^{\#})$ under the normalisation $c^{\#\prime}\beta^{\#} = I_r$, see Johansen (1996), is given by $\hat{\beta}^{\#} := \hat{\beta}_u^{\#}(c^{\#\prime}\hat{\beta}_u^{\#})^{-1}$, where $\hat{\beta}_u^{\#} := [\hat{v}_1,\ldots,\hat{v}_r]$, with $\hat{v}_i, i = 1,\ldots,r$ the eigenvectors corresponding to r largest eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_r$ of the generalised eigenvalue problem $\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0$. The maximised log-likelihood is given by

$$\ell(\hat{\beta}^{\#}) = \ell(\hat{\beta}_u^{\#}) = -\frac{T}{2}\left(p + \log|\hat{\Sigma}|\right), \qquad \hat{\Sigma} := S_{00}\textstyle\prod_{i=1}^{r}(1-\hat{\lambda}_i).$$

Under the restrictions $\operatorname{vec}\beta^{\#} = H\phi + h, \operatorname{vec}\alpha' = G\psi$, the log-likelihood (8), possibly after concentrating out Σ, can be maximised over (ϕ,ψ) using a Newton-type algorithm. Alternatively, a switching algorithm can be used that exploits the fact that expressions for the partial maximisers $\tilde{\beta}^{\#}(\alpha,\Sigma)$, $\tilde{\alpha}(\beta^{\#},\Sigma)$ and $\tilde{\Sigma}(\alpha,\beta^{\#})$ of $\ell(\alpha,\beta^{\#},\Sigma)$ are available in closed form; see Boswijk (1995) and Boswijk and Doornik (2004) for further details. Letting $\tilde{\Sigma}$ denote the PML estimator of Σ under the restrictions, the PLR statistic is then given by

$$\mathrm{LR}_T := T\log|\tilde{\Sigma}|/|\hat{\Sigma}|.$$

In order to be able to discuss hypothesis testing on α and β separately, as well as jointly, we introduce the following notation: for $H_{0\alpha}$, denote the PLR statistic by $\mathrm{LR}_T(\alpha)$; and for $H_{0\beta}$, the PLR statistic is denoted by $\mathrm{LR}_T(\beta)$; for $H_{0\alpha\beta}$, the PLR statistic is denoted by $\mathrm{LR}_T(\alpha,\beta)$.

The next lemma characterises the asymptotic behaviour of the three sufficient statistics S_{00}, $S_{10} = S_{01}'$ and S_{11}. These results will subsequently be used to characterise the limiting behaviour of the likelihood function and, hence, of the PLR statistic.

Lemma 2. *Let X_t satisfy the model (1) under Assumptions 1 and 2. Let $B_T := \operatorname{diag}(T^{-1/2}\beta_\perp, 1)$ and $S_{1\varepsilon} := S_{10} - S_{11}\beta^{\#}\alpha'$, with $S_{ij}, i,j = 0,1$ as given in (7). Then,*

$$\beta^{\#\prime}S_{11}\beta^{\#} \xrightarrow{p} \bar{\Sigma}_{\beta\beta}, \qquad \beta^{\#\prime}S_{10} \xrightarrow{p} \bar{\Sigma}_{\beta 0}, \qquad S_{00} \xrightarrow{p} \bar{\Sigma}_{00}, \tag{10}$$

and

$$T^{1/2}\beta^{\#\prime}S_{1\varepsilon} \xrightarrow{w} N(0,\Omega), \tag{11}$$

where $\bar{\Sigma}_{\beta\beta}$ *and* Ω *are positive definite matrices, defined in the Appendix, and where* $\bar{\Sigma}_{\beta 0} := \bar{\Sigma}_{\beta\beta}\alpha'$ *and* $\bar{\Sigma}_{00} := \bar{\Sigma} + \alpha\bar{\Sigma}_{\beta\beta}\alpha'$*, with* $\bar{\Sigma} := \int_0^1 \Sigma(s)ds$*. Furthermore,*

$$\left(B_T' S_{11} B_T, T^{1/2} B_T' S_{1\varepsilon}\right) \xrightarrow{w} \left(\int_0^1 G(s)G(s)'ds, \int_0^1 G(s)dM(s)'\right), \tag{12}$$

where $G(u) := (M(u)'C'\beta_\perp, 1)'$*, and*

$$B_T' S_{11} \beta^{\#} \xrightarrow{p} 0. \tag{13}$$

The expressions for $\bar{\Sigma}_{\beta\beta}$ and Ω, derived in the proof of Lemma 2 in the Appendix, imply that in general $\Omega \neq \bar{\Sigma} \otimes \bar{\Sigma}_{\beta\beta}$. This implies that $T^{1/2}(\beta^{\#\prime} S_{11} \beta^{\#})^{-1/2} \beta^{\#\prime} S_{1\varepsilon} \hat{\Sigma}^{-1/2}$, which would be a natural candidate for a standardised version of $\beta^{\#\prime} S_{1\varepsilon}$ provided that $\hat{\Sigma} \xrightarrow{p} \bar{\Sigma}$, is in general not an asymptotically standard normal random vector. The property $\Omega = \bar{\Sigma} \otimes \bar{\Sigma}_{\beta\beta}$ only holds if both $\sigma(s) = \sigma$ (a constant unconditional volatility) and $\tau_{ij} = \mathbb{I}(i = j) I_{p^2}$ for all $i, j \geq 1$ (which excludes volatility clustering). This will have implications for testing hypotheses on α in the next theorem, which gives the asymptotic null distribution of the PLR statistic, together with some additional results on the consistency and asymptotic distribution of the PML estimators. For notational convenience, we only consider the case where the restrictions on $\beta^{\#}$ do not involve the constant term ρ_1, and hence relate to β_2 only.

Theorem 1. *Under the conditions of Lemma 2:*

(a) *the PML estimators of* $(\beta^{\#}, \alpha, \Psi)$ *are consistent, i.e.,* $\hat{\beta}^{\#} \xrightarrow{p} \beta^{\#}$*,* $\hat{\alpha} \xrightarrow{p} \alpha$*,* $\hat{\Psi} \xrightarrow{p} \Psi$*, and* $\hat{\Sigma} \xrightarrow{p} \bar{\Sigma}$*;*

(b) *the asymptotic distribution of* $\hat{\beta}_2$*,* $\hat{\rho}_1$ *and* $\hat{\alpha}$ *is given by*

$$\begin{pmatrix} T(\hat{\beta}_2 - \beta_2) \\ T^{1/2}(\hat{\rho}_1 - \rho_1) \end{pmatrix} \xrightarrow{w} \left(\int_0^1 G_c(s)G_c(s)'ds\right)^{-1} \int_0^1 G_c(s)dM_\alpha(s)',$$
$$T^{1/2}(\hat{\alpha} - \alpha)' \xrightarrow{w} N(0, [I_p \otimes \bar{\Sigma}_{\beta\beta}^{-1}]\Omega[I_p \otimes \bar{\Sigma}_{\beta\beta}^{-1}]),$$

where $G_c(u) := \mathrm{diag}(\beta_\perp' c_\perp (\beta_\perp' \beta_\perp)^{-1}, 1) G(u)$ *and* $M_\alpha(u) := (\alpha'\bar{\Sigma}^{-1}\alpha)^{-1}\alpha'\bar{\Sigma}^{-1}M(u)$*;*

(c) *under* $H_{0\beta}$ *and* $H_{0\alpha}$ *respectively,*

$$\mathrm{LR}_T(\beta) \xrightarrow{w} \mathcal{LR}_\infty(\beta), \qquad \mathrm{LR}_T(\alpha) \xrightarrow{w} \mathcal{LR}_\infty(\alpha) \tag{14}$$

where

$$\begin{aligned} \mathcal{LR}_\infty(\beta) &:= \left(R_\beta \,\mathrm{vec}\left(\int_0^1 G_c(s)G_c(s)'ds\right)^{-1} \int_0^1 G_c(s)dM_\alpha(s)'\right)' \\ &\quad \times \left(R_\beta \left[(\alpha'\bar{\Sigma}^{-1}\alpha)^{-1} \otimes \left(\int_0^1 G_c(s)G_c(s)'ds\right)^{-1}\right] R_\beta'\right)^{-1} \\ &\quad \times R_\beta \,\mathrm{vec}\left(\int_0^1 G_c(s)G_c(s)'ds\right)^{-1} \int_0^1 G_c(s)dM_\alpha(s)' \end{aligned} \tag{15}$$

$$\begin{aligned} \mathcal{LR}_\infty(\alpha) &:= Z'\left(R_\alpha[I \otimes \bar{\Sigma}_{\beta\beta}^{-1}]\Omega[I \otimes \bar{\Sigma}_{\beta\beta}^{-1}]R_\alpha'\right)^{1/2} \\ &\quad \times \left(R_\alpha[\bar{\Sigma} \otimes \bar{\Sigma}_{\beta\beta}^{-1}]R_\alpha'\right)^{-1} \left(R_\alpha[I \otimes \bar{\Sigma}_{\beta\beta}^{-1}]\Omega[I \otimes \bar{\Sigma}_{\beta\beta}^{-1}]R_\alpha'\right)^{1/2} Z \end{aligned} \tag{16}$$

and where $Z \sim N(0, I_{r_\alpha})$. *Moreover, under* $H_{0\alpha\beta}$,

$$\mathrm{LR}_T(\alpha, \beta) \xrightarrow{w} \mathcal{LR}_\infty(\beta) + \mathcal{LR}_\infty(\alpha) =: \mathcal{LR}_\infty(\alpha, \beta). \tag{17}$$

Notice that $\mathcal{LR}_\infty(\beta)$ *and* $\mathcal{LR}_\infty(\alpha)$ *are mutually independent if* $\varrho = 0$, *because where this condition holds* Z *is independent of* (G_c, M_α).

REMARK 3.1. It can be seen from part (b) of Theorem 1 that the normalised estimators, $\hat{\beta}_2$, $\hat{\rho}_1$ and $\hat{\alpha}$ attain exactly the same rates of consistency under heteroskedasticity of the form given in Assumption 2 as they do under the assumption of i.i.d. shocks; cf. Johansen (1996, Chapter 13). Moreover, it is also seen from part (a) that the PML estimates of the short-run dynamic parameter matrices, $\Gamma_1, \ldots, \Gamma_{k-1}$, also retain consistency under heteroskedasticity.

REMARK 3.2. Observe from (17) that the limiting null distribution of the joint PLR test statistic for $H_{0\alpha\beta}$ is given by the sum of the limiting null distributions for the corresponding $\mathrm{LR}_T(\alpha)$ and $\mathrm{LR}_T(\beta)$ statistics, both of which can be seen to depend on nuisance parameters arising from the heteroskedasticity present in the shocks. These condition, $\varrho = 0$, for the mutual independence of these two components entails that the conditional variance matrix of z_t is independent of lagged levels, thereby ruling out any statistical leverage effects.

REMARK 3.3. Consider the limiting random variable, $\mathcal{LR}_\infty(\beta)$ in (15), relating to the restrictions on β. The stochastic integral $\int_0^1 G_c(s) dM_\alpha(s)'$ clearly plays a key role in the asymptotic distribution of $\hat{\beta}_2^{\#}$. Both $G_c(\cdot)$ and $M_\alpha(\cdot)$ can be expressed as linear combinations of the continuous-time Gaussian martingale $M(\cdot) = \int_0^{\cdot} \sigma(s) dW(s)$. This means that these two processes are independent only if their cross-variation process

$$\begin{aligned} \langle G_c, M_\alpha \rangle(u) &= A\beta_\perp' C \langle M \rangle(u) \bar{\Sigma}^{-1} \alpha (\alpha' \bar{\Sigma}^{-1} \alpha)^{-1} \\ &= A\beta_\perp' C \int_0^u \Sigma(s) ds \bar{\Sigma}^{-1} \alpha (\alpha' \bar{\Sigma}^{-1} \alpha)^{-1} \end{aligned}$$

is zero for all $u \in [0, 1]$, where $A := \mathrm{diag}(\beta_\perp' c_\perp (\beta_\perp' \beta_\perp)^{-1}, 1)$. Although the property $C\alpha = \beta_\perp (\alpha_\perp' \Gamma \beta_\perp)^{-1} \alpha_\perp' \alpha = 0$ implies that $\langle G_c, M_\alpha \rangle(1) = 0$, this property does not extend to all u, so that in general, the two processes are not independent. This in turn implies that the distribution of $\int_0^1 G_c(s) dM_\alpha(s)'$ is in general not mixed Gaussian, which is a necessary condition for a quadratic form in this stochastic integral to have a χ^2 distribution. Clearly, if $\sigma(u)$ is constant, then G_c and M_α are independent vector Brownian motions, and here it is simple to show that $\mathcal{LR}_\infty(\beta)$ is $\chi^2(r_\beta)$ distributed, as is the case for i.i.d. shocks; see Johansen (1996).

REMARK 3.4. A further example where $G_c(\cdot)$ and $M_\alpha(\cdot)$ are independent occurs where $\Sigma(u)$ can be written as the product of a constant matrix Σ and a scalar time-varying process $v(u)$, corresponding to the case of "common volatility shocks"; see Remark 2.3 of Cavaliere, Rahbek and Taylor (2010b). However, in this example $\mathcal{LR}_\infty(\beta)$ will not be $\chi^2(r_\beta)$ distributed in general. To see why, observe that $\int_0^u \Sigma(s) ds \bar{\Sigma}^{-1} = \int_0^u v(s) ds \bar{v}^{-1} I_p$, where $\bar{v} = \int_0^1 v(s) ds$, so that

$\langle G_c, M_\alpha \rangle_u = 0$ for all u, and the distribution of $\int_0^1 G_c(s)dM_\alpha(s)'$ is again mixed Gaussian, but in this case with conditional variance matrix

$$\left[(\alpha'\bar{\Sigma}^{-1}\bar{\alpha})^{-1}\alpha'\bar{\Sigma}^{-1} \otimes I_{p^\#-r}\right] \int_0^1 [\Sigma(s) \otimes G_c(s)G_c(s)']ds \left[\bar{\Sigma}^{-1}\alpha'(\alpha'\bar{\Sigma}^{-1}\bar{\alpha})^{-1} \otimes I_{p^\#-r}\right].$$

Now

$$\begin{aligned}\int_0^1 [\Sigma(s) \otimes G_c(s)G_c(s)']ds &= \Sigma \otimes \int_0^1 v(s)G_c(s)G_c(s)'ds \\ &\neq \Sigma\bar{v} \otimes \int_0^1 G_c(s)G_c(s)'ds = \bar{\Sigma} \otimes \int_0^1 G_c(s)G_c(s)'ds\end{aligned}$$

unless $v(u) = 1$ and, hence, $\Sigma(u) = \Sigma$ for all $u \in [0,1]$. Consequently, in this case $\mathcal{LR}_\infty(\beta)$ will only be $\chi^2(r_\beta)$ distributed when $\sigma(u)$ is constant. However, we will subsequently show in Section 3.2 below that the independence of $G_c(\cdot)$ and $M_\alpha(\cdot)$ is necessary and sufficient for the Wald tests based on the use of a sandwich-type PML variance matrix to deliver χ^2 inference.

REMARK 3.5. Notice that inference on $\beta^\#$ is asymptotically unaffected by the possibility of volatility clustering in z_t (where the conditional variance matrix h_t may be correlated with lagged squares and cross-products), because the matrices $\tau_{ij} = E(z_t z_t' \otimes z_{t-i} z_{t-j}')$ do not enter the right member of (15).

REMARK 3.6. Turning to the limiting distribution in (16), relating to the restrictions on α, we observe that $\mathcal{LR}_\infty(\alpha)$ is a quadratic form in a standard normal vector. However, because the weight matrix in this quadratic form is, in general, not an identity matrix, $\mathcal{LR}_\infty(\alpha)$ will not have a $\chi^2(r_\alpha)$ distribution in general. As discussed below Lemma 2, standard inference does, however, obtain if both $\sigma(u)$ is constant and z_t does not display volatility clustering, so that $\tau_{ij} = \mathbb{I}(i = j)I_{p^2}$ for all $i, j \geq 1$. More generally, however, standard asymptotic inference is expected to be delivered by the PML-based Wald test, as will be discussed in the next subsection.

REMARK 3.7. Hansen (1992a) considers asymptotic inference on β_2 in a single equation Engle and Granger (1987) co-integrating regression model where the regressors are homoskedastic $I(1)$ processes but the errors display non-stationary volatility (specifically, volatility follows an $I(1)$ process in his set-up). In the notation of the present paper, this corresponds to the case where $\alpha_\perp' M(u)$ is a Brownian motion with constant variance matrix $\alpha_\perp'\bar{\Sigma}\alpha_\perp$, and where $M_\alpha(u)$ has a time-varying (conditional) variance matrix. Hansen (1992a) shows that mixed normal inference arises if the Brownian motion driving M_α is independent of $\alpha_\perp' M(u)$ and the stochastic volatility process, which implies that $\alpha'\bar{\Sigma}^{-1}\Sigma(u)\alpha_\perp = 0$ for all $u \in [0,1]$. Notice that by a suitable choice of the matrix square root $\sigma(\cdot)$ of $\Sigma(\cdot)$, we may write

$$\begin{bmatrix} \alpha'\bar{\Sigma}^{-1} \\ \alpha_\perp' \end{bmatrix} \sigma(u) = \begin{bmatrix} \omega_{11}(u) & \omega_{12}(u) \\ 0 & \omega_{22} \end{bmatrix},$$

where $\omega_{22}\omega_{22}' = \alpha_\perp'\bar{\Sigma}\alpha_\perp$ and $\omega_{11}(u)\omega_{11}(u)' + \omega_{12}(u)\omega_{12}(u)' = \alpha'\bar{\Sigma}^{-1}\Sigma(u)\bar{\Sigma}^{-1}\alpha$. We also know that $\int_0^1 \alpha'\bar{\Sigma}^{-1}\Sigma(u)\alpha_\perp du = \int_0^1 \omega_{12}(u)\omega_{22}' du = \bar{\omega}_{12}\omega_{22}' = 0$. However, this by itself does not guarantee or require that $\omega_{12}(u) = 0$ for all u. It is the specific structure of the model considered in

Hansen (1992a) which implies that $\alpha'\bar{\Sigma}^{-1}M(u) = \int_0^u \omega_{11}(s)dW_1(s)$ and $\alpha'_\perp M(u) = \omega_{22}W_2(u)$, and hence yields the independence of the two processes. □

3.2 Wald Test

A Wald test statistic for null hypotheses as in (5) takes the form

$$\mathrm{W}_T := (R\hat{\theta} - q)'(R\widehat{\mathrm{Var}}[\hat{\theta}]R')^{-1}(R\hat{\theta} - q), \tag{18}$$

where $\widehat{\mathrm{Var}}[\hat{\theta}]$ is an estimate of the asymptotic (conditional) variance matrix of $\hat{\theta}$. As in the previous section, we denote by $\mathrm{W}_T(\alpha)$, $\mathrm{W}_T(\beta)$ and $\mathrm{W}_T(\alpha, \beta)$ test statistics for hypotheses on α only, on β only, and jointly on α and β, respectively. We consider a PML-based variance matrix, which may be expressed as follows. Let $\gamma := (\theta', (\mathrm{vec}\,\Psi')')'$, and define the following estimates of the information matrix for γ based on minus the Hessian matrix and the outer product of gradients, respectively:

$$\begin{aligned}
\mathcal{H} := -\frac{\partial^2 \ell(\gamma)}{\partial\gamma\partial\gamma'} &= T\begin{bmatrix} J'(\Sigma^{-1}\otimes M_{11})J & J'(\Sigma^{-1}\otimes M_{12}) \\ (\Sigma^{-1}\otimes M_{21})J & (\Sigma^{-1}\otimes M_{22}) \end{bmatrix}, \\
\mathcal{I} := \sum_{t=1}^T \frac{\partial \ell_t(\gamma)}{\partial\gamma}\frac{\partial \ell_t(\gamma)}{\partial\gamma'} &= \sum_{t=1}^T \begin{bmatrix} J'(\Sigma^{-1}\varepsilon_t\varepsilon_t'\Sigma^{-1}\otimes Z_{1t}Z_{1t}')J & J'(\Sigma^{-1}\varepsilon_t\varepsilon_t'\Sigma^{-1}\otimes Z_{2t}Z_{1t}') \\ (\Sigma^{-1}\varepsilon_t\varepsilon_t'\Sigma^{-1}\otimes Z_{2t}Z_{1t}')J & (\Sigma^{-1}\varepsilon_t\varepsilon_t'\Sigma^{-1}\otimes Z_{2t}Z_{2t}') \end{bmatrix},
\end{aligned}$$

where $J := \partial\,\mathrm{vec}(\beta^{\#}\alpha')/\partial\theta' = [\ (\alpha\otimes c_\perp^{\#})\quad (I_p\otimes\beta^{\#})\ ]$. Then $\widehat{\mathrm{Var}}[\hat{\theta}] = \begin{pmatrix} I_{l_\theta} & 0 \end{pmatrix}\hat{\mathcal{H}}^{-1}\hat{\mathcal{I}}\hat{\mathcal{H}}^{-1}\begin{pmatrix} I_{l_\theta} \\ 0 \end{pmatrix}$, where $l_\theta := (p^{\#} - r)r + pr = \dim\theta$, and $\hat{\mathcal{H}}$ and $\hat{\mathcal{I}}$ are as defined above, evaluated at $\gamma = \hat{\gamma}$. In these expressions, the dependence of ℓ on Σ is suppressed, and derivatives with respect to the variance parameters are not taken into account, due to the asymptotic block-diagonality of the Hessian matrix with respect to the regression and variance parameters.

In Theorem 2 we detail the asymptotic null distribution of the Wald statistic, W_T of (18). This follows an important preparatory result relating to the scaled partial sum of the PML residuals from the estimation of (1), $\hat{\varepsilon}$, given in Lemma 3.

Lemma 3. *Under the conditions of Lemma 2,* $\hat{V}_T(u) := T^{-1}\sum_{t=1}^{\lfloor Tu\rfloor}\hat{\varepsilon}_t\hat{\varepsilon}_t' \xrightarrow{p} \int_0^u \Sigma(s)ds =: V(u)$, *uniformly in* $u \in [0,1]$.

Theorem 2. *Under the conditions of Lemma 2, and under* $H_{0\alpha}$ *and* $H_{0\beta}$ *respectively,*

$$\mathrm{W}_T(\beta) \xrightarrow{w} \mathcal{W}_\infty(\beta), \qquad \mathrm{W}_T(\alpha) \xrightarrow{w} Z'Z \tag{19}$$

where $Z \sim N(0, I_{r_\alpha})$,

$$\begin{aligned}
\mathcal{W}_\infty(\beta) := &\left(R_\beta\left(I_r\otimes K^{-1}\right)\mathrm{vec}\int_0^1 G_c(s)dM_\alpha(s)'\right)' \\
&\times\left(R_\beta\left(I_r\otimes K^{-1}\right)\left\langle\mathrm{vec}\int_0^{\cdot} G_c(s)dM_\alpha(s)'\right\rangle(1)\left(I_r\otimes K^{-1}\right)R_\beta'\right)^{-1} \\
&\times R_\beta\left(I_r\otimes K^{-1}\right)\mathrm{vec}\int_0^1 G_c(s)dM_\alpha(s)'
\end{aligned}$$

and where $K := \int_0^1 G_c(s)G_c(s)'ds$,

$$\left\langle \int_0^{\cdot} \operatorname{vec} G_c(s)dM_\alpha(s)' \right\rangle (u) = \int_0^u \left[\langle M_\alpha \rangle (s) \otimes G_c(s)G_c(s)'\right] ds,$$

with $\langle M_\alpha \rangle (s) = (\alpha'\bar{\Sigma}^{-1}\alpha)^{-1}\alpha'\bar{\Sigma}^{-1}\Sigma(s)\bar{\Sigma}^{-1}\alpha(\alpha'\bar{\Sigma}^{-1}\alpha)^{-1}$. *Moreover, under* $H_{0\alpha\beta}$,

$$\mathrm{W}_T(\alpha,\beta) \xrightarrow{w} \mathcal{W}_\infty(\beta) + Z'Z =: \mathcal{W}_\infty(a,\beta), \tag{20}$$

where $\mathcal{W}_\infty(\beta)$ *and* $Z'Z$ *are independent if* $\varrho = 0$.

Remark 3.8. It is seen from (19) that, unlike the PLR test, the PML-based Wald test leads to standard asymptotic χ^2 inference on the adjustment parameters α without the need for any further assumptions. However, Wald-based inference on β is in general not mixed Gaussian in the limit, because the processes G_c and M_α which appear in $\mathcal{W}_\infty(\beta)$ are in general not independent, as discussed in Remark 3.4. However, and in contrast to the component $\mathcal{LR}_\infty(\beta)$ appearing in the asymptotic null distribution of the LR_T statistic, the quadratic form in $\mathcal{W}_\infty(\beta)$ now involves a weight matrix that is the inverse quadratic variation of the stochastic integral. This implies that $\mathcal{W}_\infty(\beta)$ will be $\chi^2(r_\beta)$ distributed when G_c and M_α are independent, as happens in the "common volatility shocks" example outlined in Remark 3.4. □

For completeness, we summarise the asymptotic null distribution of the W_T statistic in the case where G_c and M_α are independent in following corollary.

Corollary 1. *Under the conditions of Lemma 2, and if* $\alpha_\perp'\Sigma(u)\bar{\Sigma}^{-1}\alpha = 0$ *for all* $u \in [0,1]$, *then under* $H_{0\beta}$, $\mathrm{W}_T(\beta) \xrightarrow{w} \chi^2(r_\beta)$. *Moreover, if in addition* $\varrho = 0$ *then under* $H_{0\alpha\beta}$, $\mathrm{W}_T(\alpha,\beta) \xrightarrow{w} \chi^2(r_\beta + r_\alpha)$.

We conclude this section by providing a theorem detailing the consistency of the restricted PML estimators, which will subsequently be needed to prove the consistency of our proposed method of bootstrap inference outlined in the next section.

Theorem 3. *Let the conditions of Lemma 2 hold. Then, under the restrictions in (5):* $\tilde{\beta}_2 - \beta_2 = O_p(T^{-1})$, $\tilde{\rho}_1 - \rho_1 = O_p(T^{-1/2})$, $\tilde{\alpha} - \alpha = O_p(T^{-1/2})$, *and* $\tilde{\Psi} - \Psi = O_p(T^{-1/2})$. *Furthermore, denoting the associated PML residuals from restricted estimation of* (1) *by* $\tilde{\varepsilon}_t$, $\tilde{V}_T(u) := T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \tilde{\varepsilon}_t\tilde{\varepsilon}_t' \xrightarrow{p} V(u)$, *uniformly in* $u \in [0,1]$.

4 Bootstrap Inference

In this section we outline our proposed wild bootstrap-based implementations of the PLR and Wald tests from Section 3. We provide sufficient conditions for the wild bootstrap implementations of the PLR and Wald tests to be asymptotically valid under heteroskedasticity of the form given in Assumption 2, although as we will show in some cases these conditions will need to be strengthened somewhat.

The following notation will be used throughout this section: P^* and E^* respectively denote the probability and expectation conditional on the realization of the original sample. Moreover, for a given sequence X_T^* computed on the bootstrap data, with the notations $X_T^* = X + o_p^*(1)$, in probability, and $X_T^* \xrightarrow{p^*}_p X$ we mean that $P^*(|X_T^* - X| > \epsilon) \to 0$ in probability, for any $\epsilon > 0$ as $T \to \infty$. Finally, '$\xrightarrow{w^*}_p$' denotes weak convergence in probability (Giné and Zinn (1990); Hansen (1996)); that is, $X_T^* \xrightarrow{w^*}_p X$ if $\sup_{x\in\mathbb{R}} |P^*(X_T^* \leq x) - P(X \leq x)| \to_p 0$, as $T \to \infty$.

4.1 Wild Bootstrap Algorithms

We first outline our proposed algorithm which draws on the wild bootstrap principle; see, *inter alia*, Wu (1986) and Mammen (1993). The bootstrap algorithm we propose is based on the restricted Gaussian PML parameter estimates obtained by estimating the model in (1) under the null hypothesis H_0 of (5), as outlined in 3. More specifically, let $\tilde{\theta} := \{\tilde{\alpha}, \tilde{\beta}, \tilde{\rho}_1, \tilde{\Psi}^{\#}, \tilde{\mu}_2\}$ denote the restricted PML estimators of $\theta := \{\alpha, \beta, \rho_1, \Psi^{\#}, \mu_2\}$, define $\tilde{\beta}^{\#} := (\tilde{\beta}', \tilde{\rho}_1')'$. Recall that $\tilde{\varepsilon}_t := Z_{0t} - \tilde{\alpha}\tilde{\beta}^{\#\prime} Z_{1t} - \tilde{\Psi}^{\#} Z_{2t}$, with Z_{it}, $i = 0, 1, 2$, as defined in Section 2.2, denote the corresponding PML residuals from the estimation of (1) under H_0.

When applying bootstrap methods in the context of autoregressive time series models, it is common practice to check that the bootstrap data generated in Algorithm 1 will satisfy the $I(1, r)$ conditions (cf. Assumption 1); see, in particular, Swensen (2006) and Cavaliere, Rahbek and Taylor (2010a,b, 2012, 2014). In the context of the setting considered in this paper this entails checking that the equation $|\tilde{A}(z)| = 0$, with $\tilde{A}(z) := (1-z) I_p - \tilde{\alpha}\tilde{\beta}' z - \sum_{i=1}^{k-1} \tilde{\Gamma}_i z (1-z)$, has roots either equal to 1 or outside the unit circle. Where this condition is not satisfied the bootstrap samples generated in step (ii) of Algorithm 1 will contain explosive roots which could be attributable to either finite sample bias in the parameter estimates, or might indicate that the underlying model is dynamically misspecified. As such, the safe strategy would be not to use Algorithm 1 (or indeed the corresponding asymptotic tests) in such cases, although in the case where the model is correctly specified the evidence from a large simulation study reported in Cavaliere, Taylor and Trenkler (2015) suggests that, at least for the case of the bootstrap PLR tests of co-integration rank considered in Cavaliere, Rahbek and Taylor (2012, 2014), the root check can be safely ignored in practice. Reassuringly, in the Monte Carlo experiments reported in this paper we also found such violations to be extremely rare; see Section 5.

Algorithm 1 (wild bootstrap):

(i) Compute the re-centered residuals $\tilde{\varepsilon}_{c,t} := \tilde{\varepsilon}_t - T^{-1}\sum_{i=1}^{T} \tilde{\varepsilon}_t$ and construct the bootstrap errors $\varepsilon_t^* := \tilde{\varepsilon}_{c,t} w_t$, where $w_t, t = 1, \ldots, T$, is an i.i.d. sequence with $E(w_t) = 0$, $E(w_t^2) = 1$ and $E(w_t^4) < \infty$.

(ii) Construct the bootstrap sample $\{X_t^*\}$ from the recursion

$$\Delta X_t^* = \tilde{\alpha}\tilde{\beta}^{\#\prime} X_{t-1}^* + \sum_{j=1}^{k-1} \tilde{\Gamma}_j \Delta X_{t-j}^* + \tilde{\alpha}\tilde{\rho}_1' D_{1t} + \tilde{\mu}_2 D_{2t} + \varepsilon_t^*,\ t = 1, \ldots, T, \tag{21}$$

with the T bootstrap errors ε_t^* generated in step (i) and with initial values $X_t^* = X_t$ for $t = -k+1, \ldots, 0$.

(iii) Using the bootstrap sample, $\{X_t^*\}$, compute the bootstrap test statistic S_T^*, where S_T^* generically denotes either the PLR or the Wald statistic, as detailed in Section 3.1. Define the corresponding p-value as $P_T^* := 1 - G_T^*(S_T)$ with $G_T^*(\cdot)$ denoting the conditional (on the original data) cumulative density function (cdf) of S_T^*.

(iv) The wild bootstrap test of H_0 at level ξ rejects if $P_T^* \leq \xi$.

REMARK 4.1. In the context of stationary data, it is often found (for a review, see Davidson and Flachaire (2008)) that improved accuracy can be achieved by generating the pseudo-data according to an asymmetric distribution such as the Mammen (1993) distribution: $P(w_t = -0.5(\sqrt{5}-1)) = 0.5(\sqrt{5}+1)/\sqrt{5} =: \pi$, $P(w_t = 0.5(\sqrt{5}+1)) = 1-\pi$. Two other commonly used distributions are the two-point distribution $P(w_t = -1) = P(w_t = 1) = 0.5$ and an i.i.d. $N(0,1)$ sequence. The large sample properties of the resulting bootstrap tests are not affected by this choice. In simulations we found that these three gave very similar small sample performance, and so the results presented in Section 5 relate to the use of the $N(0,1)$ distribution for w_t.

REMARK 4.2. The parameter estimates and associated residuals from (1) which are used in constructing the bootstrap sample data in steps (i) and (ii), are obtained under the restriction of the null hypothesis, H_0 of (5). As suggested in Omtzigt and Fachin (2006), it would also be possible to estimate these parameters without imposing the null (i.e., using the unrestricted PML estimators detailed in Section 3.1 and in Theorem 1), and to subsequently calculate a bootstrap test statistic for the hypothesis $R\theta = R\hat{\theta}$. Unreported simulations indicate that the bootstrap based on restricted estimates is largely preferred. Hence, throughout this section for economy of discussion we will only explicitly discuss the bootstrap based on restricted estimates.

REMARK 4.3. The unknown cdf $G_T^*(\cdot)$ required in step (iii) of Algorithm 1 can be approximated by generating B (conditionally) independent bootstrap statistics, $S_{T:b}^*$, $b = 1, \ldots, B$, computed as in Algorithm 1. The simulated bootstrap p-value for S_T is then computed as $\tilde{P}_T^* := B^{-1} \sum_{b=1}^{B} \mathbb{I}(S_{T:b}^* > S_T)$, and is such that $\tilde{P}_T^* \xrightarrow{a.s.} P_T^*$ as $B \to \infty$. The choice of B is discussed by, *inter alia*, Andrews and Buchinsky (2000) and Davidson and MacKinnon (2000). □

4.2 Bootstrap Asymptotic Theory

In this section we provide results on the asymptotic properties of the bootstrap PLR and Wald statistics from Algorithm 1. In doing so we establish the conditions under which the bootstrap tests the first-order asymptotically valid.

For the bootstrap asymptotic theory related to tests of hypotheses involving α, we will need to strengthen Assumption 2 by restricting z_t (and hence ε_t) to have finite 8^+ moments.

Assumption 2'. Assumption 2 holds with $\sup_t E\left\|z_t\right\|^{4r} < \infty$ for some $r > 2$.

Before detailing the large sample behaviour of the bootstrap PLR and Wald statistics, we first need to establish two preparatory Lemmas. The first of these is the counterpart of Lemma 1 for the wild bootstrap shocks.

Lemma 4. *Let the conditions of Lemma 2 hold, and let ε_t^* be defined as in step (i) of Algorithm 1. Then,*

$$\begin{aligned}\left(M_T^*(\cdot), \int_0^1 M_T^*(s)dM_T^*(s)'\right) &:= \left(\frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t^*, \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i^*\right)\varepsilon_t^{*\prime}\right) \\ &\xrightarrow{w^*}_p \left(M(\cdot), \int_0^1 M(s)dM(s)'\right),\end{aligned}$$

where $M(\cdot) := \int_0^{\cdot}\sigma(s)dW(s)$ is a p-variate continuous martingale defined in terms of the p-variate Brownian motion $W(\cdot)$.

REMARK 4.4. Lemma 4 establishes that the two scaled cumulated functions considered of the wild bootstrap errors can replicate the limiting process of the corresponding quantities formed from the original shocks, ε_t; cf. Lemma 1. Notice, however, that for this this result to hold Assumption 1 must be satisfied, since this is needed to ensure that the restricted PML estimate, $\tilde{\theta}$, from (1) is consistent, as demonstrated in Theorem 3. □

The next lemma, which is the bootstrap counterpart of Lemma 2, characterises the asymptotic behaviour of the three bootstrap sufficient statistics, S_{00}^*, $S_{10}^* = S_{01}^{*\prime}$ and S_{11}^*. These results will subsequently be used in determining the limiting behaviour of the bootstrap PLR statistic.

Lemma 5. *Let X_t^* be generated as in Algorithm 1 and let $B_T := \mathrm{diag}(T^{-1/2}\tilde{\beta}_\perp, 1)$ and $S_{1\varepsilon}^* := S_{10}^* - S_{11}^*\tilde{\beta}^{\#}\tilde{\alpha}'$, with $S_{ij}^*, i,j = 0,1$ defined analogously to (7) but for the bootstrap data. Then, under the conditions of Lemma 2,*

$$\tilde{\beta}^{\#\prime}S_{11}^*\tilde{\beta}^{*\#} \xrightarrow{p^*}_p \bar{\Sigma}_{\beta\beta}, \qquad \tilde{\beta}^{\#\prime}S_{10}^* \xrightarrow{p^*}_p \bar{\Sigma}_{\beta 0}, \qquad S_{00}^* \xrightarrow{p^*}_p \bar{\Sigma}_{00}\ , \tag{22}$$

where $\bar{\Sigma}_{\beta 0}$, $\bar{\Sigma}_{00}$ and $\bar{\Sigma}_{\beta\beta}$ are as previously defined. Furthermore,

$$\left(B_T'S_{11}^*B_T, T^{1/2}B_T'S_{1\varepsilon}^*\right) \xrightarrow{w^*}_p \left(\int_0^1 G(s)G(s)'ds, \int_0^1 G(s)dM(s)'\right), \tag{23}$$

where $G(u) := (M(u)'C'\beta_\perp, 1)'$, and

$$B_T'S_{11}^*\tilde{\beta}^{\#} \xrightarrow{p^*}_p 0\ . \tag{24}$$

Finally, if Assumption 2 is strengthened to Assumption 2', then

$$T^{1/2}\tilde{\beta}^{\#\prime}S_{1\varepsilon}^* \xrightarrow{w^*}_p N(0, \Omega^\dagger), \tag{25}$$

where $\Omega^\dagger$ is a positive definite matrix, defined in the Appendix.

REMARK 4.5. It is useful to compare the asymptotic distributions involving S^*_{ij} in Lemma 5 with the corresponding distributions in Lemma 2. In the non-stationary directions of the system, both $\left(B'_T S_{11} B_T, T^{1/2} B'_T S_{1\varepsilon}\right)$ and its bootstrap counterpart $\left(\tilde{B}'_T S^*_{11} B_T, T^{1/2} \tilde{B}'_T S^*_{1\varepsilon}\right)$ have the same limiting distributions; compare (12) and (23). These distributions depend on the time series behaviour of the unconditional volatility of the shocks (through the function $\sigma(\cdot)$ of Assumption 2(a)) but not on the dynamics of the conditional volatility of z_t. Conversely, in the stationary directions of the system the wild bootstrap cannot, in general, replicate the correct limiting distributions. For example, the limiting variance of $T^{1/2}\beta^{*\#\prime} S^*_{1\varepsilon}$, which is given by $\Omega^\dagger$ (see equation (25)), is not equivalent to the limiting variance Ω of $T^{1/2}\beta^{\#\prime} S_{1\varepsilon}$ (see equation (11)). This disparity occurs because Ω depends on the cross moments, $E(z_t z'_t \otimes z_{t-i} z'_{t-j}) = \tau_{ij}$, which the wild bootstrap cannot replicate since, conditionally on the sample data,

$$\begin{aligned} E^*(\varepsilon^*_t \varepsilon^{*\prime}_t \otimes \varepsilon^*_{t-i} \varepsilon^{*\prime}_{t-j}) &= E^*(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t w_t^2 \otimes \tilde{\varepsilon}_{t-i} \tilde{\varepsilon}'_{t-j} w_{t-i} w_{t-j}) \\ &= \left(\tilde{\varepsilon}_t \tilde{\varepsilon}'_t E\left(w_t^2\right)\right) \otimes \left(\tilde{\varepsilon}_{t-i} \tilde{\varepsilon}'_{t-j} E\left(w_{t-i} w_{t-j}\right)\right) = 0 \end{aligned}$$

for all $i \neq j$. Therefore, it is anticipated that bootstrap PLR test statistics for hypotheses that include restrictions on α, and which therefore involve $\tilde{\beta}^{\#\prime} S^*_{1\varepsilon}$, will not be asymptotically pivotal under the null unless $\tau := \sup_{i \neq j} \|\tau_{ij}\| = 0$. □

We are now in a position to detail the asymptotic null properties of the wild bootstrap PLR statistic from Algorithm 1. Here, the bootstrap PLR statistics for tests on α, tests on β, and tests on α and β are respectively denoted by $\mathrm{LR}^*_T(\alpha)$, $\mathrm{LR}^*_T(\beta)$ and $\mathrm{LR}^*_T(\alpha, \beta)$.

Theorem 4. *Let the conditions of Lemma 5 hold. Then:*

(a) *the asymptotic distribution of $\hat{\beta}^*_2$, $\hat{\rho}^*_1$ and $\hat{\alpha}^*$ is given by*

$$\begin{pmatrix} T(\hat{\beta}^*_2 - \tilde{\beta}_2) \\ T^{1/2}(\hat{\rho}^*_1 - \tilde{\rho}_1) \end{pmatrix} \xrightarrow{w^*}_p \left(\int_0^1 G_c(s) G_c(s)' ds\right)^{-1} \int_0^1 G_c(s) dM_\alpha(s)',$$

where $G_c(u) = \mathrm{diag}(\beta'_\perp c_\perp (\beta'_\perp \beta_\perp)^{-1}, 1) G(u)$ and $M_\alpha(u) = (\alpha' \bar{\Sigma}^{-1} \alpha)^{-1} \alpha' \bar{\Sigma}^{-1} M(u)$.

If, in addition Assumption 2 is strengthened by Assumption 2′, then

$$T^{1/2}(\hat{\alpha}^* - \tilde{\alpha})' \xrightarrow{w^*}_p N(0, [I_p \otimes \bar{\Sigma}^{-1}_{\beta\beta}] \Omega^\dagger [I_p \otimes \bar{\Sigma}^{-1}_{\beta\beta}]),$$

(c) *Under $H_{0\beta}$, $\mathrm{LR}^*_T(\beta) \xrightarrow{w^*}_p \mathcal{LR}_\infty(\beta)$, and, if Assumption 2 is strengthened to Assumption 2′, then under $H_{0\alpha}$, $\mathrm{LR}^*_T(\alpha) \xrightarrow{w^*}_p \mathcal{LR}^\dagger_\infty(\alpha)$,and finally under $H_{0\alpha\beta}$,*

$$\mathrm{LR}^*_T(\alpha, \beta) \xrightarrow{w^*}_p \mathcal{LR}_\infty(\beta) + \mathcal{LR}^\dagger_\infty(\alpha) \tag{26}$$

where $\mathcal{LR}_\infty(\beta)$ is as defined in (15) and $\mathcal{LR}^\dagger_\infty(\alpha)$, which is independent of $\mathcal{LR}_\infty(\beta)$, is defined as

$$\begin{aligned} \mathcal{LR}^\dagger_\infty(\alpha) := \quad & Z' \left(R_\alpha [I \otimes \bar{\Sigma}^{-1}_{\beta\beta}] \Omega^\dagger [I \otimes \bar{\Sigma}^{-1}_{\beta\beta}] R'_\alpha\right)^{1/2} \\ & \times \left(R_\alpha [\bar{\Sigma} \otimes \bar{\Sigma}^{-1}_{\beta\beta}] R'_\alpha\right)^{-1} \left(R_\alpha [I \otimes \bar{\Sigma}^{-1}_{\beta\beta}] \Omega^\dagger [I \otimes \bar{\Sigma}^{-1}_{\beta\beta}] R'_\alpha\right)^{1/2} Z, \end{aligned}$$

where $Z \sim N(0, I_{r_\alpha})$*, independent of* (G_c, M_α)*.*

For general linear hypotheses on α and $\beta^{\#}$, the limiting null distribution of the bootstrap PLR statistic LR_T^* from Algorithm 1 can be seen to depend on the limiting variance matrix $\Omega^\dagger$, rather than on the limiting variance matrix Ω as is the case for the original LR_T statistic; cf. part (c) of Theorem 1. Consequently, the bootstrap PLR test will not have the same first-order limiting null distribution as the original PLR statistic unless $\Omega^\dagger = \Omega$. This equality holds under the condition $\tau = 0$. Where this holds, $\mathcal{LR}_\infty^\dagger(\alpha)$ and $\mathcal{LR}_\infty(\alpha)$ coincide, which means that $\mathrm{LR}_T^*(\alpha) \xrightarrow{w^*}_p \mathcal{LR}_\infty(\alpha)$. For joint hypotheses on α and β, in addition to $\tau = 0$ it is also necessary that the condition that $\varrho = 0$ holds.

For completeness we formalise this result in the following corollary, which gives the conditions under which the bootstrap PLR test is guaranteed to be first-order asymptotically valid. Here P_T^* denotes the (wild bootstrap) p-value associated with the PLR test statistic.

Corollary 2. *Let the conditions of Lemma 5 hold. Then, under* $H_{0\beta}$*,*

$$P_T^* \xrightarrow{w} U[0,1] \tag{27}$$

where $U[0,1]$ *denotes a uniform distribution on* $[0,1]$*. Under* $H_{0\alpha}$*, (27) holds provided that Assumption 2′ holds with* $\tau = 0$*. Finally, under* $H_{0\alpha\beta}$*, (27) holds provided that Assumption 2′ holds with* $\tau = 0$ *and* $\varrho = 0$*.*

Remark 4.6. Notice from the results in Corollary 2 that neither Assumption 2′ nor the restriction that either $\tau = 0$ or $\varrho = 0$ need hold when testing on $\beta^{\#}$ alone.

Remark 4.7. Under the alternative hypothesis, it is proved in Cavaliere, Nielsen and Rahbek (2015) for the i.i.d. bootstrap and under the assumption that the shocks ε_t are i.i.d., that the bootstrap PLR statistic is of $O_{p^*}(1)$, in probability. An immediate consequence of this result is that the bootstrap PLR test is consistent, due to the divergence of the standard LR_T statistic under the alternative hypothesis. While their framework covers hypotheses on β only (more specifically, hypotheses that fully specify the co-integration space), we conjecture that their result can be extended to our framework by using the asymptotic results given here and that this result will also hold for the corresponding Wald-based tests. □

We conclude this section by detailing the asymptotic null distribution of the bootstrap Wald statistic, W_T^*, from Algorithm 1. Here we let P_T^* denote the (wild bootstrap) p-value associated with the Wald test statistic.

Theorem 5. *Let the conditions of Lemma 5 hold. Then under* $H_{0\beta}$*,* $\mathrm{W}_T^*(\beta) \xrightarrow{w^*}_p \mathcal{W}_\infty(\beta)$*, where* $\mathcal{W}_\infty(\beta)$ *is as given in* (20)*. Under* $H_{0\alpha}$*, if Assumption 2' holds then* $\mathrm{W}_T^*(\alpha) \xrightarrow{w^*}_p Z'Z$*, where* $Z \sim N(0, I_{r_\alpha})$*. Under* $H_{0\alpha\beta}$*, if Assumption 2' holds and* $\varrho = 0$*, then* $\mathrm{W}_T^*(\alpha, \beta) \xrightarrow{w^*}_p \mathcal{W}_\infty(\alpha, \beta)$*, where* $\mathcal{W}_\infty(\alpha, \beta)$ *is as given in* (20)*. Consequently, in each of these three cases,* $P_T^* \xrightarrow{w} U[0,1]$*.*

REMARK 4.8. A comparison of the results in Theorem 5 and Corollary 2 highlights the result that the wild bootstrap implementation of Wald tests for hypotheses which involve α do not require the restriction $\tau = 0$ to hold in order to be (first-order) asymptotically valid, in contrast to the corresponding wild bootstrap PLR tests. This because the use of PML standard errors is sufficient to deliver asymptotically χ^2 Wald test statistics for hypotheses on α, as follows from Theorem 2, and the wild bootstrap replicates this asymptotic χ^2 null distribution. □

5 Monte Carlo Simulations

In this section we use Monte Carlo simulation methods to compare the finite sample performance of the asymptotic PLR and Wald tests from Section 3 with the corresponding wild bootstrap tests from Section 4. For comparison, we also report results for standard i.i.d. bootstrap versions of the PLR and Wald tests. We note that the i.i.d. bootstrap statistics are first order asymptotic equivalent to the standard PLR and Wald tests and hence the i.i.d. bootstrap tests are asymptotically valid if and only if the corresponding asymptotic tests are valid.

Our simulation DGP is the VAR(1) process of dimension $p = 2, 3$ and 4, with co-integrating rank $r = 1$ and

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t, \qquad \alpha' = (-0.2, 0'_{p-1}), \qquad \beta' = (1, 0'_{p-1}).$$

where 0_m denotes an m-vector of zeros. The process is initialised at $X_0 = 0$ and we consider sample sizes of $T = 100, 200$ and 400. A restricted constant term is included in the estimation. All experiments are run over $10,000$ Monte Carlo replications using $B = 499$ bootstrap repetitions. In the context of Algorithm 1, any samples violating the root check conditions in step (ii) are discarded. For each bootstrap procedure the observed (rounded) frequency of such violations was below 0.1% in all cases.

As in Section 2, the errors are defined as $\varepsilon_t = \sigma_t z_t$, z_t being a martingale difference sequence with unit unconditional variance matrix. Three versions of the unconditional variance matrix $\Sigma_t = \sigma_t \sigma'_t$ are considered. Specifically, with ι_m denoting an m-vector of ones, we consider

$$\begin{aligned} \Sigma_t^{(1)} &= (1-\rho) I_p + \rho \iota_p \iota'_p =: \bar{\Sigma} \\ \Sigma_t^{(2)} &= v_t \bar{\Sigma} \\ \Sigma_t^{(3)} &= (v_t - \rho) I_p + \rho \iota_p \iota'_p = (v_t - 1) I_p + \bar{\Sigma} \end{aligned}$$

which we label Cases 1, 2 and 3, respectively. In each case, we set $\rho = 0.4$ (implying a moderate degree of correlation between the components of ε_t) and $v_t = 2\, \mathbb{I}_{[0,\frac{1}{3})}\left(\frac{t}{T}\right) + 0.5\, \mathbb{I}_{[\frac{1}{3},1]}\left(\frac{t}{T}\right)$, such that the time-varying volatility factor v_t displays a negative shift a third of the way through the sample.[3] Notice that the volatility factor is normalised such that $\bar{v} := \int_0^1 v(s)\, ds = 1$,

[3]This particular structure for v_t is chosen to mimic the pattern of unconditional volatility seen in the interest rate data used in the empirical example in Section 6.

which implies that $\bar{\Sigma}^{(i)} = \bar{\Sigma}$, $i = 1, 2, 3$, where $\bar{\Sigma}^{(i)}$ is defined analogously to the definition of $\bar{\Sigma}$ in Lemma 2. For $\Sigma_t = \Sigma_t^{(1)}$, the errors ε_t are unconditionally homoskedastic. For $\Sigma_t = \Sigma_t^{(2)}$, a common time-varying factor affects the whole unconditional variance matrix. In this case, $\alpha_\perp' \Sigma(u) \bar{\Sigma}^{-1} \alpha = 0$, and the estimator $\hat{\beta}_2$ of the co-integrating vector is asymptotically mixed Gaussian; cf. Remark 3.5. Finally, for $\Sigma_t = \Sigma_t^{(3)}$ there are no common unconditional volatility factors, and so $\hat{\beta}_2$ is not mixed Gaussian. Regarding z_t, this is generated as the linear map

$$z_t = \Lambda e_t \tag{28}$$

where Λ is an invertible $p \times p$ matrix which is constant over time, with the p components of $e_t := (e_{1t}, ..., e_{pt})'$ independent across $i = 1, ..., p$. Without loss of generality, we set $\Lambda = I_p$. Each of the p components e_{it} is specified to follow a stationary GARCH$(1, 1)$ process driven by standard normal innovations; ie, $e_{it} = h_{it}^{1/2} \xi_{it}$, $i = 1, \ldots, p$, where ξ_{it} is i.i.d. $N(0, 1)$, independent across i, and $h_{it} = (1 - d_0 - d_1) + d_0 e_{i,t-1}^2 + d_1 h_{i,t-1}$, $t = 0, \ldots, T$. Results are reported for $(d_0, d_1) \in \{(0.0, 0.0), (0.05, 0.94)\}$; notice that: (a) for the former the errors are i.i.d. Gaussian; (b) for the latter d_0 and d_1 are chosen such that $d_0 + d_1$ is close to unity as tends to be observed with financial data, yet still satisfies an eight order moment assumption, see He and Teräsvirta (1999). In both cases, $\tau_{ij} = 0$ for all $i \neq j$; cf. Assumption 2 (b).[4]

We report results for tests of the following hypotheses

$$H_{0\beta} : \beta_2 = 0, \qquad H_{0\alpha} : \alpha_2 = 0,$$

where $\beta_2 = (0_{p-1} : I_{p-1})\beta$ and $\alpha_2 = (0_{p-1} : I_{p-1})\alpha$, together with tests of the joint hypothesis $H_{0\alpha\beta} = H_{0\beta} \cap H_{0\alpha}$. The general algorithms to obtain restricted estimates discussed in Section 3.1 are not needed here, as the the PLR test statistics for these specific hypotheses admit a closed-form expression, see Johansen (1996).

In each case we report empirical rejection frequencies [ERFs] under the null hypothesis for the tests which reject for large values of the PLR and Wald statistics when compared to: (i) asymptotic (χ^2) critical values; (ii) wild bootstrap critical values; (iii) i.i.d. bootstrap critical values, where the bootstrap errors in step (i) of Algorithm 1 are instead obtained by i.i.d. sampling from the re-centered $\{\varepsilon_t\}_{t=1}^T$. All tests are run at the nominal 5% significance level.

Insert Tables 1–3 about here

Results are reported in Tables 1, 2 and 3 for $p = 2$, $p = 3$ and $p = 4$, respectively. Entries in italics correspond to tests which are not asymptotically valid; cf. Sections 3 and 4. The following observations can be drawn from these results:

[4] We also ran additional Monte Carlo simulations for analogous conditionally heteroskedastic DGPs but which do not satisfy the restriction that $\tau_{ij} = 0$ for all $i \neq j$. We found almost no differences from the results reported here. This is perhaps not too surprising in the light of the observations made in point 5 of the summary of the simulations results below.

1. The results highlight the tendency for significant oversize in the standard asymptotic PLR and Wald tests, even in cases where these are asymptotically valid. For example, in Case 1 (i.i.d.) for $p = 3$ and $T = 100$, the ERF of the asymptotic tests range from 13.4% to 49.0%. Even for $T = 400$, the asymptotic Wald test on β shows ERFs in excess of 10%. For cases where the asymptotic tests are valid, both the i.i.d. and wild bootstrap versions of these perform well, with ERFs significantly closer to the 5% nominal level than is seen for the asymptotic tests.

2. In general, where the tests are known to be asymptotically valid, the observed size distortions are seen to decrease with the sample size T, as would be expected.

3. Other things being equal, the observed size distortions (of both the asymptotically valid and invalid tests) display a clear dependence on the dimension of the system, p. In particular, these tend to increase significantly with p. For example, the asymptotic PLR test for $H_{0\beta}$ in Case 3 with no GARCH, for $T = 100$ displays an ERF of 20.3% for $p = 2$, rising to 41.5% for $p = 3$ and 61.6% for $p = 4$. In this case, the ERFs of the corresponding wild bootstrap PLR test are 5.4% for $p = 2$, 6.7% for $p = 3$ and 7.4% for $p = 4$. The advantage of the wild bootstrap over the i.i.d. bootstrap in terms of finite sample size control also becomes increasingly evident the higher the dimension of the system, consistent with the observation that the distortions seen in the corresponding asymptotic tests become more pronounced as p increases, other things equal.

4. Of the bootstrap tests considered, the wild bootstrap PLR test appears, on average, to perform best, particularly so for the joint test of $H_{0\alpha\beta}$. For example, for $p = 4$ in Case 2 with no GARCH and $T = 100$ ($T = 400$), the wild bootstrap PLR test has an ERF of 4.8% (5.1%) while the wild bootstrap Wald test has an ERF of 12.9% (6.9%). In this example, the invalidity of the i.i.d. bootstrap PLR test is clearly demonstrated, while the i.i.d. bootstrap Wald test has an ERF of 14.9% for $T = 100$ reducing only to 8.0% for $T = 400$. It is also noteworthy that even for tests of $H_{0\alpha}$ (where the Wald statistic has an asymptotically pivotal limiting distribution under all of the DGPs considered, while the limiting null distribution of the PLR statistic is pivotal only under Case 1 when no GARCH is present) the wild bootstrap PLR test displays superior finite sample control to either the i.i.d. or wild bootstrap Wald tests in the vast majority of the results reported in Tables 1–3. Although an asymptotic refinement might be conjectured for the i.i.d. Wald test here, this does not appear to translate into better finite sample size control than the wild bootstrap PLR test.

5. A comparison of the results for the i.i.d. and the GARCH(1,1) cases suggests little apparent differences between the two. This is to be expected for tests of $H_{0\beta}$, since these tests are asymptotically valid in both cases. Only very small differences are seen in those tests involving $H_{0\alpha}$ that are asymptotically invalid under GARCH dynamics, suggesting

that, at least for the GARCH(1,1) model considered here, the impact on finite sample behaviour is rather limited.

The foregoing results all pertain to the behaviour of the tests under the null hypothesis. Unreported simulations under the alternative hypothesis show that the size-corrected power of wild bootstrap tests is only slightly smaller than that of the corresponding asymptotic tests; a similar result was found by Cavaliere, Nielsen and Rahbek (2015).

6 Empirical Application

In this section we apply the methods developed in the previous sections to an empirical model of the term structure of interest rates in the US. We use monthly observations over the period 1970:1–2009:12 on government zero yields $y_t(\tau)$ for maturities $\tau = 3, 12, 36, 60, 120$ (measured in months), and hence we consider a VAR model for $X_t = (y_t(3), y_t(12), y_t(36), y_t(60), y_t(120))'$. The zero yields have been constructed from the CRSP un-smoothed Fama and Bliss (1987) forward rates, see Diebold and Li (2006). The maturities have been chosen such that the dimension of the VAR model stays manageable ($p = 5$), and yet a reasonable coverage of the short, middle and long end of the term structure is obtained.

The econometric analysis of term structure data in recent years has been dominated by factor models, in particular the dynamic Nelson-Siegel model of Diebold and Li (2006). In this model, the time-series behaviour of $y_t(\tau)$ is described by the sum of level, slope and curvature factors (f_{1t} through f_{3t}), each multiplied by their factor exposures:

$$y_t(\tau) = f_{1t} + \frac{1 - e^{-\lambda\tau}}{\lambda\tau} f_{2t} + \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) f_{3t}, \tag{29}$$

where the shape parameter λ could be time-varying but in practice is often taken as constant (Diebold and Li (2006) set $\lambda = 0.0609$). An idiosyncratic error term is implicitly needed to fit the model to the data.

Depending on the unit root properties of f_{1t}, f_{2t} and f_{3t}, the model has clear co-integration implications. If all three factors have a unit root (and are not co-integrated), then a VAR model of dimension $p = 5$ should contain 3 common trends and hence $r = p - 3 = 2$ co-integrating relations. If the level and slope factors have a unit root, but the curvature factor is stable, then $r = 3$; this is found by Diebold and Li (2006). Finally, if only the level factor has a unit root, then $r = 4$ and $\beta'_\perp = (1, \ldots, 1)$, so that $\beta' X_t$ consists of spreads $y_t(\tau) - y_t(3)$ for $\tau > 3$. This hypothesis also arises as the so-called (weak-form) expectations hypothesis of interest rates; see Campbell and Shiller (1987).

We estimate a VAR(2) model with a constant term for X_t, using observations on the first two months of 1970 as starting values; hence, the estimation sample is 1970:3–2009:12, with $T = 478$. The lag order $k = 2$ is selected by the Hannan-Quinn information criterion; the Schwarz (Bayesian) criterion selects $k = 1$, but a first-order VAR model displays some rather

large first-order residual autocorrelation coefficients. The VAR(2) has much smaller, though apparently significant residual autocorrelation. However, this significance may be partly driven by the time-varying volatility; see Godfrey and Tremayne (2005). The Lagrange-multiplier F-tests for first- and second-order residual (vector) autocorrelation in the VAR(2) model have asymptotic p-values of 0.004 and 0.005; however, the corresponding wild bootstrap p-values are 0.32 and 0.42, respectively. This already illustrates the large impact that non-stationary volatility may have on the validity of standard asymptotic inference methods, for these data.

To visualise the possible presence of non-stationary volatility, we plot time series of VAR(2) residuals, as well as the corresponding variance profiles, in Figure 1. The variance profiles, see Caveliere and Taylor (2007), are plots of $\hat{V}_{T,ii}(u)/\hat{V}_{T,ii}(1)$ against $u \in [0,1]$, where $\hat{V}_{T,ii}(u) := T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \hat{\varepsilon}_{it}^2$. Deviations of this function from the diagonal (45 degree) line indicate the presence of persistent changes in volatility. We focus on the first, third and fifth residual, corresponding to the 3-month, 3-year and 10-year yields. In all series we may clearly distinguish a period of relatively high volatility around the late 1970s and early 1980s, and a period of low volatility since the mid-1980s (associated with the Great Moderation). This pattern is most pronounced in the short rate residuals, and dampened in the longer-maturity residuals. In addition to unconditional heteroskedasticity, we tested and found evidence for the presence of conditional heteroskedasticity, after correcting for shifts in the unconditional variance.

Insert Figure 1 and Table 4 about here

Table 4 displays the trace test statistics for co-integration rank (with restricted constant), together with asymptotic, standard bootstrap and wild bootstrap p-values. The asymptotic p-values are obtained using the procedure given in MacKinnon, Haug and Michelis (1999); the standard (sieve-type) bootstrap follows Cavaliere, Rahbek and Taylor (2012), whereas the wild bootstrap procedure is implemented as in Cavaliere, Rahbek and Taylor (2014); in all cases we take $B = 999$ bootstrap replications. The standard bootstrap p-values are included to assess to what extent the difference between the asymptotic and wild bootstrap results are due to the correction for non-stationary volatility, or due to finite sample problems only.

We observe that there is strong evidence against $r = 0$ or $r = 1$, regardless of the p-value method used. The asymptotic and standard bootstrap p-values also lead us to reject $r = 2$ (at the 5% level), but the wild bootstrap leads to weaker evidence for $r \geq 3$. Rejection of $r = 3$ against $r \geq 4$ is not supported by any of the methods at conventional significance levels, although the asymptotic p-value would lead us to consider this possibility more seriously than the wild bootstrap p-value. We conclude that a co-integration rank of $r = 3$ is supported by the data (although only at the 10% significance level for the wild bootstrap procedure). In the remainder of this section, we analyse some hypotheses on the co-integration parameters β and adjustment coefficients α. Because this exercise is intended as an illustration of the procedures proposed in this paper, we do so both in the context of the model with $r = 4$, where the weak-form expectations hypothesis of interest rates of Campbell and Shiller (1987) holds, and

for the model with $r = 3$, suggested by the rank tests in Table 1.

The unrestricted (but normalised) estimates of $\beta^{\#}$ and α in the model with $r = 4$, together with the PML standard errors, are as follows:

$$\hat{\beta}^{\#} = \begin{pmatrix} \underset{(0.03)}{-1.02} & \underset{(0.07)}{-0.99} & \underset{(0.10)}{-0.95} & \underset{(0.12)}{-0.87} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \underset{(0.16)}{-0.27} & \underset{(0.43)}{-0.89} & \underset{(0.56)}{-0.42} & \underset{(0.67)}{-2.11} \end{pmatrix}, \quad \hat{\alpha} = \begin{pmatrix} \underset{(0.24)}{0.11} & \underset{(0.49)}{-0.15} & \underset{(0.48)}{0.22} & \underset{(0.14)}{-0.07} \\ \underset{(0.16)}{-0.48} & \underset{(0.34)}{0.40} & \underset{(0.37)}{-0.13} & \underset{(0.12)}{-0.02} \\ \underset{(0.12)}{-0.30} & \underset{(0.26)}{-0.02} & \underset{(0.28)}{0.11} & \underset{(0.09)}{-0.02} \\ \underset{(0.12)}{-0.34} & \underset{(0.24)}{0.19} & \underset{(0.24)}{-0.10} & \underset{(0.08)}{0.02} \\ \underset{(0.11)}{-0.26} & \underset{(0.22)}{-0.06} & \underset{(0.21)}{0.20} & \underset{(0.07)}{-0.09} \end{pmatrix}.$$

Note that Theorem 2 implies that the PML standard errors of $\hat{\alpha}_{ij}$ (and hence Wald tests based on these standard errors) are asymptotically valid. However, the simulation results in Section 5 show that the wild bootstrap version of the PLR test has much better finite sample properties, which is why we will use this test for hypotheses on α.

We observe that $\hat{\beta}^{\#}$ is fairly close to the theoretical value implied by the expectations hypothesis, or equivalently by the hypothesis of a dynamic Nelson-Siegel model with stable slope and curvature factors. In the notation of Section 2.2, this hypothesis and the normalisation used above correspond to

$$\bar{c}^{\#} = c^{\#} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \bar{c}^{\#}_{\perp} = c^{\#}_{\perp} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad R_\beta = I_4 \otimes \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad q_\beta = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}.$$

The PLR statistic for the expectations hypothesis (with restricted estimates computed using the switching algorithm discussed in Section 3.1) is equal to 8.01, with an asymptotic p-value of 0.09, but a wild bootstrap p-value of 0.34. Therefore, this hypothesis cannot be rejected (conditional on a co-integration rank of $r = 4$), although the asymptotic p-value would cast some doubt on it.

A possible hypothesis of interest on α is that its first row is zero. This corresponds to the hypothesis of weak exogeneity of the short rate $X_{1t} = y_t(3)$ for the co-integration parameters. Equivalently, it corresponds to the hypothesis that the single common trend (the non-stationary level factor) is fully driven by the disturbance from the first equation of the VAR model ($\alpha'_\perp \varepsilon_t = \varepsilon_{1t}$). In the notation of Section 2.2, this corresponds to $R_\alpha = \begin{bmatrix} I_4 & 0 \end{bmatrix}$ and $q_\alpha = 0$. The PLR test for this hypothesis has an asymptotic p-value of 0.25 and a bootstrap p-value of 0.68, so this hypothesis cannot be rejected.

The PLR test for the joint hypothesis (stationary spreads and weak exogeneity of $y_t(3)$) has an asymptotic p-value of 0.08 and a wild bootstrap p-value of 0.54. This gives another example

of a hypothesis where asymptotic and bootstrap inference might yield different conclusions, at least if a significance level of 10% were adopted.

We now consider hypotheses in the model with $r = 3$ co-integrating vectors. Here the spreads $s_t(\tau) := y_t(\tau) - y_t(3)$ are not (all) stable, but a hypothesis of interest is that the co-integrating relations can be expressed in terms of these, which requires $\beta'(1, \ldots, 1) = 0$. The PLR statistic for this hypothesis has an asymptotic p-value of 0.24, and a wild bootstrap p-value of 0.46, so this hypothesis is not rejected. The PML estimates under this restriction are

$$\tilde{\beta} = \begin{pmatrix} -1.007 & -0.71 & -0.44 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \underset{(0.07)}{0.007} & \underset{(0.12)}{-0.29} & \underset{(0.11)}{-0.56} \end{pmatrix}, \quad \tilde{\alpha} = \begin{pmatrix} \underset{(0.24)}{0.13} & \underset{(0.50)}{-0.21} & \underset{(0.49)}{0.28} \\ \underset{(0.16)}{-0.48} & \underset{(0.34)}{0.42} & \underset{(0.38)}{-0.11} \\ \underset{(0.12)}{-0.30} & \underset{(0.26)}{0.06} & \underset{(0.29)}{0.09} \\ \underset{(0.12)}{-0.34} & \underset{(0.23)}{0.22} & \underset{(0.24)}{-0.10} \\ \underset{(0.11)}{-0.25} & \underset{(0.21)}{-0.09} & \underset{(0.21)}{0.21} \end{pmatrix}.$$

For example, the second co-integrating vector implies that the three-year spread $s_t(36)$ minus 0.29 times the 10-year spread $s_t(120)$ is stable. The particular normalisation of β chosen in this representation is inspired by the fact that the one-year spread $s_t(12)$ is close to being stable (as the estimate of 0.007 indicates). This would mean that the coefficient a in a stable relation of the form of $s_t(\tau) - a \cdot s_t(12)$ is not well defined, as it is normalised on the wrong variable.

Finally, it may be of interest to test whether the co-integrating relation found here is in agreement with a dynamic Nelson-Siegel model of the form (29) with $\lambda = 0.0609$. This requires that the matrix β is orthogonal to both the vector of level factor exposures (a vector of ones), and the vector of slope factor exposures. It is easily seen that this corresponds to the hypothesis

$$H_0 : \beta' X_t = \begin{pmatrix} s_t(12) - 0.26 \cdot s_t(120) \\ s_t(36) - 0.65 \cdot s_t(120) \\ s_t(60) - 0.83 \cdot s_t(120) \end{pmatrix}.$$

The PLR test for this hypothesis (in the unrestricted model with $r = 3$) has a wild bootstrap p-value of 0.01, which leads us to reject this hypothesis. A simpler model, in which the slope factor exposure increases linearly with the horizon τ, corresponds to a hypothesis that cannot be rejected, with a wild bootstrap p-value of 0.24.

7 Conclusions

We have investigated the impact of time-varying volatility in co-integrated VAR models on the standard methods of estimation (PML estimators) and hypothesis testing (PLR tests) on the coefficients of the co-integrating relations (β) and the adjustment coefficients (α) outlined in Johansen (1996). In particular, for a very general model which allows both conditional and unconditional heteroskedasticity of a quite general form, we have shown that the PML estimates

remain consistent but that their limiting distributions depend in general on nuisance parameters arising from the underlying volatility process. The latter is also shown to be the case for the limiting null distributions of the PLR statistics, with the implication that the resulting PLR tests can have true size significantly in excess of the nominal significance level when based on conventional χ^2 critical values. Solutions to this problem based on Wald tests and on the use of the wild bootstrap were subsequently discussed. The conditions under which these methods yield asymptotically valid inference were provided, with the wild bootstrap implementation of the Wald test shown to require only a relatively mild strengthening of the necessary moment conditions for this to be obtained under the class of time-varying volatility processes considered. Monte Carlo evidence was reported for a variety of conditionally and unconditionally heteroskedastic models which suggested that the proposed bootstrap co-integration tests perform well in finite samples largely avoiding the oversize problems that can occur with the standard tests, the latter being worse, other things equal, the higher the dimension of the system. Finally, an application to the term structure of interest rates in the US was used to highlight the differences that can occur in practice between standard and bootstrap inferences regarding hypotheses on the co-integrating vectors and adjustment coefficients.

We end with a suggestion for further research. Our aim has been to deliver tests on α and β based on standard PML estimates (i.e. those which obtain for Gaussian i.i.d. errors) but which are (asymptotically) robust to a wide class of heteroskedasticity in the errors. Where heteroskedasticity is present, the PML estimates are not efficient and consequently tests derived using the true ML estimates would be anticipated to be more efficient. As with the approach we have taken here, one would, however, probably not want to use an estimation method which assumed a particular parametric model for the volatility process, since this would be expected to perform poorly (quite possibly worse than PML-based tests) if an inappropriate volatility model was specified. However, under rather stronger conditions on the volatility process than are considered here, it may be possible to develop tests which are asymptotically efficient using an adaptive approach, based on non-parametric estimation of the sequence of unconditional variance matrices. This has been developed for the univariate case by Boswijk (2005). Because the asymptotic distributions of the resulting statistics are not expected to be free of nuisance parameters, this should be combined with the wild bootstrap as in the present paper.

A Appendix

A.1 Preliminary Results

The following Lemma will be used throughout.

Lemma A.1. *Let ε_t be defined as in Assumption 2. Then, as $T \to \infty$,*

(i) $T^{-1}\sum_{t=1}^{T} E\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right) = T^{-1}\sum_{t=1}^{T}\sigma_t h_t \sigma_t' \xrightarrow{p} \bar{\Sigma} := \int_0^1 \Sigma\left(s\right) ds;$

(ii) $T^{-1}\sum_{t=1}^{T}\left(E\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right)\otimes\varepsilon_{t-i}\right)=T^{-1}\sum_{t=1}^{T}(\sigma_t h_t\sigma_t'\otimes\sigma_{t-i}z_{t-i})\xrightarrow{p}0$ *for* $i\geq 1$*;*

(iii) $T^{-1}\sum_{t=1}^{T}(E\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right)\otimes\varepsilon_{t-i}\varepsilon_{t-j}')=T^{-1}\sum_{t=1}^{T}(\sigma_t h_t\sigma_t'\otimes\sigma_{t-i}z_{t-i}z_{t-j}'\sigma_{t-j}')$
$\xrightarrow{p}\int_0^1\left[\sigma\left(s\right)\otimes\sigma\left(s\right)\right]\tau_{ij}\left[\sigma\left(s\right)\otimes\sigma\left(s\right)\right]'ds$*, for* $i,j\geq 1$.

Proof. For part (i), notice that $\frac{1}{T}\sum_{t=1}^{T}\sigma_t h_t\sigma_t'=\frac{1}{T}\sum_{t=1}^{T}\sigma_t\left(h_t-I_p\right)\sigma_t'+\frac{1}{T}\sum_{t=1}^{T}\sigma_t\sigma_t'$, where $T^{-1}\sum_{t=1}^{T}\sigma_t\sigma_t'\to\bar{\Sigma}$. For the term $T^{-1}\sum_{t=1}^{T}\sigma_t\left(h_t-I_p\right)\sigma_t'$ we can proceed as in Theorem A.1 of Cavaliere and Taylor (2009), using $\sup_t\|h_t\|<\infty$ (which is implied by the moment assumption on ε_t) and that $\|\sigma\left(u\right)\sigma\left(u\right)'\|$ is a càdlàg process in $D[0,1]$. Notice that Cavaliere and Taylor (2009) require the stochastic term (h_t-I_p) to be a mixingale, but their proof of Theorem A.1 goes through by replacing this assumption with Assumption 2(b) i. of Section 2. Parts (ii) and (iii) follow similarly. □

A.2 Results for The Asymptotic Test Statistics

Proof of Lemma 1. Assumption 2(b) implies that z_t satisfies a functional central limit theorem (Brown, 1971, Theorem 3), since the higher moment assumption implies a conditional Lindeberg condition. Therefore

$$W_T(\cdot):=\frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor T\cdot\rfloor}z_t\xrightarrow{w}W(\cdot),\tag{A.1}$$

where W is a p-dimensional standard Brownian motion. Next, define $\sigma_T(u):=\sigma_{\lfloor Tu\rfloor+1}$, $0\leq u<1$, and $\sigma_T(1):=\sigma_T$, such that $M_T(\cdot):=\int_0^{\cdot}\sigma_T(s)dW_T(s)=\frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\sigma_t z_t=\frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\sigma(\frac{t}{T})z_t$. We also define the following approximant

$$\tilde{M}_T(\cdot):=\int_0^{\cdot}\sigma(s)dW_T(s)=\frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor T\cdot\rfloor}\sigma^{-}(\tfrac{t}{T})z_t,$$

where $\sigma^{-}\left(0\right)=\sigma\left(0\right)$ and, for any $u\in(0,1]$, $\sigma^{-}\left(u\right):=\lim_{s\uparrow u}\sigma\left(u\right)$ (notice that $\sigma^{-}\left(u\right)$ exists as $\sigma\left(u\right)$ is càdlàg). Because $\sup_T T^{-1}\sum_{t=1}^{T}E(\|z_t\|^2)=p<\infty$, it follows from Theorem 2.1 of Hansen (1992b) that

$$(W_T,\tilde{M}_T)\xrightarrow{w}(W,M).\tag{A.2}$$

Therefore, in order to derive the weak limit of M_T it suffices to prove that $\tilde{M}_T$ and M_T are sufficiently close in the sup norm, in the sense that

$$\sup_{u\in[0,1]}\left\|\tilde{M}_T(u)-M_T(u)\right\|=o_p\left(1\right),\tag{A.3}$$

see Billingsley (1968, Theorem 4.1). To see (A.3), notice that

$$\begin{aligned}\sup_{u\in[0,1]}\left\|\tilde{M}_T(u)-M_T(u)\right\| &= \sup_u\left\|T^{-1/2}\textstyle\sum_{t=1}^{[Tu]}(\sigma^{-}(\frac{t}{T})-\sigma(\frac{t}{T}))z_t\right\|\\ &\leq T^{-1/2}\textstyle\sum_{t=1}^{T}\left\|\sigma(\frac{t}{T})-\sigma^{-}(\frac{t}{T})\right\|\|z_t\|\\ &\leq \max_{t=1,\ldots,T}T^{-1/2}\|z_t\|\times\textstyle\sum_{t=1}^{T}\left\|\sigma(\frac{t}{T})-\sigma^{-}(\frac{t}{T})\right\|=:J_T^{(1)}\times J_T^{(2)}\end{aligned}$$

where $J_T^{(1)}$ and $J_T^{(2)}$ are implicitly defined. In order to analyse $J_T^{(2)}$, let $f_{ij,t} := \left|\sigma_{ij}\left(\frac{t}{T}\right) - \sigma_{ij}^{-}\left(\frac{t}{T}\right)\right|$ with $f_{ij,t} \leq \bar{f}_{ij} := 2\sup_u |\sigma_{ij}(u)|$, where $\bar{f}_{ij} < \infty$ since $\sigma(\cdot)$ – and therefore $\sigma_{ij}(\cdot)$ – is càdlàg. Then, we can write

$$\sum_{t=1}^{T} f_{ij,t} = \sum_{t=1}^{T} f_{ij,t}\mathbb{I}(f_{ij,t} > \tfrac{1}{T}) + \sum_{t=1}^{T} f_{ij,t}\mathbb{I}(f_{ij,t} \leq \tfrac{1}{T}) \leq \bar{f}_{ij}\sum_{t=1}^{T}\mathbb{I}(f_{ij,t} > \tfrac{1}{T}) + 1.$$

Since $\sigma_{ij}(\cdot)$ is càdlàg, the number $K_{ij,T}$ of jumps such that $\left|\sigma_{ij}(u) - \sigma_{ij}^{-}(u)\right| > 1/T$ is finite for every T (see e.g. Davidson (1994), p. 458), and therefore $\sum_{t=1}^{T} f_{ij,t} \leq \bar{f}_{ij}K_{ij,T} + 1 = O(1)$, for any T. Since σ_t has finite dimension $p \times p$, it also holds that $J_T^{(2)} \leq p\max_{i,j=1,\ldots,p} f_{ij,t} = O(1)$. As far as $J_T^{(1)}$ is concerned, we have that $J_T^{(1)} \to_p 0$ since for any $\epsilon > 0$, by Bonferroni's and Markov's inequalities

$$\begin{aligned} P\left(J_T^{(2)} > \epsilon\right) &= P\left(\max_{t=1,\ldots,T} T^{-1/2}\|z_t\| > \epsilon\right) = P\left(\max_{t=1,\ldots,T}\|z_t\| > \epsilon T^{1/2}\right) \\ &\leq \sum_{t=1}^{T} P\left(\|z_t\| > \epsilon T^{1/2}\right) \leq \sum_{t=1}^{T}\frac{E(\|z_t\|^4)}{\epsilon^4 T^2} \leq \frac{\sup_t E(\|z_t\|^4)}{\epsilon^4 T} \to 0 \end{aligned}$$

since $\sup_t E(\|z_t\|^4) < \infty$ by Assumption 2(b)iv. Since $J_T^{(1)} = o_p(1)$ and $J_T^{(2)} = O(1)$, the convergence in (A.3) holds and therefore

$$(W_T(\cdot), M_T(\cdot)) \xrightarrow{w} (W(\cdot), M(\cdot)), \tag{A.4}$$

as required.

Next, recalling that $\int_0^1 M_T(s)dM_T(s)' = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\varepsilon_i\right)\varepsilon_t'$, and using the result that

$$\sup_T \frac{1}{T}\sum_{t=1}^{T} E(\|\varepsilon_t\|^2) = \sup_T \frac{1}{T}\sum_{t=1}^{T}\left\|\sigma(\tfrac{t}{T})\right\|^2 < \infty,$$

because σ is càdlàg and hence bounded, applying Theorem 2.1 of Hansen (1992b) once more, we find

$$\operatorname{vec}\int_0^1 M_T(s)dM_T(s)' \xrightarrow{w} \operatorname{vec}\int_0^1 M(s)dM(s)',$$

jointly with (A.4). □

Proof of Lemma 2. Define $\mathbb{X}_{\beta t}^{\#} = (Z_{\beta t+1}', Z_{2t+1}')'$, with $Z_{\beta t} = \beta^{\#\prime}Z_{1t}$; therefore, $\mathbb{X}_{\beta t}^{\#}$ equals $\mathbb{X}_{\beta t}$ except that the mean ρ_1' has been subtracted from $\beta' X_{t-1}$. This means that $\mathbb{X}_{\beta t}^{\#} = \Phi\mathbb{X}_{\beta,t-1}^{\#} + F\varepsilon_t$, see Section 2.1, a stable vector process (because $\rho(\Phi) < 1$) with zero mean, except for the effect of fixed starting values. Because $\left\|\Phi^i\right\| \leq c\lambda^i$ for some positive constant c and $|\lambda| < 1$, we therefore may write $\mathbb{X}_{\beta t}^{\#} = \Phi^t\mathbb{X}_{\beta 0}^{\#} + \sum_{i=0}^{t-1}\Phi^i F\varepsilon_{t-i} = \sum_{i=0}^{t-1}\Phi^i F\varepsilon_{t-i} + o_p(1)$, where the $o_p(1)$ term will be neglected in the following.

First, as in Lemma A.1 of Cavaliere, Rahbek and Taylor (2010b) we have that

$$\begin{pmatrix} \beta^{\#\prime}M_{11}\beta^{\#} & \beta^{\#\prime}M_{12} \\ M_{21}\beta^{\#} & M_{22} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{X}_{\beta,t-1}^{\#}\mathbb{X}_{\beta,t-1}^{\#\prime} \xrightarrow{p} \sum_{i=0}^{\infty}\Phi^i F\bar{\Sigma}F'\Phi^{i\prime} =: \begin{pmatrix} \bar{M}_{\beta\beta} & \bar{M}_{\beta 2} \\ \bar{M}_{2\beta} & \bar{M}_{22} \end{pmatrix}, \tag{A.5}$$

from which we find $\beta^{\#\prime}S_{11}\beta^{\#} = \beta^{\#\prime}M_{11}\beta^{\#} - \beta^{\#\prime}M_{12}M_{22}^{-1}M_{21}\beta^{\#} \xrightarrow{p} \bar{M}_{\beta\beta} - \bar{M}_{\beta 2}\bar{M}_{22}^{-1}\bar{M}_{2\beta} =: \bar{\Sigma}_{\beta\beta}$. The convergence in (A.5) is obtained as follows.

Consider

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\mathbb{X}_{\beta,t-1}^{\#}\mathbb{X}_{\beta,t-1}^{\#\prime} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j=1}^{t-1}\Phi^{i-1}F\varepsilon_{t-i}\varepsilon_{t-j}'F'\Phi^{j-1\prime} \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\Phi^{i-1}F\sigma_{t-i}h_{t-i}\sigma_{t-i}'F'\Phi^{i-1\prime} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\Phi^{i-1}F(\varepsilon_{t-i}\varepsilon_{t-i}' - \sigma_{t-i}h_{t-i}\sigma_{t-i}')F'\Phi^{i-1\prime} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j=1,j\neq i}^{t-1}\Phi^{i-1}F\varepsilon_{t-i}\varepsilon_{t-j}'F'\Phi^{j-1\prime} =: A_T + B_T + C_T,
\end{aligned}$$

where A_T, B_T and C_T are implicitly defined. Using Lemma A.1 (a), we find that

$$A_T = \sum_{i=1}^{T-1}\Phi^{i-1}F\left(\frac{1}{T}\sum_{t=1}^{T-i}\sigma_t h_t\sigma_t'\right)F'\Phi^{i-1\prime} \xrightarrow{p} \sum_{i=1}^{\infty}\Phi^i F\bar{\Sigma}F'\Phi^{i\prime}.$$

Next, $B_T = \sum_{i=1}^{T-1}\Phi^{i-1}F\left(\frac{1}{T}\sum_{t=1}^{T-i}\sigma_t(z_t z_t' - h_t)\sigma_t'\right)F'\Phi^{i-1\prime} = o_p(1)$, by the martingale law of large numbers, using the fact that z_t has finite fourth moment, σ_t is a bounded sequence, and $\left\|\Phi^i\right\| \leq c\lambda^i$. Using the fact that $E(\varepsilon_{t-i}\varepsilon_{t-j}'|\mathcal{F}_{t-k}) = 0$ with $k = \max(i,j)+1$ and that $\varepsilon_{t-i}\varepsilon_{t-j}'$ has bounded variance because σ_t and τ_{ij} are bounded, it follows that $C_T = o_p(1)$, which proves (A.5).

By the same approach, we have

$$\begin{pmatrix} \beta^{\#\prime}M_{1\varepsilon} \\ M_{2\varepsilon} \end{pmatrix} := \frac{1}{T}\sum_{t=1}^{T}\mathbb{X}_{\beta,t-1}^{\#}\varepsilon_t' \xrightarrow{p} 0, \tag{A.6}$$

so that $\beta^{\#\prime}S_{10} = \beta^{\#\prime}S_{1\varepsilon} + \beta^{\#\prime}S_{11}\beta^{\#}\alpha' = \beta^{\#\prime}M_{1\varepsilon} - \beta^{\#\prime}M_{12}M_{22}^{-1}M_{2\varepsilon} + \beta^{\#\prime}S_{11}\beta^{\#}\alpha' \xrightarrow{p} \bar{\Sigma}_{\beta\beta}\alpha' =: \bar{\Sigma}_{\beta 0}$, and, defining $S_{\varepsilon\varepsilon} := M_{\varepsilon\varepsilon} - M_{\varepsilon 2}M_{22}^{-1}M_{2\varepsilon}$ with $M_{\varepsilon\varepsilon} := T^{-1}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t'$,

$$\begin{aligned}
S_{00} &= \alpha\beta^{\#\prime}S_{10} + S_{\varepsilon 0} \\
&= \alpha\beta^{\#\prime}S_{10} + S_{\varepsilon\varepsilon} + S_{\varepsilon 1}\beta^{\#}\alpha' \\
&= \alpha\beta^{\#\prime}S_{10} + M_{\varepsilon\varepsilon} - M_{\varepsilon 2}M_{22}^{-1}M_{2\varepsilon} + M_{\varepsilon 1}\beta^{\#}\alpha' - M_{\varepsilon 2}M_{22}^{-1}M_{21}\beta^{\#}\alpha' \\
&\xrightarrow{p} \alpha\bar{\Sigma}_{\beta 0} + \bar{\Sigma} = \alpha\bar{\Sigma}_{\beta\beta}\alpha' + \bar{\Sigma} =: \bar{\Sigma}_{00}.
\end{aligned}$$

This proves (10).

To prove (11), we start with $\operatorname{vec}\mathbb{X}_{\beta,t-1}^{\#}\varepsilon_t' = \operatorname{vec}\Phi^{t-1}\mathbb{X}_{\beta 0}^{\#}\varepsilon_t' + \sum_{i=1}^{t-1}\operatorname{vec}\Phi^{i-1}F\varepsilon_{t-i}\varepsilon_t'$. For any fixed $n \in \mathbb{N}$, consider the decomposition

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}\operatorname{vec}\mathbb{X}_{\beta,t-1}^{\#}\varepsilon_t' = S_T^{(n)} + R_T^{(n)} + R_{0T}, \tag{A.7}$$

where (taking $z_t = 0$ and hence $\varepsilon_t = 0$ for $t \leq 0$)

$$S_T^{(n)} := \frac{1}{T^{1/2}} \sum_{t=1}^{T} \sum_{i=1}^{n} \operatorname{vec} \Phi^{i-1} F \varepsilon_{t-i} \varepsilon_t' = \sum_{i=1}^{n} (I_p \otimes \Phi^{i-1} F) \frac{1}{T^{1/2}} \sum_{t=1}^{T} (\sigma_t \otimes \sigma_{t-i})(z_t \otimes z_{t-i}),$$

$$R_T^{(n)} := \frac{1}{T^{1/2}} \sum_{t=1}^{T} \sum_{i=n+1}^{t-1} \operatorname{vec} \Phi^{i-1} F \varepsilon_{t-i} \varepsilon_t', \qquad R_{0T} := \frac{1}{T^{1/2}} \sum_{t=1}^{T} \operatorname{vec} \Phi^{t-1} \mathbb{X}_{\beta 0}^{\#} \varepsilon_t'.$$

We will focus first on the limit of $S_T^{(n)}$ as $T \to \infty$ for fixed n, and then let $n \to \infty$ to obtain the limiting distribution of the right-hand side of (A.7), where we need to show that $\|R_{0T}\| = o_p(1)$ and $\lim_{n\to\infty} \limsup_{T\to\infty} P(\|R_T^{(n)}\| > \epsilon) = 0$ for all $\epsilon > 0$, see Proposition 6.3.9 of Brockwell and Davis (1991).

Consider the process

$$U_t^{(n)} = \begin{pmatrix} (z_t \otimes z_{t-1}) \\ \vdots \\ (z_t \otimes z_{t-n}) \end{pmatrix}.$$

Assumption 2 (b) implies that $U_t^{(n)}$ is a vector martingale difference sequence, with conditional variance matrix satisfying $\frac{1}{T}\sum_{t=1}^{T} E(U_t^{(n)} U_t^{(n)\prime} | \mathcal{F}_{t-1}) \xrightarrow{p} \tau^{(n)} = E(U_t^{(n)} U_t^{(n)\prime})$, where $\tau^{(n)}$ is a matrix with blocks τ_{ij}, $i, j = 1, \ldots, n$. Furthermore, the finite $(4r)$th moment of z_t implies that $U_t^{(n)}$ has a finite $(2r)$th moment, $r > 1$, which in turn implies a Lindeberg condition, such that $U_t^{(n)}$ satisfies the invariance principle of Brown (1971), i.e.,

$$W_T^{(n)}(\cdot) = \begin{pmatrix} W_{1T}(\cdot) \\ \vdots \\ W_{nT}(\cdot) \end{pmatrix} := \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor T \cdot \rfloor} U_t^{(n)} \xrightarrow{w} W^{(n)}(\cdot) = \begin{pmatrix} W_1(\cdot) \\ \vdots \\ W_n(\cdot) \end{pmatrix},$$

where $W^{(n)}$ is a vector Brownian motion with variance matrix $\tau^{(n)}$, independent of W because $E(U_t^{(n)} z_t') = 0$ by Assumption 2 (b) ii.

Next, we may write $S_T^{(n)} = \sum_{i=1}^{n} (I_p \otimes \Phi^{i-1} F) \int_0^1 [\sigma_T(s) \otimes \sigma_T(s - i/T)] dW_{iT}(s)$, which by Theorem 2.1 of Hansen (1992b) converges weakly to

$$S^{(n)} := \sum_{i=1}^{n} (I_p \otimes \Phi^{i-1} F) \int_0^1 [\sigma(s) \otimes \sigma(s)] dW_i(s) \sim N(0, V^{(n)}),$$

with $V^{(n)} := \sum_{i=1}^{n} \sum_{j=1}^{n} (I_p \otimes \Phi^{i-1} F) \int_0^1 [\sigma(s) \otimes \sigma(s)] \tau_{ij} [\sigma(s) \otimes \sigma(s)]' ds (I_p \otimes \Phi^{j-1} F)'$. Because $\sigma(u)$ and τ_{ij} are bounded and $\rho(\Phi) < 1$, $\lim_{n\to\infty} V^{(n)} = V := V^{(\infty)}$, and as $n \to \infty$, $S^{(n)} \xrightarrow{w} N(0, V)$.

It is easily seen that the assumptions imply that R_{0T} has mean zero and variance of order T^{-1}, so that $R_{0T} = o_p(1)$. Similarly, for each fixed T and n, $R_T^{(n)}$ has mean zero and variance matrix $\sum_{i=n+1}^{T-1} \sum_{j=n+1}^{T-1} (I_p \otimes \Phi^{i-1}) \left[\frac{1}{T} \sum_{t=i+1}^{T} (\sigma_t \otimes F\sigma_{t-i}) \tau_{ij} (\sigma_t \otimes F\sigma_{t-j})' \right] (I_p \otimes \Phi^{j-1})'$, the

matrix norm of which is less than

$$\sum_{i=n+1}^{T-1}\sum_{j=n+1}^{T-1}\left\|I_p\otimes\Phi^{i-1}\right\|\frac{1}{T}\sum_{t=i+1}^{T}\left\|(\sigma_t\otimes F\sigma_{t-i})\tau_{ij}(\sigma_t\otimes F\sigma_{t-j})'\right\|\left\|I_p\otimes\Phi^{j-1}\right\|$$
$$\leq \sum_{i=n+1}^{\infty}\sum_{j=n+1}^{\infty}p\left\|\Phi\right\|^{i+j-2}\max_{n<i,j<t<T,}\left\|(\sigma_t\otimes F\sigma_{t-i})\tau_{ij}(\sigma_t\otimes F\sigma_{t-j})'\right\|,$$

and since the final factor is bounded because $\sigma(u)$ and τ_{ij} are bounded, this converges to 0 as $n\to\infty$. Therefore, using Chebyshev's inequality, $\lim_{n\to\infty}\limsup_{T\to\infty}P(\|R_T^{(n)}\|>\epsilon)=0$, for all $\epsilon>0$, and hence $T^{-1/2}\sum_{t=1}^{T}\mathrm{vec}\,\mathbb{X}^{\#}_{\beta,t-1}\varepsilon_t'\xrightarrow{w}N(0,V)$. From the definition of $\beta^{\#\prime}S_{1\varepsilon}$, this in turn implies $T^{1/2}\beta^{\#\prime}S_{1\varepsilon}\xrightarrow{w}N(0,\Omega)$, with $\Omega=\left[I:-\bar{M}_{\beta 2}\bar{M}_{22}^{-1}\right]V\left[I:-\bar{M}_{\beta 2}\bar{M}_{22}^{-1}\right]'$. This proves (11).

To prove (12) and (13), we note that from (3) with $\mu D_t=\alpha\rho_1'$ and hence $C\mu D_t=0$, we have

$$B_T'Z_{1t}=\begin{pmatrix}T^{-1/2}(\beta_\perp'C\sum_{i=1}^{t-1}\varepsilon_i+\beta_\perp'S_{t-1}+\beta_\perp C_0)\\ 1\end{pmatrix}.$$

Letting $G_T(\cdot)=B_T'Z_{1,\lfloor T\cdot\rfloor+1}$, we have $G_T(\cdot)\xrightarrow{w}G(\cdot)$, using the fact that $\beta_\perp'S_t$ is linear in $\mathbb{X}_{\beta t}$, and $T^{-1/2}\mathbb{X}_{\beta\lfloor T\cdot\rfloor}=o_p(1)$. Therefore, using the continuous mapping theorem [CMT] and Theorem 2.1 of Hansen (1992b), we have that

$$\begin{aligned}\left(B_T'M_{11}B_T,B_TM_{1\varepsilon}\right) &= \left(\int_0^1 G_T(s)G_T(s)'ds,\int_0^1 G_T(s)dM_T(s)'\right)\\ &\xrightarrow{w} \left(\int_0^1 G(s)G(s)'ds,\int_0^1 G(s)dM(s)'\right).\end{aligned}\tag{A.8}$$

The fact that Z_{2t} is a linear process in ε_t with exponentially decaying weights implies that

$$B_T'M_{12}=O_p(T^{-1/2}),\qquad B_T'M_{11}\beta^{\#}=O_p(T^{-1/2}),\tag{A.9}$$

which together with (A.5) and (A.6), implies that (A.8)–(A.9) also holds with M_{11} and $M_{1\varepsilon}$ replaced by S_{11} and $S_{1\varepsilon}$, respectively. □

Proof of Theorem 1. Consistency of $\hat{\beta}^{\#}$ follows from the limiting behaviour of the concentrated pseudo-log-likelihood (9), together with the results of Lemma 2. This is analysed in detail in Lemmas 13.1 and Theorem 13.3 of Johansen (1996), which can be applied directly to the present case. It also follows that $\hat{\beta}_2-\beta_2=O_p(T^{-1})$ and $\hat{\rho}_1-\rho=O_p(T^{-1/2})$, and this in turn implies that $\hat{\alpha}=S_{01}\hat{\beta}^{\#}(\hat{\beta}^{\#\prime}S_{11}\hat{\beta}^{\#})^{-1}=S_{01}\beta^{\#}(\beta^{\#\prime}S_{11}\beta^{\#})^{-1}+o_p(1)\xrightarrow{p}\bar{\Sigma}_{0\beta}\bar{\Sigma}_{\beta\beta}^{-1}=\alpha$, and similarly $\hat{\Sigma}=S_{00}-S_{01}\hat{\beta}^{\#}(\hat{\beta}^{\#\prime}S_{11}\hat{\beta}^{\#})^{-1}\hat{\beta}^{\#\prime}S_{10}=S_{\varepsilon\varepsilon}-S_{\varepsilon 1}\beta^{\#}(\beta^{\#\prime}S_{11}\beta^{\#})^{-1}\beta^{\#\prime}S_{1\varepsilon}+o_p(1)\xrightarrow{p}\bar{\Sigma}$. The PML estimator of Ψ is readily obtained from the un-concentrated pseudo-log-likelihood, and this leads to $\hat{\Psi}=M_{22}^{-1}(M_{20}-M_{21}\hat{\beta}^{\#}\hat{\alpha}')=\Psi+M_{22}^{-1}\left[M_{21}(\beta^{\#}\alpha'-\hat{\beta}^{\#}\hat{\alpha}')+M_{2\varepsilon}\right]\xrightarrow{p}\Psi$. This proves part (a).

For the asymptotic distributions of the estimators, we use the fact that consistency at the appropriate rate, and a sufficient degree of differentiability, allows the following result based on

a second-degree Taylor approximation of ℓ:

$$D_T^{-1}(\hat{\theta}-\theta) = \left(-D_T'\frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta'}D_T\right)^{-1} D_T'\frac{\partial\ell(\theta)}{\partial\theta} + o_p(1), \tag{A.10}$$

where θ is the parameter vector and D_T is a matrix chosen such that the matrix in parentheses converges to a finite non-singular limit; see the proof of Theorem 3 for details of the expansion in (A.10) applied to the more general case of restricted estimators. Choosing

$$\theta = \begin{pmatrix} \operatorname{vec}\beta_2^{\#} \\ \operatorname{vec}\alpha' \end{pmatrix}, \qquad D_T = \begin{pmatrix} I_r \otimes D_{\beta T} & 0 \\ 0 & T^{-1/2}I_{pr} \end{pmatrix}, \qquad D_{\beta T} = \begin{pmatrix} T^{-1}I_{(p-r)} & 0 \\ 0 & T^{-1/2} \end{pmatrix},$$

we have

$$\begin{aligned} D_T'\frac{\partial\ell(\theta)}{\partial\theta} &= \begin{pmatrix} T\operatorname{vec}(D_{\beta T}'c_\perp^{\#\prime}S_{1\varepsilon}\bar{\Sigma}^{-1}\alpha) \\ T^{1/2}\operatorname{vec}(\beta^{\#\prime}S_{1\varepsilon}\bar{\Sigma}^{-1}) \end{pmatrix} \xrightarrow{w} \begin{pmatrix} \operatorname{vec}\int_0^1 AG(s)dM(s)'\bar{\Sigma}^{-1}\alpha \\ N(0,[\bar{\Sigma}^{-1}\otimes I_r]\Omega[\bar{\Sigma}^{-1}\otimes I_p] \end{pmatrix}, \qquad (A.11) \\ -D_T'\frac{\partial\ell^2(\theta)}{\partial\theta\partial\theta}D_T &= \begin{pmatrix} (\alpha'\bar{\Sigma}^{-1}\alpha)\otimes(TD_{\beta T}'c_\perp^{\#\prime}S_{11}c_\perp^{\#}D_{\beta T}) & (\alpha'\bar{\Sigma}^{-1})\otimes(T^{1/2}D_{\beta T}'c_\perp^{\#\prime}S_{11}\beta^{\#}) \\ (\bar{\Sigma}^{-1}\alpha)\otimes(T^{1/2}\beta^{\#\prime}S_{11}c_\perp^{\#}D_{\beta T}) & \bar{\Sigma}^{-1}\otimes(\beta^{\#\prime}S_{11}\beta^{\#}) \end{pmatrix} \\ &\quad + o_p(1) \\ &\xrightarrow{w} \begin{pmatrix} (\alpha'\bar{\Sigma}^{-1}\alpha)\otimes\left(\int_0^1 AG(s)G(s)'dsA'\right) & 0 \\ 0 & \bar{\Sigma}^{-1}\otimes\bar{\Sigma}_{\beta\beta} \end{pmatrix}, \qquad (A.12) \end{aligned}$$

where $A = \operatorname{diag}(c_\perp'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}, 1)$; this matrix arises from

$$T^{1/2}D_{\beta T}'c_\perp^{\#}Z_{1t} = \begin{pmatrix} T^{-1/2}c_\perp'X_{t-1} \\ 1 \end{pmatrix} = AB_T'Z_{1t} + o_p(1),$$

because $c_\perp'X_{t-1} = c_\perp'\beta(\beta'\beta)^{-1}\beta'X_{t-1} + c_\perp'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'X_{t-1}$. Note that $\bar{\Sigma}$ is the pseudo-true value of Σ; formally, one should include $\operatorname{vech}\Sigma$ in the parameter vector θ, or use the pseudo-log-likelihood concentrated with respect to Σ, but this will lead to the same result, since the asymptotic information is block-diagonal with respect to $\operatorname{vech}\Sigma$ and the other parameters. Combining (A.10), (A.11) and (A.12) leads to part (b).

For part (c), let $\tilde{\theta}$ denote the PMLEs under the restrictions in (5). Consistency of $\tilde{\theta}$, which is proved formally in Theorem 3, together with a second-order Taylor series expansion leads to the following expression of the PLR test statistic under H_0:

$$\mathrm{LR}_T := -2\left(\ell(\tilde{\theta}) - \ell(\hat{\theta})\right) = (\tilde{\theta}-\hat{\theta})'\hat{\mathcal{H}}(\tilde{\theta}-\hat{\theta}) + o_p(1),$$

where $\hat{\mathcal{H}} = -\partial^2\ell(\hat{\theta})/\partial\theta\partial\theta'$. Standard derivations involving restricted ML estimation lead to $D_T^{-1}(\tilde{\theta}-\hat{\theta}) = -D_T^{-1}\hat{\mathcal{H}}^{-1}R'(R\hat{\mathcal{H}}^{-1}R')^{-1}R(\hat{\theta}-\theta) + o_p(1)$. Let $\tilde{D}_T := \operatorname{diag}(T^{-1}I_{r_\beta}, T^{-1/2}I_{r_\alpha})$, so that

$$\tilde{D}_T^{-1}RD_T \to R. \tag{A.13}$$

Here we use the fact that the restrictions do not involve ρ_1, so that $R_\beta\hat{\beta}_2^{\#}$ only involves linear combinations of the $O_p(T^{-1})$-consistent estimator $\hat{\beta}_2$. Then

$$\begin{aligned} \mathrm{LR}_T &= (\hat{\theta}-\theta)'(D_TR'\tilde{D}_T^{-1})\left[(\tilde{D}_T^{-1}RD_T)(D_T\hat{\mathcal{H}}D_T)^{-1}(D_TR'\tilde{D}_T^{-1})\right]^{-1} \\ &\quad \times(\tilde{D}_T^{-1}RD_T)D_T^{-1}(\hat{\theta}-\theta) + o_p(1), \end{aligned}$$

and combining this with (A.10), (A.11), (A.12) and (A.13), this leads to the results in part (c), where $Z \sim N(0, I_{r_\alpha})$ is obtained as the limit in distribution of

$$\left(R_\alpha[I \otimes \bar{\Sigma}_{\beta\beta}^{-1}]\Omega[I \otimes \bar{\Sigma}_{\beta\beta}^{-1}]R_\alpha'\right)^{-1/2} T^{1/2} R_\alpha \operatorname{vec}(\hat{\alpha} - \alpha)'.$$

Because the asymptotic distribution of $T^{1/2}\operatorname{vec}(\hat{\alpha} - \alpha)'$ is defined from the Brownian motions W_i, see the proof of Lemma 2, and these are independent of W and hence (G_c, M_α), it follows that if $\varrho = 0$, then Z is also independent of these.

The proof of (14) follows as a specialization when $H_0 = H_{0\alpha}$ or $H_0 = H_{0\beta}$. □

Proof of Lemma 3. The proof follows the same approach as the proof of Lemma A.5 of Cavaliere, Rahbek and Taylor (2010b). As indicated there, it suffices to prove $\hat{V}_T(u) \xrightarrow{p} V(u)$ for fixed $u \in [0,1]$, which is then strengthened to uniform convergence because $\hat{V}_T(u)$ is monotonically increasing and $V(u)$ is continuous in u. Letting $V_T(u) = T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \varepsilon_t \varepsilon_t'$, we have $\hat{V}_T(u) - V_T(u) = o_p(1)$ because of consistency (at the appropriate rate) of $\hat{\gamma}$. Moreover,

$$\begin{aligned} V_T(u) &= T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \varepsilon_t \varepsilon_t' = T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \sigma_t z_t z_t' \sigma_t' \\ &= T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \sigma_t \left(z_t z_t' - h_t\right)\sigma_t' + T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \sigma_t \left(h_t - I_p\right)\sigma_t' + T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \sigma_t \sigma_t' \\ &=: A_{1T}(u) + A_{2T}(u) + A_{3T}(u), \end{aligned}$$

with $A_{1T}(\cdot)$, $A_{2T}(\cdot)$ and $A_{3T}(\cdot)$ defined implicitly. Now consider each of these terms. First, since $(z_t z_t' - h_t)$ is a martingale difference sequence with bounded fourth order moments and σ_t is non-stochastic and bounded, $\sigma_t(z_t z_t' - h_t)\sigma_t'$ is uncorrelated over time and $A_{1T}(u) \xrightarrow{p} 0$ by a standard application of Chebyshev's inequality. Second, since by Assumption 2 (i) $T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor}(h_t - I_p) \xrightarrow{p} 0$ and $\sigma_t = \sigma(t/T)$ with $\sigma(\cdot)$ càdlàg, by Lemma A.1 we have that $A_{2T}(u) \xrightarrow{p} 0$. Finally, Assumption 2 (a) implies that $A_{3T}(u) = \frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor}\Sigma(t/T) \to \int_0^u \Sigma(s)ds$. Taken together, these results imply that $V_T(u) \xrightarrow{p} \int_0^u \Sigma(s)ds$, as required. □

Proof of Theorem 2. The results will follow from the asymptotic properties $\hat{\mathcal{I}}$, which is defined from sums of the type $\sum_{t=1}^T(\hat{\varepsilon}_t\hat{\varepsilon}_t' \otimes Z_{it}Z_{jt}')$, $i,j = 1,2$. Consider first

$$\frac{1}{T}\sum_{t=1}^T(\hat{\varepsilon}_t\hat{\varepsilon}_t' \otimes B_T' Z_{1t} Z_{1t}' B_T) = \int_0^1 [d\hat{V}_T(s) \otimes G_T(s)G_T(s)'],$$

where the right-hand side should be read as a matrix with blocks $\int_0^1 G_T(s)G_T(s)'d\hat{V}_{T,ij}(s)$. Because Lemma 3 and the results in the proof of Lemma 1 imply $(G_T(\cdot), \hat{V}_T(\cdot)) \xrightarrow{w} (G(\cdot), V(\cdot))$, we find $\frac{1}{T}\sum_{t=1}^T(\hat{\varepsilon}_t\hat{\varepsilon}_t' \otimes B_T' Z_{1t} Z_{1t}' B_T) \xrightarrow{w} \int_0^1 [dV(s) \otimes G(s)G(s)'] = \int_0^1 [\Sigma(s) \otimes G(s)G(s)']ds$. Note that $V(\cdot)$ is a deterministic matrix function of bounded variation, so that the limiting integral is a Lebesgue-Stieltjes integral, and the result follows from the CMT. This in turn implies

$$\sum_{t=1}^T(\hat{\varepsilon}_t\hat{\varepsilon}_t' \otimes D_{\beta T}' c_\perp^{\#\prime} Z_{1t} Z_{1t}' c_\perp^{\#} D_{\beta T}) \xrightarrow{w} \int_0^1 [\Sigma(s) \otimes G_c(s)G_c(s)']ds, \tag{A.14}$$

where $D_{\beta T}$ and $G_c(\cdot)$ are as defined in (the proof of) Theorem 1. Next, consider

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t\hat{\varepsilon}_t'\otimes\mathbb{X}_{\beta,t-1}^{\#}\mathbb{X}_{\beta,t-1}^{\#\prime})=\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t\varepsilon_t'\otimes\mathbb{X}_{\beta,t-1}^{\#}\mathbb{X}_{\beta,t-1}^{\#\prime})+o_p(1),$$

where $\mathbb{X}_{\beta t}^{\#}$ is as defined in the proof of Lemma 2, and the asymptotic negligibility of the estimation error in $\hat{\varepsilon}_t$ follows from consistency at the appropriate rate of $\hat{\gamma}$, and the fact that $\mathbb{X}_{\beta,t-1}^{\#}$ is a stable process with bounded fourth moments. Using again $\mathbb{X}_{\beta,t-1}^{\#}=\Phi^{t-1}\mathbb{X}_{\beta 0}^{\#}+\sum_{i=1}^{t-1}\Phi^{i-1}F\varepsilon_{t-i}$, we find

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t\varepsilon_t'\otimes\mathbb{X}_{\beta,t-1}^{\#}\mathbb{X}_{\beta,t-1}^{\#\prime}) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j=1}^{t-1}(\varepsilon_t\varepsilon_t'\otimes\Phi^{i-1}F\varepsilon_{t-i}\varepsilon_{t-j}'F'\Phi'^{j-1})\\ &\quad+\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t\varepsilon_t'\otimes\Phi^{t-1}\mathbb{X}_{\beta 0}^{\#}\mathbb{X}_{\beta 0}^{\#\prime}\Phi'^{t-1})\\ &\quad+\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}(\varepsilon_t\varepsilon_t'\otimes[\Phi^{t-1}\mathbb{X}_{\beta 0}^{\#}\varepsilon_{t-i}'F'\Phi'^{i-1}+\Phi^{i-1}F\varepsilon_{t-i}\mathbb{X}_{\beta 0}^{\#\prime}\Phi'^{t-1}]).\end{aligned}$$

The second right-hand side term is $o_p(1)$ by L_1 convergence, because $\varepsilon_t\varepsilon_t'$ has bounded mean, and $\rho(\Phi)<1$. Similarly, the third term is $o_p(1)$, as $\varepsilon_t\varepsilon_t'\otimes\varepsilon_{t-i}$ has bounded mean by Assumption 2 (b) ii. (As before, take $z_t=0$ and hence $\varepsilon_t=0$ for $t\leq 0$.) By Assumption 2 (b) iii and the law of large numbers, we have $T^{-1}\sum_{t=1}^{T}(z_tz_t'\otimes z_{t-i}z_{t-j})\xrightarrow{p}\tau_{ij}$. Consequently,

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t\hat{\varepsilon}_t'\otimes\mathbb{X}_{\beta,t-1}^{\#}\mathbb{X}_{\beta,t-1}^{\#\prime})\\ &\quad\xrightarrow{p}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(I_p\otimes\Phi^{i-1}F)\int_0^1[\sigma(s)\otimes\sigma(s)]\tau_{ij}[\sigma(s)\otimes\sigma(s)]'ds(I_p\otimes\Phi^{j-1}F)',\quad\text{(A.15)}\end{aligned}$$

which equals $V=V^{(\infty)}$ defined in the proof of Lemma 2. Following the same approach, it can be shown that

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T}(\hat{\varepsilon}_t\hat{\varepsilon}_t'\otimes D_{\beta T}c_{\perp}^{\#}Z_{1t}\mathbb{X}_{\beta,t-1}^{\#\prime})\xrightarrow{p}0.\qquad\text{(A.16)}$$

The result of Theorem 2 now follows from (A.14)–(A.16), combined with (A.10)–(A.12) and the results of Theorem 1. □

Proof of Theorem 3. The proof mimics the proof of consistency given in Theorem 4.7 of Kristensen and Rahbek (2013) [KR13], and is based on an application of KR13 (Lemma D.1), where a third-order Taylor expansion of the log-likelihood function is employed. In fact, the proof we give below implies also weak convergence of the QMLE $\hat{\vartheta}$, see KR13 (Lemma D.2 with $v_T:=T$). With $\vartheta:=\left(\phi',\psi'\right)'\in\mathbb{R}^{l_\phi\times l_\psi}$, the criterion function in Lemma D.1 in KR13 reduces here to, cf. (6), $Q_T(\vartheta):=-\frac{1}{T}\ell(\vartheta)$ for which we need to evaluate first-, second- and third-order differentials. Moreover, in terms of Lemma D.1, set $U_T=V_T:=\sqrt{T}D_{\vartheta T}$, where $D_{\vartheta T}$ is a normalization matrix defined below.

Note initially that under H_0', $\operatorname{vec}(\alpha') = G\psi + g$ and $\operatorname{vec}\left(\beta^{\#}\right) = H\phi + h$, and therefore, using in particular $\beta^{\#} = \bar{c}^{\#} + c_{\perp}^{\#}\beta_2^{\#}$,

$$d\operatorname{vec}\left(\beta_2^{\#}\right) = \left(I_r \otimes \bar{c}_{\perp}^{\#\prime}\right) H d\phi \quad \text{and} \quad d\operatorname{vec}\left(\alpha'\right) = G d\psi, \tag{A.17}$$

Hence it follows immediately that the first- and second-order derivatives of $\ell(\cdot)$, and hence of $Q_T(\cdot)$, from the proof of Theorem 1 can be applied using classic rules for Jacobians as is done next.

Consider the first-order differentials of $\ell(\vartheta)$ in the direction $d\vartheta$, $d\ell(\vartheta; d\vartheta)$. We find that

$$-\tfrac{1}{T} d\ell(\vartheta; d\vartheta) = \operatorname{tr}\left\{\Sigma^{-1}\alpha d\beta^{\#\prime}\left[S_{10} - S_{11}\beta^{\#}\alpha'\right]\right\} + \operatorname{tr}\left\{\Sigma^{-1} d\alpha \beta^{\#\prime}\left[S_{10} - S_{11}\beta^{\#}\alpha'\right]\right\},$$

which evaluated at $\vartheta = \vartheta_0$, gives

$$\begin{aligned} -\tfrac{1}{T} d\ell(\vartheta_0; d\vartheta) &= \operatorname{tr}\left\{c_{\perp}^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\alpha d\beta_2^{\#\prime}\right\} + \operatorname{tr}\left\{\beta^{\#\prime} S_{1\varepsilon}\Sigma^{-1} d\alpha\right\} \\ &= d\operatorname{vec}\left(\beta_2^{\#}\right)' \operatorname{vec}\left(c_{\perp}^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\alpha\right) + d\operatorname{vec}\left(\alpha'\right) \operatorname{vec}\left(\beta^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\right). \end{aligned}$$

Observe, using standard matrix calculus, that $d\operatorname{vec}(\alpha')\operatorname{vec}\left(\beta^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\right) = d\psi' G' \operatorname{vec}\left(\beta^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\right)$, and, similarly, $d\operatorname{vec}\left(\beta_2^{\#}\right)' \operatorname{vec}\left(c_{\perp}^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\alpha\right) = d\phi' H' \operatorname{vec}\left(\bar{c}_{\perp}^{\#} c_{\perp}^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\alpha\right)$. Inserting the normalization matrix $D_{\beta T}$ from (A.9), we find,

$$d\phi' H' \operatorname{vec}\left(\bar{c}_{\perp}^{\#} D_{\beta T}' c_{\perp}^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\alpha\right) = d\phi' H'\left(I_r \otimes \bar{c}_{\perp}^{\#} D_{\beta T}'\right) \operatorname{vec}\left(c_{\perp}^{\#\prime} S_{1\varepsilon}\Sigma^{-1}\alpha\right).$$

That is, the normalisation $D_{\phi T} := \left(I_r \otimes D_{\beta T}\bar{c}_{\perp}^{\#\prime}\right) H$ replaces $(I_r \otimes D_{\beta T})$ in (A.9). Likewise, $\sqrt{T} G' \operatorname{vec}\left(\beta^{\#\prime} S_{1\varepsilon}\bar{\Sigma}^{-1}\right)$ converges, such that with $D_{\vartheta T} := \operatorname{diag}\left(D_{\phi T}, T^{-1/2} I_{l_\psi}\right)$ replacing D_T in (A.9), we have that

$$D_{\vartheta T}'\left[-\tfrac{1}{T}\tfrac{\partial \ell(\vartheta)}{\partial \vartheta}\right]\Big|_{\vartheta=\vartheta_0} \xrightarrow{w} \begin{pmatrix} H' \operatorname{vec}\left(\bar{c}_{\perp}^{\#} A \int_0^1 G(s) dM(s)' \bar{\Sigma}^{-1}\alpha\right) \\ N\left(0, G'\left[\bar{\Sigma}^{-1} \otimes I_r\right]\Omega\left[\bar{\Sigma}^{-1} \otimes I_r\right] G\right) \end{pmatrix}.$$

Consider next the second-order differentials of $\ell(\vartheta)$ in the directions $d\vartheta$ and $d\bar{\vartheta}$, $d^2\ell\left(\vartheta; d\vartheta, d\bar{\vartheta}\right)$. Observe that by definition,

$$\begin{aligned} -\tfrac{1}{T} d^2\ell\left(\vartheta; d\vartheta, d\bar{\vartheta}\right) &= \operatorname{tr}\left\{\Sigma^{-1} d\bar{\alpha} d\beta^{\#\prime}\left[S_{10} - S_{11}\beta^{\#}\alpha'\right]\right\} - \operatorname{tr}\left\{\Sigma^{-1}\alpha d\beta^{\#\prime} S_{11}\beta^{\#} d\bar{\alpha}'\right\} \\ &\quad - \operatorname{tr}\left\{\alpha'\Sigma^{-1}\alpha d\beta^{\#\prime} S_{11} d\bar{\beta}^{\#}\right\} + \operatorname{tr}\left\{\Sigma^{-1} d\alpha d\bar{\beta}^{\#\prime}\left[S_{10} - S_{11}\beta^{\#}\alpha'\right]\right\} \\ &\quad + \operatorname{tr}\left\{\Sigma^{-1} d\alpha \beta^{\#\prime} S_{11}\beta^{\#} d\bar{\alpha}'\right\} - \operatorname{tr}\left\{\alpha'\Sigma^{-1} d\alpha \beta^{\#\prime} S_{11} d\bar{\beta}^{\#}\right\} \end{aligned}$$

and, hence, at $\vartheta = \vartheta_0$,

$$\begin{aligned} -\tfrac{1}{T} d^2\ell\left(\vartheta_0; d\phi, d\phi\right) &= -\operatorname{tr}\left\{\alpha'\Sigma^{-1}\alpha d\beta^{\#\prime} S_{11} d\beta^{\#}\right\} \\ &= -\operatorname{vec}\left(d\beta_2^{\#}\right)'\left(\alpha'\Sigma^{-1}\alpha \otimes c_{\perp}^{\#\prime} S_{11} c_{\perp}^{\#}\right) \operatorname{vec}\left(d\beta_2^{\#}\right) \\ &= -d\phi' H'\left(\alpha'\Sigma^{-1}\alpha \otimes \bar{c}_{\perp}^{\#} c_{\perp}^{\#\prime} S_{11} c_{\perp}^{\#}\bar{c}_{\perp}^{\#\prime}\right) H d\phi. \end{aligned}$$

As before, inserting $D_{\phi T}$ we find,

$$\begin{aligned} & d\phi' D'_{\phi T}\left(\alpha'\Sigma^{-1}\alpha\otimes\bar{c}_{\perp}^{\#}c_{\perp}^{\#\prime}S_{11}c_{\perp}^{\#}\bar{c}_{\perp}^{\#\prime}\right)D_{\phi T}Hd\phi \\ & \quad = d\phi' H'\left(\alpha'\Sigma^{-1}\alpha\otimes\bar{c}_{\perp}^{\#}\left[D'_{\beta T}c_{\perp}^{\#\prime}S_{11}c_{\perp}^{\#}D_{\beta T}\right]\bar{c}_{\perp}^{\#\prime}\right)Hd\phi \end{aligned}$$

which converges weakly, see (A.10). Likewise,

$$\begin{aligned} -\tfrac{1}{T}d^2\ell\left(\vartheta_0;d\psi,d\psi\right) &= -\operatorname{tr}\left\{\Sigma^{-1}d\alpha\beta^{\#\prime}S_{11}\beta^{\#}d\alpha'\right\} \\ &= -d\psi' G'\left(\Sigma^{-1}\otimes\beta^{\#\prime}S_{11}\beta^{\#}\right)Gd\psi, \end{aligned}$$

which converges in probability, see (A.10). Finally, $-\frac{1}{T}d^2\ell\left(\vartheta_0;d\phi,d\psi\right)=\operatorname{tr}\left\{\Sigma^{-1}d\alpha d\beta^{\#\prime}S_{1\varepsilon}\right\}-\operatorname{tr}\left\{\Sigma^{-1}\alpha d\beta^{\#\prime}S_{11}\beta^{\#}d\alpha'\right\}$, where, for example, the last term can be rewritten as,

$$\begin{aligned} \operatorname{tr}\left\{\Sigma^{-1}\alpha d\beta^{\#\prime}S_{11}\beta^{\#}d\alpha'\right\} &= \operatorname{vec}\left(d\beta_2^{\#}\right)'\left(\alpha'\Sigma^{-1}\otimes c_{\perp}^{\#\prime}S_{11}\beta^{\#}\right)\operatorname{vec}\left(d\alpha'\right) \\ &= d\phi' H'\left(I_r\otimes\bar{c}_{\perp}^{\#}\right)\left(\alpha'\Sigma^{-1}\otimes c_{\perp}^{\#\prime}S_{11}\beta^{\#}\right)Gd\psi, \end{aligned}$$

and $D_{\beta T}$ can be inserted as above. Collecting terms, we find, using (A.10), that

$$\begin{aligned} & D'_{\vartheta T}\left(-\tfrac{\partial^2\ell(\vartheta)}{\partial\vartheta\partial\vartheta'}\right)\Big|_{\vartheta=\vartheta_0,\Sigma=\bar{\Sigma}}D_{\vartheta T}\xrightarrow{w} \\ & \begin{pmatrix} H'\left(\alpha'\Sigma^{-1}\alpha\otimes\bar{c}_{\perp}^{\#}\left[\int_0^1 AG(s)G(s)'A'ds\right]\bar{c}_{\perp}^{\#\prime}\right)H & 0 \\ 0 & G'\left(\bar{\Sigma}^{-1}\otimes\Sigma_{\beta\beta}\right)G \end{pmatrix}. \end{aligned}$$

Finally, consider third-order differentials of $\ell\left(\vartheta\right)$, $d^3\ell\left(\cdot\right)$ say, in the directions $d\vartheta$, $d\bar{\vartheta}$ and $d\vartheta^*$. Using Lemma D.1 in KR13, the supremum of the norm of $\frac{1}{T}d^3\ell\left(\cdot\right)$ needs to be bounded in probability uniformly over $\vartheta\in\mathcal{N}_T\left(\vartheta_0\right)$, where the sequence of neighbourhoods is given by, $\mathcal{N}_T\left(\vartheta_0\right)=\left\{\vartheta\mid\Upsilon_T^{1/2}\left\|\vartheta-\vartheta_0\right\|\leq\epsilon\right\}$. Here $\Upsilon_T^{1/2}:=T^{-1/2}D_{\vartheta T}^{-1}$ and (some small) $\epsilon>0$. In terms of $d\vartheta$ in the differential, $d\vartheta$ should be normalised as $\Upsilon_T^{-1/2}d\vartheta$; and likewise for $\bar{\vartheta}$, ϑ^*. Now, the only non-zero third-order differentials of $\frac{1}{T}\ell(\vartheta)$ have the form, $\operatorname{tr}\{d\alpha^{*\prime}\Sigma^{-1}d\bar{\alpha}d\beta_2^{\#\prime}c_{\perp}^{\#\prime}S_{11}\beta\}$ and $\operatorname{tr}\{\alpha'\Sigma^{-1}d\bar{\alpha}d\beta_2^{\#\prime}c_{\perp}^{\#\prime}S_{11}c_{\perp}^{\#}d\beta_2^{\#*}\}$. Consequently, using the identities in (A.17), and in particular that in $\mathcal{N}_T\left(\vartheta_0\right)$, $\phi-\phi_0=T^{-1/2}h$, with $\left\|h\right\|\leq\epsilon$, it follows that,

$$T^{-1}\sup_{\vartheta\in\mathcal{N}_T(\vartheta_0)}\left\|d^3\ell\left(\vartheta;\Upsilon_T^{-1/2}d\vartheta,\Upsilon_T^{-1/2}d\vartheta\Upsilon_T^{-1/2}d\bar{\vartheta},\Upsilon_T^{-1/2}d\vartheta^*\right)\right\|,$$

is bounded by $C_T\left\|d\vartheta\right\|\left\|d\bar{\vartheta}\right\|\left\|d\vartheta^*\right\|\left\|\frac{1}{T}S_{11}\right\|=O_p\left(1\right)$, since $\left\|\frac{1}{T}S_{11}\right\|$ converges weakly, and C_T contains (standard) terms converging in probability. The consistency $\tilde{V}_T$ follows by similar arguments as in the proof of Lemma 3. □

A.3 Results for the bootstrap tests

Proof of Lemma 4. The proof follows as in the proof of Lemma A.5 in Cavaliere, Rahbek and Taylor (2010b) by showing that the conditional variance of $M_T^*(\cdot)$ satisfies

$$E^*\left(M_T^*(u)M_T^*(u)'\right)=\frac{1}{T}\sum_{t=1}^{\lfloor Tu\rfloor}\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}=\frac{1}{T}\sum_{t=1}^{\lfloor Tu\rfloor}\tilde{\varepsilon}_t\tilde{\varepsilon}'_t+o_p\left(1\right)\xrightarrow{p}\int_0^u\Sigma\left(s\right)ds.$$

The latter convergence is established in Theorem 3 and this completes the proof. □

Proof of Lemma 5. As done for the proof of Lemma 2, define $\mathbb{X}_{\beta t}^{\#*} = (Z_{\beta t+1}^{*\prime}, Z_{2t+1}^{*\prime})'$, with $Z_{\beta t}^{*} = \tilde{\beta}^{\#\prime} Z_{1t}^{*}$; therefore, $\mathbb{X}_{\beta t}^{\#*}$ equals $\mathbb{X}_{\beta t}^{*}$ except that $\tilde{\rho}_1' D_{1t}$ has been added to $\tilde{\beta}' X_{t-1}^{*}$. This means that $\mathbb{X}_{\beta t}^{\#*} = \tilde{\Phi}\mathbb{X}_{\beta,t-1}^{\#*} + \tilde{F}\varepsilon_t^{*}$, where $\tilde{\Phi}$ and $\tilde{F}$ are defined as Φ and F but with the restricted estimators replacing the population parameter values. Moreover, since $\|\Phi^i\| \leq c\lambda^i$ for some positive constant c and $|\lambda| < 1$ (see the proof of Lemma 2), the consistency of the restricted estimators implies that for T large enough, $\|\tilde{\Phi}^i\| \leq c\lambda^i$. The initial value is $\mathbb{X}_{\beta,0}^{\#*} = \mathbb{X}_{\beta,0}^{\#}$, as the bootstrap sample is initialised at the original data. We may therefore write $\mathbb{X}_{\beta t}^{\#*} = \tilde{\Phi}^t \mathbb{X}_{\beta 0}^{\#*} + \sum_{i=0}^{t-1} \tilde{\Phi}^j \tilde{F}\varepsilon_{t-i}^{*} = \tilde{\Phi}^t \mathbb{X}_{\beta 0}^{\#} + \sum_{i=0}^{t-1} \tilde{\Phi}^i \tilde{F}\varepsilon_{t-i}^{*} = \sum_{i=0}^{t-1} \tilde{\Phi}^i \tilde{F}\varepsilon_{t-i}^{*} + o_p(1)$, where the $o_p(1)$ term will be neglected in the following.

First, as in Lemma A.7 of Cavaliere, Rahbek and Taylor (2010b) we have that

$$\begin{pmatrix} \tilde{\beta}^{\#\prime} M_{11}^{*} \tilde{\beta}^{\#} & \tilde{\beta}^{\#\prime} M_{12} \\ M_{21}\tilde{\beta}^{\#} & M_{22} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{X}_{\beta,t-1}^{\#*} \mathbb{X}_{\beta,t-1}^{\#*\prime} \xrightarrow{p^*}_{p} \sum_{i=0}^{\infty} \Phi^i F \bar{\Sigma} F' \Phi^{i\prime} =: \begin{pmatrix} \bar{M}_{\beta\beta} & \bar{M}_{\beta 2} \\ \bar{M}_{2\beta} & \bar{M}_{22} \end{pmatrix}, \tag{A.18}$$

from which we find $\hat{\beta}^{\#\prime} S_{11}^{*} \hat{\beta}^{\#} = \hat{\beta}^{\#\prime} M_{11}^{*} \hat{\beta}^{\#} - \hat{\beta}^{\#\prime} M_{12}^{*} M_{22}^{*-1} M_{21}^{*} \hat{\beta}^{\#} \xrightarrow{p^*}_{p} \bar{M}_{\beta\beta} - \bar{M}_{\beta 2}\bar{M}_{22}^{-1}\bar{M}_{2\beta} = \bar{\Sigma}_{\beta\beta}$. The convergence in (A.18) follows by the triangle inequality using the following two results. First, as $\varepsilon_t^{*}\varepsilon_t^{*\prime} = \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' w_t^2$ and hence $E^{*}(\varepsilon_t^{*}\varepsilon_t^{*\prime}) = \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'$, we have that

$$\begin{aligned}
E^{*}\left(\frac{1}{T}\sum_{t=1}^{T} \mathbb{X}_{\beta,t-1}^{\#*}\mathbb{X}_{\beta,t-1}^{\#*\prime}\right) &= \frac{1}{T}\sum_{t=1}^{T} E^{*}\left(\mathbb{X}_{\beta,t-1}^{\#*}\mathbb{X}_{\beta,t-1}^{\#*\prime}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1} \tilde{\Phi}^{i-1}\tilde{F} E^{*}\left(\varepsilon_{t-i}^{*}\varepsilon_{t-i}^{*\prime}\right)\tilde{F}'\tilde{\Phi}^{i-1\prime} \\
&= \sum_{i=1}^{T-1} \tilde{\Phi}^{i-1}\tilde{F}\left(\frac{1}{T}\sum_{t=1}^{T-i} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\right)\tilde{F}'\tilde{\Phi}^{i-1\prime} \\
&= \sum_{i=1}^{T-1} \tilde{\Phi}^{i-1}\tilde{F}\left(\frac{1}{T}\sum_{t=1}^{T} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\right)\tilde{F}'\tilde{\Phi}^{i-1\prime} + \sum_{i=1}^{T-1} \tilde{\Phi}^{i-1}\tilde{F}\left(\frac{1}{T}\sum_{t=T-i+1}^{T} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\right)\tilde{F}'\tilde{\Phi}^{i} \\
&\xrightarrow{p} \sum_{i=0}^{\infty} \Phi^i F \bar{\Sigma} F' \Phi^{i\prime},
\end{aligned}$$

since, by consistency of the restricted estimators and Lemma A.1(i), $\sum_{i=1}^{T-1} \tilde{\Phi}^{i-1}\tilde{F}\left(T^{-1}\sum_{t=1}^{T} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\right)\tilde{F}'\tilde{\Phi}^{i-1} \xrightarrow{p} \sum_{i=0}^{\infty} \Phi^i F\bar{\Sigma}F'\Phi^{i\prime}$, and furthermore

$$\begin{aligned}
\left\|\sum_{i=1}^{T-1} \tilde{\Phi}^{i-1}\tilde{F}\left(\frac{1}{T}\sum_{t=T-i+1}^{T} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\right)\tilde{F}'\tilde{\Phi}^{i-1}\right\| &\leq \left\|\tilde{F}\right\|^2 \left(\sum_{i=0}^{T-2} \left\|\tilde{\Phi}^i\right\|^2 i\right)\frac{1}{T}\max_{t=1,\ldots,T}\left\|\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\right\| \\
&\leq c_T \frac{1}{T}\max_{t=1,\ldots,T}\left\|\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\right\| = o_p(1),
\end{aligned}$$

as $c_T = O_p(1)$ and

$$P\left(\frac{1}{T}\max_{t=1,\ldots,T}\left\|\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\right\|\geq\epsilon\right) \leq \sum_{t=1}^{T}P\left(\left\|\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\right\|\geq\epsilon T\right)$$
$$\leq \frac{\sup_t E\left(\left\|\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\right\|^2\right)}{\epsilon^2 T}=\frac{E\left(\left(\tilde{\varepsilon}'_{c,t}\tilde{\varepsilon}_{c,t}\right)^2\right)}{\epsilon^2 T}\to 0$$

as ε_t has finite 4^+ order moments.

Second, using the fact that $\varepsilon_t^*\varepsilon_t^{*\prime}-E^*\left(\varepsilon_t^*\varepsilon_t^{*\prime}\right)=\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\left(w_t^2-1\right)$, we have that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{X}_{\beta,t-1}^{\#*}\mathbb{X}_{\beta,t-1}^{\#*\prime}-E^*\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{X}_{\beta,t-1}^{\#*}\mathbb{X}_{\beta,t-1}^{\#*\prime}\right)$$
$$=\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\tilde{\Phi}^{i-1}\tilde{F}\left(\varepsilon_{t-i}^*\varepsilon_{t-i}^{*\prime}-E^*\left(\varepsilon_{t-i}^*\varepsilon_{t-i}^{*\prime}\right)\right)\tilde{F}'\tilde{\Phi}^{i-1\prime}+R_T^*$$
$$=\sum_{i=0}^{T-1}\tilde{\Phi}^{i-1}\tilde{F}\left(\frac{1}{T}\sum_{t=1}^{T-i}\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\left(w_t^2-1\right)\right)\tilde{F}'\tilde{\Phi}^{i-1\prime}+R_T^*, \tag{A.19}$$

where $R_T^*:=T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j=1,j\neq i}^{t-1}\tilde{\Phi}^{i-1}\tilde{F}(\varepsilon_{t-i}^*\varepsilon_{t-j}^{*\prime})\tilde{F}'\tilde{\Phi}^{j-1\prime}$. To show that (A.19) is of $o_p^*(1)$ in probability, notice first that

$$\operatorname{vec}\left(\sum_{i=1}^{T-1}\tilde{\Phi}^{i-1}\tilde{F}\left(\frac{1}{T}\sum_{t=1}^{T-i}\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\left(w_t^2-1\right)\right)\tilde{F}'\tilde{\Phi}^{i-1\prime}\right)$$
$$=\left(\sum_{i=1}^{T-1}\left(\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\otimes\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\frac{1}{T}\sum_{t=1}^{T-i}\operatorname{vec}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\right)\left(w_t^2-1\right)\right)\right)$$
$$=\left(\left(\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i=1}^{T-t}\left(\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\otimes\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\right)\operatorname{vec}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\right)\left(w_t^2-1\right)\right)\right),$$

which (conditionally on the data) is the average of an independent sequence with variance

$$\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{i=1}^{T-t}\left(\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\otimes\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\right)\operatorname{vec}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\right)\operatorname{vec}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\right)'\left(\sum_{i=1}^{T-t}\left(\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\otimes\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\right)'$$

which is of $o_p(1)$ as $\sum_{i=1}^{T-t}\left(\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\otimes\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)=O_p(1)$ and ε_t has finite fourth order moments. Similarly it can be shown that R_T^* is of $o_p^*(1)$, in probability.

In the same way of (A.18) it can be proved that

$$\begin{pmatrix}\tilde{\beta}^{\#\prime}M_{1\varepsilon}^*\\ M_{2\varepsilon}^*\end{pmatrix}:=\frac{1}{T}\sum_{t=1}^{T}\mathbb{X}_{\beta,t-1}^{\#*}\varepsilon_t^{*\prime}\xrightarrow{p^*}_p 0, \tag{A.20}$$

which also implies, as in the proof of Lemma 2, that $\tilde{\beta}^{\#\prime}S_{10}^*\xrightarrow{p^*}_p\bar{\Sigma}_{\beta\beta}\alpha'=:\bar{\Sigma}_{\beta 0}$, and $S_{00}^*\xrightarrow{p^*}_p\bar{\Sigma}_{00}$. This proves (22).

To prove (23) and (24), as in the proof of Lemma 2 notice that we have

$$\tilde{B}_T'Z_{1t}^*=\begin{pmatrix}T^{-1/2}(\tilde{\beta}'_\perp\tilde{C}\sum_{i=1}^{t-1}\varepsilon_i^*+\tilde{\beta}'_\perp S_{t-1}^*+\tilde{\beta}'_\perp\tilde{C}_0)\\ 1\end{pmatrix}+o_p^*(1)$$

in probability (see also Lemma A.4 in Cavaliere, Rahbek and Taylor (2010b) for details on the $o_p^*(1)$ term). Letting $G_T^*(\cdot) = \tilde{B}_T' Z_{1,\lfloor T\cdot\rfloor+1}^*$, we have $G_T^*(\cdot) \overset{w^*}{\to}_p G(\cdot)$. Using the CMT and Theorem 2.1 of Hansen (1992b), together with Lemma 4, we have that

$$\begin{aligned}\left(\tilde{B}_T' M_{11}^* \tilde{B}_T, \tilde{B}_T M_{1\varepsilon}^*\right) &= \left(\int_0^1 G_T^*(s) G_T^*(s)' ds, \int_0^1 G_T^*(s) dM_T^*(s)'\right) \\ &\overset{w}{\to}_p \left(\int_0^1 G(s)G(s)'ds, \int_0^1 G(s)dM(s)'\right).\end{aligned}$$

Because Z_{2t}^* is a linear process in ε_t^* with exponentially decaying weights, $\tilde{B}_T' M_{12}^* = O_p^*(T^{-1/2})$ and $\tilde{B}_T' M_{11}^* \tilde{\beta}^{\#} = O_p^*(T^{-1/2})$, in probability, which together with (A.18) and (A.20), implies that (A.8)–(A.9) also holds with M_{11}^* and $M_{1\varepsilon}^*$ replaced by S_{11}^* and $S_{1\varepsilon}^*$, respectively.

Finally, to prove (25), we start with

$$\begin{aligned}Y_t^* &:= \operatorname{vec} \mathbb{X}_{\beta,t-1}^{\#*} \varepsilon_t^{*\prime} = \operatorname{vec}\left(\tilde{\Phi}^{t-1}\mathbb{X}_{\beta 0}^{\#*} + \sum_{i=1}^{t-1}\tilde{\Phi}^{i-1}\tilde{F}\varepsilon_{t-i}^*\right)\varepsilon_t^{*\prime} \\ &= \operatorname{vec}\left(\tilde{\Phi}^{t-1}\mathbb{X}_{\beta 0}^{\#*}\varepsilon_t^{*\prime}\right) + \operatorname{vec}\left(\sum_{i=1}^{t-1}\tilde{\Phi}^{i-1}\tilde{F}\varepsilon_{t-i}^*\varepsilon_t^{*\prime}\right) \\ &= \operatorname{vec}\left(\tilde{\Phi}^{t-1}\mathbb{X}_{\beta 0}^{\#*}\varepsilon_t^{*\prime}\right) + \sum_{i=1}^{t-1}\left(\varepsilon_t^* \otimes \tilde{\Phi}^{i-1}\tilde{F}\varepsilon_{t-i}^*\right) \\ &= \operatorname{vec}\left(\tilde{\Phi}^{t-1}\mathbb{X}_{\beta 0}^{\#*}\varepsilon_t^{*\prime}\right) + \sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\varepsilon_t^* \otimes \varepsilon_{t-i}^*\right) =: Y_{0,t}^* + Y_{1,t}^*\end{aligned}$$

with $Y_{0,t}^*$ and $Y_{1,t}^*$ defined implicitly. We now prove that show that $T^{-1/2}\sum_{t=1}^T Y_t^*$ satisfies (in probability) a CLT by showing (i) that $T^{-1/2}\sum_{t=1}^T Y_{1,t}^*$ satisfies (in probability) a CLT and (ii) that $T^{-1/2}\sum_{t=1}^T Y_{0,t}^* = o_p^*(1)$ (in probability).

Part (i). With $\mathcal{F}_t^* := \sigma(w_1, w_2, \ldots, w_t)$, notice that, conditionally on the data, $E^*\left(Y_{1,t}^* | \mathcal{F}_{t-1}^*\right) = 0$ and hence $\{Y_{1,t}^*, \mathcal{F}_t^*\}$ is a vector martingale difference sequence. We prove that a central limit theorem holds on $T^{-1/2}\sum_{t=1}^T Y_{1,t}^*$ by proving that its conditional variance converges in probability and that (conditionally on the data) the Lindeberg condition holds.

First, since for $i \geq 1$, $E^*\left(\left(\varepsilon_t^* \otimes \varepsilon_{t-i}^*\right)\left(\varepsilon_t^* \otimes \varepsilon_{t-i}^*\right)' | \mathcal{F}_{t-1}^*\right) = E^*\left(\varepsilon_t^*\varepsilon_t^{*\prime} \otimes \varepsilon_{t-i}^*\varepsilon_{t-i}^{*\prime} | \mathcal{F}_{t-1}^*\right) = \left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \varepsilon_{t-i}^*\varepsilon_{t-i}^{*\prime}\right)$, we have that

$$\begin{aligned}E^*\left(Y_{1,t}^* Y_{1,t}^{*\prime} | \mathcal{F}_{t-1}^*\right) &= \sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right) E^*\left(\left(\varepsilon_t^*\varepsilon_t^{*\prime} \otimes \varepsilon_{t-i}^*\varepsilon_{t-i}^{*\prime}\right) | \mathcal{F}_{t-1}^*\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)' \\ &= \sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{t-i}\tilde{\varepsilon}_{t-i}'\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\ &\quad + \sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{t-i}\tilde{\varepsilon}_{t-i}'\eta_{t-i}\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right),\end{aligned}$$

where $\eta_t := w_t^2 - 1$. Averaging across t yields

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} E^*\left(Y_t^* Y_t^{*\prime} | \mathcal{F}_{t-1}^*\right) \\
&\quad = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\
&\qquad + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\eta_{t-i}\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\
&\quad =: V_{0,T} + V_{1,T}^*,
\end{aligned}
$$

where $V_{0,T}$ and $V_{1,T}^*$ are implicitly defined. Notice that $V_{1,T}^* = o_p^*(1)$, in probability, since by Chebychev's inequality

$$
\begin{aligned}
V_{1,T}^* &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\eta_{t-i}\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\
&= \frac{1}{T}\sum_{t=1}^{T-1}\sum_{i=1}^{T-t}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\tilde{\varepsilon}_{c,t+i}\tilde{\varepsilon}_{c,t+i}' \otimes \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\eta_t\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\
&= o_p^*(1) \text{ in probability,}
\end{aligned}
$$

as η_t is (conditionally) independent and ε_t has finite 8^+ moments.

Regarding $V_{0,T}$, we have that

$$
\begin{aligned}
V_{0,T} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\
&= \sum_{i=0}^{T-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\frac{1}{T}\sum_{t=1}^{T-i}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\
&= V_{0,T}^{(n)} + R_{0,T}^{(n)},
\end{aligned}
$$

where, for any $n \leq T-1$, we set

$$
V_{0,T}^{(n)} := \sum_{i=0}^{n}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\frac{1}{T}\sum_{t=1}^{T-i}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right),
$$

and

$$
\begin{aligned}
R_{0,T}^{(n)} &:= \sum_{i=n+1}^{T-1}\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\frac{1}{T}\sum_{t=i+1}^{T}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \otimes \tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\right)\left(I_p \otimes \left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\
&= \sum_{i=n+1}^{T-1}\left(I_p \otimes \left(\Phi^{i-1}F\right)\right)\frac{1}{T}\sum_{t=i+1}^{T}\left(\varepsilon_t\varepsilon_t' \otimes \varepsilon_{t-i}\varepsilon_{t-i}'\right)\left(I_p \otimes \left(\Phi^{i-1}F\right)'\right) + o_p(1),
\end{aligned}
$$

where the $o_p(1)$ term does not depend on n. We now derive the limit of $V_{0,T}^{(n)}$ as $T \to \infty$ for fixed n, and then let $n \to \infty$ to obtain the limiting distribution of $V_{0,T}$, where we also need to show that $\lim_{n\to\infty} \limsup_{T\to\infty} P(\|R_{0,T}^{(n)}\| > \epsilon) = 0$ for all $\epsilon > 0$, as in the proof of Lemma 2.

For fixed n,

$$\begin{aligned} V_{0,T}^{(n)} &= \sum_{i=0}^{n}\left(I_p\otimes\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\right)\frac{1}{T}\sum_{t=1}^{T-i}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\otimes\tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\right)\left(I_p\otimes\left(\tilde{\Phi}^{i-1}\tilde{F}\right)'\right) \\ &\xrightarrow{p} \sum_{i=0}^{n}\left(I_p\otimes\left(\Phi^{i-1}F\right)\right)\int_0^1[\sigma(s)\otimes\sigma(s)]\tau_{ii}[\sigma(s)\otimes\sigma(s)]'ds\left(I_p\otimes\left(\Phi^{i-1}F\right)'\right) =: V^{(n)\dagger} \end{aligned}$$

since for any fixed i, we have that

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T-i}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\otimes\tilde{\varepsilon}_{c,t-i}\tilde{\varepsilon}_{c,t-i}'\right) = \frac{1}{T}\sum_{t=1}^{T-i}\left(\varepsilon_t\varepsilon_t'\otimes\varepsilon_{t-i}\varepsilon_{t-i}'\right)+o_p(1) \\ &\quad= \frac{1}{T}\sum_{t=1}^{T-i}\left(\left(\varepsilon_t\varepsilon_t'-\sigma_t h_t\sigma_t'\right)\otimes\varepsilon_{t-i}\varepsilon_{t-i}'\right)+\frac{1}{T}\sum_{t=1}^{T-i}\left(\sigma_t h_t\sigma_t'\otimes\sigma_{t-i}z_{t-i}z_{t-i}'\sigma_{t-i}'\right)+o_p(1) \\ &\quad\to_p \int_0^1[\sigma(s)\otimes\sigma(s)]\tau_{ii}[\sigma(s)\otimes\sigma(s)]'ds \end{aligned}$$

by Lemma A.1(iii) and as $\left(\left(\varepsilon_t\varepsilon_t'-\sigma_t h_t\sigma_t'\right)\otimes\varepsilon_{t-i}\varepsilon_{t-i}'\right)$ is a martingale difference array with bounded 1^+ moments, so that a WLLN applies (see e.g. Hall and Heyde (1980, Theorem 2.13 (i))). Next, we have that

$$\lim_{n\to\infty}V^{(n)\dagger}=\sum_{i=0}^{\infty}\left(I_p\otimes\left(\Phi^{i-1}F\right)\right)\int_0^1[\sigma(s)\otimes\sigma(s)]\tau_{ii}[\sigma(s)\otimes\sigma(s)]'ds\left(I_p\otimes\left(\Phi^{i-1}F\right)'\right)=:V^{\dagger}.$$

Finally we have that, for T large enough,

$$\begin{aligned} P(\|R_{0,T}^{(n)}\|>\epsilon) &= P\left(\left\|\sum_{i=n+1}^{T-1}\left(I_p\otimes\left(\Phi^{i-1}F\right)\right)\frac{1}{T}\sum_{t=i+1}^{T}\left(\varepsilon_t\varepsilon_t'\otimes\varepsilon_{t-i}\varepsilon_{t-i}'\right)\left(I_p\otimes\left(\Phi^{i-1}F\right)'\right)\right\|>\epsilon\right) \\ &\le P\left(\sum_{i=n+1}^{T-1}\left\|\left(I_p\otimes\left(\Phi^{i-1}F\right)\right)\right\|^2\left\|\frac{1}{T}\sum_{t=i+1}^{T}\left(\varepsilon_t\varepsilon_t'\otimes\varepsilon_{t-i}\varepsilon_{t-i}'\right)\right\|>\epsilon\right) \\ &\le \frac{E\left(\sum_{i=n+1}^{T-1}\left\|\left(I_p\otimes\left(\Phi^{i-1}F\right)\right)\right\|^2\left\|\frac{1}{T}\sum_{t=i+1}^{T}\left(\varepsilon_t\varepsilon_t'\otimes\varepsilon_{t-i}\varepsilon_{t-i}'\right)\right\|\right)}{\epsilon} \\ &= \frac{\sum_{i=n+1}^{T-1}\left\|\left(I_p\otimes\left(\Phi^{i-1}F\right)\right)\right\|^2 E\left\|\frac{1}{T}\sum_{t=i+1}^{T}\left(\varepsilon_t\varepsilon_t'\otimes\varepsilon_{t-i}\varepsilon_{t-i}'\right)\right\|}{\epsilon}\le\frac{K\lambda^{2n}}{\epsilon}\to 0 \end{aligned}$$

as $n\to\infty$. This completes the proof that $T^{-1}\sum_{t=1}^{T}E^*\left(Y_{1,t}^*Y_{1,t}^{*\prime}|\mathcal{F}_{t-1}^*\right)\xrightarrow{p^*}_p V^{\dagger}$.

We prove the (conditional) Lindberg conditions by showing that $T^{-2}\sum_{t=1}^{T}E^*\left(\|Y_{1,t}^*\|^4\right)\to_p 0$. First, we have that, since ε_t^* is independent under P^*,

$$\begin{aligned} E^*\left\|Y_{1,t}^*\right\|^4 &= E^*\left\|\varepsilon_t^*\otimes\left(\sum_{i=1}^{t-1}\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\varepsilon_{t-i}^*\right)\right\|^4 = E^*\left\|\varepsilon_t^*\right\|^4 E^*\left\|\sum_{i=1}^{t-1}\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\varepsilon_{t-i}^*\right\|^4 \\ &\le K\left\|\tilde{\varepsilon}_{c,t}\right\|^4 E^*\left\|\sum_{i=1}^{t-1}\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\varepsilon_{t-i}^*\right\|^4 \end{aligned}$$

with $K=E|w_t^4|$. Moreover,

$$
\begin{aligned}
&E^*\left\|\sum_{i=1}^{t-1}\left(\tilde{\Phi}^{i-1}\tilde{F}\right)\varepsilon^*_{t-i}\right\|^4 \\
&\quad = E^*\left(\left(\sum_{i=1}^{t-1}\varepsilon^{*\prime}_{t-i}\tilde{F}'\tilde{\Phi}^{i-1\prime}\right)\left(\sum_{i=1}^{t-1}\tilde{\Phi}^{i-1}\tilde{F}\varepsilon^*_{t-i}\right)\right)^2 \\
&\quad = E^*\left(\sum_{i=1}^{t-1}\sum_{j=1}^{t-1}\varepsilon^{*\prime}_{t-i}\tilde{F}'\tilde{\Phi}^{i-1\prime}\tilde{\Phi}^{j-1}\tilde{F}\varepsilon^*_{t-j}\right)^2 \\
&\quad \leq E^*\left(\sum_{i=1}^{t-1}\sum_{j=1}^{t-1}\mathrm{tr}\left(\tilde{F}'\tilde{\Phi}^{i-1\prime}\tilde{\Phi}^{j-1}\tilde{F}\right)\left(\varepsilon^{*\prime}_{t-i}\varepsilon^*_{t-j}\right)\right)^2 \\
&\quad = \sum_{i=1}^{t-1}\sum_{j=1}^{t-1}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}\mathrm{tr}\left(\tilde{F}'\tilde{\Phi}^{i-1\prime}\tilde{\Phi}^{j-1}\tilde{F}\right)\mathrm{tr}\left(\tilde{F}'\tilde{\Phi}^{k-1\prime}\tilde{\Phi}^{l-1}\tilde{F}\right)E^*\left(\varepsilon^{*\prime}_{t-i}\varepsilon^*_{t-j}\varepsilon^{*\prime}_{t-l}\varepsilon^*_{t-k}\right).
\end{aligned}
$$

Since

$$
\begin{aligned}
E^*\left(\varepsilon^{*\prime}_{t-i}\varepsilon^*_{t-j}\varepsilon^{*\prime}_{t-l}\varepsilon^*_{t-k}\right) &= E^*\left(w_{t-i}w_{t-j}w_{t-k}w_{t-l}\right)\left(\tilde{\varepsilon}'_{c,t-i}\tilde{\varepsilon}_{c,t-j}\tilde{\varepsilon}'_{c,t-l}\tilde{\varepsilon}_{c,t-k}\right) \\
&\leq K\mathbb{I}\left(k=i,l=j\right)\left(\tilde{\varepsilon}'_{c,t-i}\tilde{\varepsilon}_{c,t-j}\right)^2
\end{aligned}
$$

we have that

$$
\frac{1}{T^2}\sum_{t=1}^{T}E^*\left(\|Y^*_{1,t}\|^4\right) \leq K\frac{1}{T^2}\sum_{t=1}^{T}\|\tilde{\varepsilon}_{c,t}\|^4\sum_{i=1}^{t-1}\sum_{j=1}^{t-1}\mathrm{tr}\left(\tilde{F}'\tilde{\Phi}^{i-1\prime}\tilde{\Phi}^{j-1}\tilde{F}\right)^2\left(\tilde{\varepsilon}'_{c,t-i}\tilde{\varepsilon}_{c,t-j}\right)^2 = O_p\left(T^{-1}\right),
$$

since ε_t has finite 8^+ moments and $\mathrm{tr}\left(\tilde{F}'\tilde{\Phi}^{i-1\prime}\tilde{\Phi}^{j-1}\tilde{F}\right)^2$ is exponentially decaying, see above.

Part (ii). To show this part if suffices to notice that, since $\mathbb{X}^{\#*}_{\beta 0} = \mathbb{X}^{\#}_{\beta 0}$ due to the initialization of the bootstrap sample, we have that (conditionally on the original sample)

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}Y^*_{0,t} &= T^{-1/2}\sum_{t=1}^{T}\mathrm{vec}\left(\tilde{\Phi}^{t-1}\mathbb{X}^{\#*}_{\beta 0}\varepsilon^{*\prime}_t\right) = T^{-1/2}\sum_{t=1}^{T}\mathrm{vec}\left(\tilde{\Phi}^{t-1}\mathbb{X}^{\#}_{\beta 0}\varepsilon^{*\prime}_t\right) \\
&= T^{-1/2}\sum_{t=1}^{T}\left(\varepsilon^*_t\otimes\tilde{\Phi}^{t-1}\mathbb{X}^{\#}_{\beta 0}\right)
\end{aligned}
$$

has variance $\frac{1}{T^2}\sum_{t=1}^{T}\left(E^*(\varepsilon^*_t\varepsilon^{*\prime}_t)\otimes\left(\tilde{\Phi}^{t-1}\mathbb{X}^{\#}_{\beta 0}\mathbb{X}^{\#\prime}_{\beta 0}\tilde{\Phi}^{t-1\prime}\right)\right) = \frac{1}{T^2}\sum_{t=1}^{T}\left(\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}'_{c,t}\otimes\tilde{\Phi}^{t-1}\mathbb{X}^{\#}_{\beta 0}\mathbb{X}^{\#\prime}_{\beta 0}\tilde{\Phi}^{t-1\prime}\right)$ $\xrightarrow{p^*}_p 0$, under the stated assumptions. This result completes the proof of (25), since, from the definition of $\tilde{\beta}^{\#\prime}S^*_{1\varepsilon}$, it implies $T^{1/2}\tilde{\beta}^{\#\prime}S^*_{1\varepsilon} \xrightarrow{w}_p N(0,\Omega^\dagger)$, with $\Omega^\dagger := \left[I : -\bar{M}_{\beta 2}\bar{M}_{22}^{-1}\right]V^\dagger\left[I : -\bar{M}_{\beta 2}\bar{M}_{22}^{-1}\right]'$. □

Proof of Theorem 4. It follows using the same steps as in the proof of Theorem 1, using the wild bootstrap results of Lemma 5. □

Proof of Corollary 2. Theorems 4 and 5 imply that, uniformly in probability, the (conditional) cdf of the bootstrap statistic LR^*_T satisfies $G^*_T(\cdot) \to F(\cdot)$, with F the cdf of the limiting

distribution of LR_T. This implies that, under the null hypothesis, P_T^* converges weakly to $U[0,1]$, see Hansen (2000, proof of Theorem 5). □

Proof of Theorem 5. It follows using the same steps as in the proof of Theorem 2, using the wild bootstrap results of Lemma 5 and the fact that, under the assumptions of the Theorem, we have that, as $T \to \infty$, $\hat{V}_T^*(u) := T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \hat{\varepsilon}_t^* \hat{\varepsilon}_t^{*\prime} = T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \varepsilon_t^* \varepsilon_t^{*\prime} + o_p^*(1) \xrightarrow{p^*}_p \int_0^u \Sigma(s)ds =: V(u)$, uniformly in $u \in [0,1]$. This result can be proved by noticing that $\frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor} \varepsilon_t^* \varepsilon_t^{*\prime} = \frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \left(w_t^2 - 1\right) + \frac{1}{T}\sum_{t=1}^{\lfloor Tu \rfloor} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \to_p V(u)$, as, by Lemma 3, $T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}' \to_p V(u)$ and, conditionally on the data, $\tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\left(w_t^2 - 1\right)$ is an independent sequence and $T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \tilde{\varepsilon}_{c,t}\tilde{\varepsilon}_{c,t}'\left(w_t^2 - 1\right)$ converges to 0 by standard arguments. □

References

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

Billingsley, P. (1968), *Convergence of Probability Measures*. John Wiley, New York.

Boswijk, H.P. (1995), Identifiability of cointegrated systems. Tinbergen Institute Discussion Paper # 95-78.

Boswijk, H.P. (2005), Adaptive testing for a unit root with nonstationary volatility, UvA-Econometrics Discussion Paper 2005/07.

Boswijk, H.P. and J.A. Doornik (2004), Identifying, estimating and testing restricted cointegrated systems: An overview, *Statistica Neerlandica* 58, 440–465.

Brockwell, P.J. and R.A. Davis (1991), *Time Series: Theory and Methods* (2nd Edition). Springer, New York.

Brown, B.M. (1971), Martingale central limit theorems, *Annals of Mathematical Statistics* 42, 59–66.

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

Cavaliere, G., H.B. Nielsen and A. Rahbek (2015), Bootstrap testing of hypotheses on cointegrating relations in VAR models, *Econometrica*, forthcoming.

Cavaliere, G., A. Rahbek and A.M.R. Taylor (2010a), co-integration rank testing under conditional heteroskedasticity, *Econometric Theory* 26, 1719–1760.

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

Cavaliere, G., A. Rahbek and A.M.R. Taylor (2012), Bootstrap determination of the co-integration rank in vector autoregressive models, *Econometrica* 80, 1721–1740.

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

Cavaliere, G., A.M.R. Taylor and C. Trenkler (2015), Bootstrap Co-integration Rank Testing: The Effect of Bias-Correcting Parameter Estimates, *Oxford Bullettin of Economics and Statistics*, forthcoming.

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

Cavaliere G. and A.M.R. Taylor (2007), Heteroskedastic time series with a unit root, *Econometric Theory* 25, 1228–1276.

Chung, H. and J.Y. Park (2007), Nonstationary nonlinear heteroskedasticity in regression, *Journal of Econometrics* 137, 230–259.

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

Davidson, R. and Flachaire, E. (2008). The wild bootstrap, tamed at last, *Journal of Econometrics* 146, 162–169.

Davidson, R. and MacKinnon, J.G. (2000). Bootstrap tests: How many bootstraps?, *Econometric Reviews* 19, 55–68.

Deo, R.S. (2000). Spectral tests of the martingale hypothesis under conditional heteroskedasticity, *Journal of Econometrics* 99, 291–315.

Diebold, F.X. and C. Li (2006), Forecasting the term structure of government bond yields, *Journal of Econometrics* 130, 337–364.

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

Fama, E. and R. Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.

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

Godfrey, L.G. and A.R. Tremayne (2005), The wild bootstrap and heteroskedasticity-robust tests for serial correlation in dynamic regression models, *Computational Statistics & Data Analysis* 49, 377–395.

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

Hall, P. and C.C. Heyde (1980), *Martingale Limit Theory and its Application*. Academic Press, New York.

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), Sample splitting and threshold estimation, *Econometrica* 68, 575–603.

He, C. and T. Teräsvirta (1999), Properties of moments of a family of GARCH processes,

Journal of Econometrics 92, 173–192.

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. (1995), Identifying restrictions of linear equations — with applications to simultaneous equations and cointegration, *Journal of Econometrics* 69, 111–132.

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

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.

Kristensen, D. and A. Rahbek (2010), Testing and inference in nonlinear cointegrating vector error-correction models, *Econometric Theory*, forthcoming.

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.

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

Omtzigt, P. and S. Fachin (2006), The size and power of bootstrap and Bartlett-corrected tests of hypotheses on the cointegrating vectors, *Econometric Reviews* 25, 41–60.

Park, J.Y. (2002), Nonstationary nonlinear heteroskedasticity, *Journal of Econometrics* 110, 383–415.

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

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

Table 1: Monte Carlo rejection frequencies for $p = 2$

GARCH	$\Sigma_t^{(i)}$		$H_{0\beta} : \beta_2 = 0$						$H_{0\alpha} : \alpha_2 = 0$						$H_0 = H_{0\beta} \cap H_{0\alpha}$					
			PLR			Wald			PLR			Wald			PLR			Wald		
N/Y	i	T	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS
		100	11.1	5.6	5.3	18.8	5.6	6.3	8.9	5.8	5.9	15.0	6.2	6.9	9.9	5.3	5.1	24.5	5.6	7.0
	1	200	7.2	4.8	4.9	11.2	4.8	5.0	6.3	5.2	5.3	9.1	5.2	5.5	6.9	4.9	4.9	12.8	4.8	5.5
		400	5.9	4.8	4.6	7.7	4.9	4.9	5.8	5.2	5.2	7.1	5.2	5.3	5.9	5.0	5.0	8.7	4.9	5.3
		100	*18.7*	*11.2*	5.9	21.8	7.3	6.0	*16.2*	*11.9*	6.0	17.4	7.7	7.0	*19.7*	*12.5*	5.5	31.9	10.5	8.1
N	2	200	*12.6*	*8.9*	4.9	13.1	6.3	5.5	*12.3*	*10.8*	5.5	11.3	7.0	6.1	*14.9*	*11.6*	4.9	17.0	7.5	6.0
		400	*10.8*	*9.3*	5.0	9.0	5.5	5.3	*11.1*	*10.2*	4.9	7.2	5.5	5.1	*13.3*	*11.7*	4.9	10.1	6.1	5.4
		100	*20.3*	*12.6*	5.4	*22.1*	*7.6*	6.1	*16.0*	*11.6*	5.6	16.9	7.3	6.8	*20.4*	*13.4*	5.0	*31.4*	*10.3*	8.1
	3	200	*14.1*	*10.8*	5.0	*13.2*	*6.6*	5.5	*12.6*	*10.7*	5.1	11.6	6.8	5.9	*16.1*	*12.7*	4.9	*16.7*	*7.6*	5.9
		400	*11.6*	*10.1*	5.2	*8.5*	*5.5*	5.3	*11.3*	*10.5*	4.8	7.4	5.4	5.1	*13.9*	*12.4*	4.9	*9.8*	*6.1*	5.6
		100	11.4	5.8	5.5	18.7	5.8	6.2	*9.2*	*5.9*	5.9	15.1	6.3	7.2	*10.4*	*5.6*	5.5	25.0	6.0	7.2
	1	200	7.8	5.2	5.1	11.4	4.9	5.3	*6.6*	*5.5*	5.2	9.0	5.3	5.5	*7.4*	*5.3*	5.1	13.0	5.0	5.5
		400	6.1	5.0	4.8	7.6	4.8	4.9	*6.0*	*5.3*	5.2	6.8	5.0	5.2	*6.3*	*5.2*	5.0	8.6	5.0	5.3
		100	*19.4*	*11.9*	6.0	21.9	7.3	6.2	*16.8*	*12.7*	6.2	17.9	8.2	7.1	*20.3*	*13.2*	5.6	32.3	11.0	8.4
Y	2	200	*13.0*	*9.7*	5.0	13.2	6.3	5.5	*12.5*	*11.0*	5.3	11.3	6.8	6.0	*15.4*	*12.2*	5.2	17.1	7.5	6.0
		400	*10.9*	*9.1*	5.2	8.8	5.5	5.2	*11.4*	*10.6*	4.9	7.5	5.7	5.2	*13.6*	*12.1*	5.2	10.1	6.1	5.5
		100	*21.0*	*13.2*	5.7	*22.2*	*7.7*	6.2	*16.7*	*12.3*	5.8	17.0	7.9	7.2	*21.7*	*14.3*	5.2	*32.1*	*11.1*	8.2
	3	200	*14.6*	*11.1*	4.7	*13.3*	*6.9*	5.7	*12.8*	*11.0*	5.1	11.3	6.9	5.8	*16.6*	*13.1*	5.0	*16.8*	*7.9*	6.2
		400	*12.0*	*10.4*	5.3	*8.8*	*5.7*	5.6	*11.5*	*10.7*	5.1	7.7	5.5	5.2	*15.0*	*13.1*	5.1	*9.8*	*5.9*	5.5

Notes: p denotes the dimension of the system. $\beta_2 = (0_{p-1} : I_{p-1})\beta$ and $\alpha_2 = (0_{p-1} : I_{p-1})\alpha$, so $H_{0\beta}$ ($H_{0\alpha}$) restricts the last $p-1$ components of β (α) to zero. PLR and Wald refer to the pseudo-likelihood ratio and PML-based Wald tests. The columns labelled Asy, IBS and WBS refer to rejection frequencies using χ^2, i.i.d. bootstrap and wild bootstrap critical values, respectively, all at the 5% nominal level. Entries in italics correspond to tests which are not asymptotically valid. T is the sample size; the three choices for $\Sigma_t^{(i)}, i = 1, 2, 3$ are described in Section 5. The column GARCH N/Y indicates whether (Y) or not (N) the errors were generated from univariate GARCH processes with parameter values indicated in Section 5.

Table 2: Monte Carlo rejection frequencies for $p = 3$

GARCH	$\Sigma_t^{(i)}$		$H_{0\beta} : \beta_2 = 0$						$H_{0\alpha} : \alpha_2 = 0$						$H_0 = H_{0\beta} \cap H_{0\alpha}$					
			PLR			Wald			PLR			Wald			PLR			Wald		
N/Y	i	T	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS
		100	17.5	5.8	5.7	32.9	6.0	6.5	13.4	6.4	6.2	24.0	6.2	8.6	14.6	5.2	4.9	49.0	5.3	8.6
	1	200	10.0	5.3	5.4	19.2	5.6	6.5	8.0	5.4	5.4	13.9	5.7	6.8	8.9	5.2	5.2	24.4	5.5	7.0
		400	7.3	5.3	5.4	11.4	5.4	5.9	6.0	4.9	5.0	8.3	5.2	5.4	6.9	5.1	5.1	13.1	5.1	6.0
		100	*31.4*	*15.0*	6.5	37.3	7.7	6.7	*26.2*	*16.7*	6.8	26.6	8.8	8.6	*31.9*	*16.2*	5.6	59.9	12.2	10.4
N	2	200	*17.8*	*11.4*	5.7	21.8	7.2	6.4	*17.9*	*14.1*	5.9	17.6	8.4	7.6	*20.9*	*15.0*	5.5	31.0	9.6	7.2
		400	*14.0*	*10.7*	5.3	13.1	6.3	5.6	*14.8*	*12.8*	5.3	9.8	6.1	5.7	*17.8*	*14.7*	5.1	16.0	7.0	6.2
		100	*41.5*	*23.3*	6.7	*39.4*	*10.6*	6.9	*31.6*	*21.7*	7.2	26.7	9.7	8.2	*41.4*	*24.5*	5.7	*63.5*	*16.3*	10.8
	3	200	*24.9*	*17.0*	5.6	*23.4*	*8.0*	5.9	*20.2*	*16.5*	5.6	18.2	9.6	8.0	*27.8*	*20.4*	5.3	*33.3*	*12.1*	7.4
		400	*19.3*	*15.7*	5.4	*13.5*	*6.7*	6.1	*16.4*	*14.7*	5.1	10.3	6.6	5.8	*23.4*	*19.8*	5.0	*16.9*	*7.7*	6.3
		100	18.3	6.4	6.1	33.0	6.4	7.0	*13.5*	*6.4*	6.3	23.6	6.2	8.2	*14.9*	*5.7*	5.2	48.7	5.8	9.4
	1	200	10.6	5.8	5.5	19.5	5.9	6.4	*8.4*	*5.9*	5.6	14.0	5.9	6.9	*9.4*	*5.8*	5.4	24.7	5.5	6.8
		400	7.4	5.2	4.9	11.4	5.2	5.6	*6.3*	*5.3*	5.0	8.6	5.4	5.7	*7.0*	*5.5*	5.1	13.0	5.0	5.8
		100	*31.9*	*15.8*	6.6	36.3	8.5	6.9	*27.6*	*17.7*	6.5	27.5	9.4	8.7	*32.7*	*17.4*	5.4	60.2	13.8	11.2
Y	2	200	*19.1*	*12.4*	5.8	22.5	7.4	6.2	*17.8*	*14.0*	5.4	17.6	8.2	7.3	*22.0*	*15.7*	5.4	31.8	10.1	7.4
		400	*14.4*	*11.0*	5.4	13.1	6.1	5.8	*15.1*	*13.6*	5.2	10.7	6.8	6.4	*18.4*	*15.3*	5.0	16.5	7.3	6.3
		100	*40.7*	*23.9*	6.7	*39.2*	*10.8*	6.8	*31.9*	*22.3*	6.7	27.1	10.1	8.4	*41.9*	*25.5*	5.5	*63.0*	*17.6*	11.4
	3	200	*25.4*	*17.5*	5.5	*24.0*	*8.3*	6.2	*20.3*	*16.3*	5.2	18.1	8.9	7.2	*28.0*	*20.9*	5.0	*34.2*	*12.5*	7.4
		400	*19.8*	*15.9*	5.3	*13.9*	*6.6*	5.7	*16.7*	*15.0*	5.2	11.2	7.4	6.1	*23.7*	*20.1*	4.7	*17.5*	*7.8*	5.9

Notes: See Notes to Table 1.

Table 3: Monte Carlo rejection frequencies for $p = 4$

GARCH	$\Sigma_t^{(i)}$		$H_{0\beta} : \beta_2 = 0$						$H_{0\alpha} : \alpha_2 = 0$						$H_0 = H_{0\beta} \cap H_{0\alpha}$					
			PLR			Wald			PLR			Wald			PLR			Wald		
N/Y	i	T	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS	Asy	IBS	WBS
		100	28.7	6.7	6.3	48.6	6.6	7.4	20.3	7.6	7.8	34.2	6.4	9.8	22.6	5.7	5.1	73.0	5.5	11.3
	1	200	13.1	5.3	5.3	28.9	5.5	6.5	9.3	5.3	5.2	20.1	5.9	7.5	10.9	5.0	5.0	38.4	4.9	7.0
		400	8.5	5.1	5.3	15.8	5.5	6.1	6.9	5.3	5.5	11.3	5.7	6.4	7.7	5.3	5.2	19.3	5.5	6.6
		100	*46.9*	*20.2*	6.5	50.4	8.9	7.3	*40.3*	*23.7*	7.8	36.6	9.9	9.3	*46.9*	*21.0*	4.8	81.7	14.9	12.9
N	2	200	*24.5*	*13.3*	5.5	33.6	8.0	6.5	*22.4*	*16.0*	5.5	24.4	8.6	7.9	*27.6*	*16.8*	5.0	48.9	11.5	8.2
		400	*16.2*	*11.7*	5.6	18.0	6.9	6.4	*17.9*	*15.1*	5.8	13.2	7.2	6.7	*21.8*	*16.8*	5.1	23.9	8.0	6.9
		100	*61.6*	*36.3*	7.4	*53.1*	*12.3*	6.8	*48.7*	*34.8*	8.5	34.8	11.8	8.5	*62.7*	*37.2*	5.2	*85.8*	*25.0*	13.5
	3	200	*37.8*	*25.0*	5.7	*36.8*	*10.7*	6.9	*29.2*	*22.3*	5.4	26.4	11.2	8.5	*41.1*	*28.8*	4.8	*55.4*	*17.9*	8.8
		400	*26.2*	*19.9*	5.2	*19.3*	*7.8*	6.1	*21.8*	*18.8*	5.1	14.4	8.0	6.6	*32.0*	*25.9*	5.1	*26.6*	*9.8*	6.6
		100	28.3	7.5	6.8	47.3	6.8	7.5	*20.8*	*8.6*	7.7	33.2	6.9	10.2	*23.0*	*6.3*	5.4	71.8	5.8	11.1
	1	200	13.5	5.5	5.1	29.2	5.5	6.7	*10.0*	*6.1*	5.6	20.1	6.2	7.5	*11.0*	*5.1*	4.7	39.0	5.3	7.6
		400	7.9	4.8	4.6	15.2	4.8	5.5	*7.1*	*5.2*	5.0	10.5	5.5	5.7	*7.5*	*4.9*	4.8	18.4	4.6	5.7
		100	*46.3*	*20.7*	6.8	49.8	8.9	7.1	*39.6*	*24.3*	7.8	36.1	10.3	9.3	*46.6*	*21.7*	4.8	81.1	15.5	12.7
Y	2	200	*25.9*	*13.8*	5.6	33.9	8.2	6.9	*23.8*	*17.2*	5.6	24.8	9.5	8.7	*29.0*	*18.4*	4.9	49.2	12.0	8.3
		400	*16.4*	*11.8*	4.8	17.3	6.3	5.6	*17.9*	*15.3*	5.1	13.1	7.0	6.3	*21.7*	*16.5*	4.6	23.3	7.4	5.9
		100	*61.4*	*37.2*	7.2	*52.6*	*11.9*	6.4	*49.0*	*35.6*	8.6	35.0	12.1	8.8	*62.4*	*37.9*	5.3	*85.8*	*25.8*	13.2
	3	200	*39.1*	*26.3*	5.8	*37.7*	*10.6*	7.0	*30.0*	*23.2*	6.0	26.7	11.7	9.0	*42.5*	*30.7*	5.1	*55.7*	*18.7*	9.5
		400	*26.0*	*20.6*	5.2	*19.1*	*7.4*	6.2	*21.7*	*18.7*	5.1	14.8	8.1	6.7	*31.5*	*26.2*	4.6	*25.8*	*10.2*	6.7

Notes: See Notes to Table 1.

Table 4: Trace test statistics (Q_r) for co-integration rank (r), with asymptotic (p_{Asy}), standard bootstrap (p_{IBS}) and wild bootstrap (p_{WBS}) p-values.

r	Q_r	p_{Asy}	p_{IBS}	p_{WBS}
0	191.6	0.000	0.000	0.000
1	105.3	0.000	0.000	0.000
2	41.66	0.009	0.007	0.087
3	17.44	0.117	0.127	0.286
4	2.664	0.645	0.658	0.795

Figure 1: Time series (ei) and variance profiles (vi) of VAR(2) residuals, $i = 1, 3, 5$.