

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Determining the cointegration rank in heteroskedastic VAR models of unknown order

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Cavaliere, G., De Angelis, L., Rahbek, A., Taylor, A.M. Robert (2018). Determining the cointegration rank in heteroskedastic VAR models of unknown order. *ECONOMETRIC THEORY*, 34(2), 349-382 [10.1017/S0266466616000335].

Availability:

This version is available at: <https://hdl.handle.net/11585/565788> since: 2021-02-22

Published:

DOI: <http://doi.org/10.1017/S0266466616000335>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript published in “Econometric Theory”; (Cambridge University Press), *Determining the Cointegration Rank in Heteroskedastic VAR Models of Unknown Order*, Cavaliere, Giuseppe; De Angelis, Luca; Rahbek, Anders; Robert Taylor, A M .

The final published version is available online at: <https://doi.org/10.1017/S0266466616000335>

© [2016]. This manuscript version is made available under the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) 4.0 International License
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please

Determining the Cointegration Rank in Heteroskedastic VAR Models of Unknown Order*

Giuseppe Cavaliere^a, Luca De Angelis^a, Anders Rahbek^b and A.M.Robert Taylor^c

^a Department of Statistical Sciences, University of Bologna

^b Department of Economics, University of Copenhagen

^c Essex Business School, University of Essex

August 18, 2016

Abstract

We investigate the asymptotic and finite sample properties of a number of methods for estimating the cointegration rank in integrated vector autoregressive systems of unknown autoregressive order driven by heteroskedastic shocks. We allow for both conditional and unconditional heteroskedasticity of a very general form. We establish the conditions required on the penalty functions such that standard information criterion-based methods, such as the Bayesian information criterion [BIC], when employed either sequentially or jointly, can be used to consistently estimate both the cointegration rank and the autoregressive lag order. In doing so we also correct errors which appear in the proofs provided for the consistency of information-based estimators in the homoskedastic case by Aznar and Salvador (2002). We also extend the corpus of available large sample theory for the conventional sequential approach of Johansen (1995) and the associated wild bootstrap implementation thereof of Cavaliere, Rahbek and Taylor (2014) to the case where the lag order is unknown. In particular, we show that these methods remain valid under heteroskedasticity and an unknown lag length provided the lag length is first chosen by a consistent method, again such as the BIC. The relative finite sample properties of the different methods discussed are investigated in a Monte Carlo simulation study. The two best performing methods in this study are a wild bootstrap implementation of the Johansen (1995) procedure implemented with BIC selection of the lag length and joint IC approach (cf. Phillips, 1996) which uses the BIC to jointly select the lag order and the cointegration rank.

Keywords: Cointegration rank; Information criteria; Wild bootstrap; Trace statistic; Lag length; Heteroskedasticity.

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

*We are grateful to the Editor, Peter Phillips, the Co-Editor, Michael Jansson, and two anonymous referees for their helpful comments on earlier versions of this paper. This work was supported by the Danish Council for Independent Research Sapere Aude | DFF Advanced Grant [grant number 12-124980], the Economic and Social Research Council [grant number ES/M01147X/1], and the Italian Ministry of Education, University and Research (MIUR), PRIN project “Multivariate statistical models for risk assessment”. Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. E-mail: rtaylor@essex.ac.uk

1 Introduction

Cointegration testing, and reduced rank testing more generally, are research areas where Richard Smith has made many important contributions (see, for example, Camba-Mendez, Kapetanios, Smith and Weale, 2003, Pesaran, Shin and Smith, 2000, 2001, and Robin and Smith, 2000), and this paper contributes further to this literature, focusing on methods for determining the cointegration rank of vector autoregressive [VAR] systems integrated of order one, $I(1)$. This is often done using a test-based sequential procedure based on the asymptotic (pseudo-) likelihood ratio [PLR] test of Johansen (1995). However, even where the autoregressive lag order is known, this approach can be significantly over-sized in finite samples, particularly so where the shocks contain strong serial correlation; see, in particular, Johansen, (2002) and the references therein. The PLR tests can also display significant upward size distortions in the presence of conditional heteroskedasticity (Lee and Tse, 1996; Cavaliere, Rahbek and Taylor, 2010a [CRT2010a, hereafter]) or non-stationary heteroskedasticity (Cavaliere, Rahbek and Taylor, 2010b [CRT2010b, hereafter]) which characterise many key macroeconomic and financial variables. As a result, assuming a known autoregressive lag length, Cavaliere, Rahbek and Taylor (2012, 2014) [CRT2012, CRT2014, hereafter] develop bootstrap implementations of the PLR tests; for earlier contributions see, *inter alia*, van Giersbergen (1996) and Swensen (2006).

There is a close link between (P)LR tests and information criteria [IC]-based methods for determining the cointegration rank (and, indeed, the autoregressive lag length); see Chao and Phillips (1999). The efficacy of IC-based methods for determining the cointegrating rank has also been explored by a number of authors; for early contributions see Phillips (1996), Phillips and McFarland (1997) and Lütkepohl and Poskitt (1998). For a known lag length, Cavaliere, De Angelis, Rahbek and Taylor (2015) [CDRT, hereafter] investigate selecting the cointegrating rank using the Bayesian IC [BIC] (Rissanen, 1978; Schwarz, 1978), the Akaike IC [AIC] (Akaike, 1974) and the Hannan-Quinn IC [HQC] (Hannan and Quinn, 1979). They find that BIC provides a useful complement to the bootstrap sequential procedures. Under conditions which rule out heteroskedasticity in the shocks, Aznar and Salvador (2002) and Kapetanios (2000) establish the requirements needed on the penalty function for consistent estimation of cointegration rank by IC-based methods and show that both BIC and HQC satisfy these conditions. Kapetanios (2004) establishes the inconsistency of the AIC-based approach, showing that its estimate of the rank displays a severe upward bias. For the case of a known lag length, CDRT extend these results to the case of heteroskedastic shocks, showing that the BIC and HQC approaches remain consistent.

Assuming a known lag length in the cointegrated VAR [CVAR] model is unreasonable in practice. The PLR tests can suffer from serious size distortions, even asymptotically, if a lag order smaller than the true order is used, while fitting unnecessary lags compromises finite sample power; see, among others, Boswijk and Franses (1992), Cheung and Lai (1993), Haug (1996) and Lütkepohl and Saikkonen (1999). Consequently, the practitioner needs to estimate both lag length and cointegration rank. This is often done using a two-step procedure, where the autoregressive lag length is first estimated and then subsequently employed in a PLR-based cointegration rank determination second step. The first

step will often involve using either an IC or a sequence of likelihood ratio lag specification tests. In the latter case these are often coupled with model misspecification tests on the residuals from the fitted model. Under the assumption of homoskedastic shocks, Lütkepohl and Saikkonen (1999) show that PLR tests where the lag length is estimated by a consistent IC, such as BIC or HQC, have the same asymptotic distributions as PLR tests computed with knowledge of the true lag order. Moreover, again for homoskedastic shocks, simulation results in Kascha and Trenkler (2011) suggest that an i.i.d. bootstrap implementation of this two-stage procedure works well.

Using IC-based approaches, the lag length and cointegration rank can be determined jointly or sequentially. Phillips (1996) considers the former using a procedure based on the posterior IC of Phillips and Ploberger (1996). Chao and Phillips (1999) establish the consistency of this estimator under the assumption that the CVAR is driven by homoskedastic shocks. Phillips and McFarland (1997) consider analogous BIC-based approaches. Fitting the correct autoregressive lag length has less influence, at least in large samples, on the behaviour of the resulting cointegration rank IC-based estimator than for the sequential PLR-based estimator. Cheng and Phillips (2009) propose a semi-parametric IC-based cointegrating rank estimator using a lag length of one in the CVAR. They show that the resulting BIC and HQC procedures consistently estimate the cointegration rank (under standard conditions on the penalty function) even when the shocks follow a general short memory homoskedastic process; Cheng and Phillips (2012) show that this result also holds under unconditional heteroskedasticity of the form considered in CRT2010b. However, we show that using an incorrect lag length can seriously reduce the finite-sample efficacy of IC-based methods.

Important gaps therefore exist relating to the theory of cointegration rank determination in the context of CVARs of unknown autoregressive lag length driven by potentially heteroskedastic shocks. A major contribution of this paper is to fill those gaps. For shocks which can display stationary conditional heteroskedasticity and/or non-stationary unconditional heteroskedasticity, we provide sufficient conditions for IC-based procedures to be consistent, when the lag length is unknown. We discuss the cases where the lag length and cointegration rank are jointly determined and where a sequential procedure is employed. As part of establishing large sample results for the latter, we also demonstrate (correcting errors in the corresponding proof provided by Aznar and Salvador, 2002) that the cointegration rank can be consistently estimated (using standard penalty functions) under heteroskedasticity regardless of the lag length used. This result implies that the consistency results given in Cheng and Phillips (2009, 2012) also apply for any choice of the lag length and confirms their conjecture that the result also holds under conditionally heteroskedastic shocks. In the same heteroskedastic setting we also extend the available large sample theory for the standard and wild bootstrap PLR tests to cover the case of an unknown lag order in the CVAR, establishing the conditions under which using an IC-based method to first select the lag length prior to computing the (bootstrap) PLR tests leads to the same large sample properties as hold under a known lag length.

The remainder of the paper is organised as follows. Section 2 details our reference heteroskedastic CVAR model. Section 3 outlines the various cointegration rank determination methods which we consider and section 4 analyses the large sample properties of these methods. The results from a

large number of Monte Carlo experiments are discussed in section 5 in order to provide a detailed comparison of the relative finite sample performances of the various methods. Section 6 concludes. All proofs are contained in the Appendix. The tabulated results from the Monte Carlo simulation experiments are reported in the accompanying online Supplementary Appendix. Additional simulation results and an empirical application to the term structure of interest rates in the US can also be found in the accompanying working paper, Cavaliere, De Angelis, Rahbek and Taylor (2014).

2 The Heteroskedastic Cointegrated VAR Model

Consider the p -dimensional process $\{X_t\}$ which satisfies the k -th order reduced rank VAR model:

$$\Delta X_t = \alpha \beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \alpha \rho' D_t + \phi d_t + \varepsilon_t, \quad t = 1, \dots, T \quad (2.1)$$

where $X_t := (X_{1t}, \dots, X_{pt})'$ and the initial values, X_{1-k}, \dots, X_0 , are taken to be fixed in the statistical analysis. Let k_0 denote the true value of the autoregressive lag length k in (2.1). In the context of (2.1) we assume that the standard ‘I(1, r_0) conditions’ hold, where $r_0 \in \{0, \dots, p\}$ denotes the cointegration rank of the system (see also CRT2012); that is, the characteristic polynomial associated with (2.1) has $p - r_0$ roots equal to 1 with all other roots lying outside the unit circle, and where α and β have full column rank r_0 .

The innovation process $\varepsilon_t := (\varepsilon_{1t}, \dots, \varepsilon_{pt})'$ is assumed to satisfy the general assumption introduced by Boswijk, Cavaliere, Rahbek and Taylor (2016):

Assumption \mathcal{H} The innovations $\{\varepsilon_t\}$ are defined as $\varepsilon_t := \sigma_t z_t$, where σ_t is non-stochastic and satisfies $\sigma_t := \sigma(t/T)$ for all $t = 1, \dots, T$, where $\sigma(\cdot) \in \mathcal{D}_{\mathbb{R}^{p \times p}}[0, 1]$, with $\mathcal{D}_{\mathbb{R}^{m \times n}}[0, 1]$ used to denote the space of $m \times n$ matrices of càdlàg functions on $[0, 1]$ equipped with the Skorokhod metric, and where $\Sigma(u) := \sigma(u) \sigma(u)'$ is assumed to be positive definite for all $u \in [0, 1]$; z_t forms a martingale difference sequence with respect to the filtration \mathcal{F}_t with conditional variance matrix $h_t := E(z_t z_t' | \mathcal{F}_{t-1})$, satisfying $\sup_t E \|z_t\|^{4r} < \infty$, for some $r > 1$, and $T^{-1} \sum_{t=1}^T h_t \xrightarrow{p} E(z_t z_t') = I_p$, where: I_k denotes the $k \times k$ identity matrix, \xrightarrow{p} denotes convergence in probability as $T \rightarrow \infty$, and for any vector, x , $\|x\|$ denotes the usual Euclidean norm, $\|x\| := (x'x)^{1/2}$.

Remark 1: Assumption \mathcal{H} implies that ε_t is a p -dimensional vector martingale difference sequence with respect 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 $\Sigma_t := E(\varepsilon_t \varepsilon_t') = \sigma_t \sigma_t' > 0$. As such, Assumption \mathcal{H} combines the assumptions of CRT2010a and CRT2010b 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, and hence only conditional heteroskedasticity) and $h_t = I_p$ (so that $\Sigma_{t|t-1} = \Sigma_t = \Sigma(t/T)$, allowing only unconditional non-stationary volatility). As discussed in CRT2010b, the latter 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. The former allows for a wide range of conditionally heteroskedastic processes including stochastic volatility and generalised autoregressive-conditional heteroskedasticity (GARCH) processes; see CRT2010a and Boswijk *et al.* (2016) for further discussion.

The deterministic variables in (2.1) are taken to satisfy one of the following cases (see, e.g., Johansen, 1995): (i) $D_t = 0$, $d_t = 0$ (no deterministic); (ii) $D_t = 1$, $d_t = 0$ (restricted constant); or (iii) $D_t = t$, $d_t = 1$ (restricted linear trend). We will take the form of the deterministic component (but not the values of the parameters which characterise it) to be given. For the case where ε_t is i.i.d. with mean vector zero, variance matrix Σ and finite fourth moments, Aznar and Salvador (2002) also discuss a sequential strategy where as a first step the lag order and cointegration rank in (2.1) are determined by a joint information criterion of the form outlined in section 3.1 below, but then undertake additional steps to determine the form of the deterministic component. While we do not consider the choice of the form of the deterministic component in this paper, the strategy outlined by Aznar and Salvador (2002) could also be extended straightforwardly to the present context.

3 Methods for the Determining the Cointegration Rank when the Lag Length is Unknown

In this section we briefly discuss available methods for cointegration rank determination in the context of (2.1) where the true autoregressive lag, k_0 , is unknown to the practitioner. In order to motivate why the methods we outline in this section are important in practice, we first illustrate, by means of a small Monte Carlo exercise, the impact that the value of the lag length used has on the ability of information-based and PLR based procedures to determine the cointegration rank. In particular, we take as our simulation DGP the VAR model of dimension $p = 4$, defined in (5.1) of section 5, whose true autoregressive lag length, k_0 , is equal to two. Three cases are considered for the true cointegrating rank: $r_0 = 0$, $r_0 = 1$, and $r_0 = 2$. To keep the DGP simple at this stage we take the innovation process, ε_t , to be i.i.d. Gaussian. Table 1, which can be found in the accompanying online Supplementary Appendix, reports the frequencies with which the wild bootstrap PLR based procedure of CRT2014 and a BIC-based procedure, which minimises the quantity $\text{BIC}(k, r)$ defined in (3.5) below over $r = 0, 1, \dots, p$, select the true cointegration rank, in each case fixing the lag length used at each of $k = 1, \dots, 6$. In the context of the simulation DGP in (5.1) we consider samples of sizes $T = 100$ and 400 and set the lag parameter matrix $\Gamma_1 := \gamma I_4$ with $\gamma \in \{0.2, 0.5, 0.8\}$. The results are obtained using 10,000 Monte Carlo replications and 399 bootstrap replications.

The results in Table 1 show quite clearly that the value of the autoregressive lag length used can have a sizeable influence on the efficacy of these procedures in practice. As might be expected, for both the PLR- and BIC-based procedures, the best results are obtained when the lag length used is equal to the true lag length, $k_0 = 2$. It is also interesting to observe that both under-specifying and over-specifying the lag length can significantly reduce the efficacy of both procedures. In the case

of the BIC-based procedure, the performance is worst when the lag length is either under-specified ($k = 1$) or heavily over-specified ($k = 5, 6$). As an example, while the BIC procedure computed for the correct lag length ($k = 2$) selects the correct rank of zero 96.0% of the time when $T = 100$ and $\gamma = 0.8$ this drops to 80.0% when a lag length of $k = 6$ is used and collapses to as little as 1.4% when a lag length of $k = 1$ is used; notice that setting $k = 1$ yields the semi-parametric BIC-based rank estimator proposed in Cheng and Phillips (2009, 2012). Indeed, at least as far as the results in Table 1 are concerned, setting $k = 1$ yields the least reliable estimator. We will investigate this further in section 5.

The simulation exercise above demonstrates that in practice one will need a good estimate of the lag length in order to obtain a reliable procedure for estimating the cointegration rank. We shall explore two methods for doing this. The first approach, which will be detailed in section 3.1, is to jointly minimise an information criterion over both the cointegration rank and the autoregressive lag length. The second approach, which will be detailed in section 3.2, and is regularly used in practice in connection with the PLR-based procedure, is to use a two-step approach. Here one first selects the lag length, often by either an information criterion or by statistical specification tests on the lag length (as noted in section 1, it is common practice to also conduct various misspecification tests on the residuals from the fitted model), without imposing a reduced rank structure. One then determines the cointegration rank (either by a sequential approach based on PLR or bootstrap PLR tests or by using an information criterion) based on fixing the lag length at the value estimated in the first step. The two-step procedure clearly affords greater flexibility in that different methods can be used for determining the lag order and/or the cointegration rank; for example, the lag length used in a PLR-based sequential procedure could be selected in the first step by the BIC.¹

In what follows, we will use the generic notation $\hat{k}_{IC}(r)$ to denote an estimator of the autoregressive lag length, k_0 , obtained via an information criterion when the cointegration rank is set to be r ; that is, $\hat{k}_{IC}(r) := \arg \min_{k=1, \dots, K} IC(k, r)$, where $IC(k, r)$ is used to denote the value taken by a given information criterion when imposing rank r and lag length k on the estimation of (2.1). Similarly, $\hat{r}_{IC}(k)$ will be used to denote the estimator of the cointegration rank, r_0 , obtained by an information criterion imposing a lag order equal to k ; that is, $\hat{r}_{IC}(k) := \arg \min_{r=0, \dots, p} IC(k, r)$. Finally, we use $\hat{r}_{PLR}(k)$ and $\hat{r}_{PLR}^*(k)$ to denote the rank estimator using, respectively, the sequential PLR-based and wild bootstrap PLR-based procedures, in each case for a lag length k .

3.1 Joint Determination of the Lag Length and Cointegration Rank

Following Phillips and McFarland (1997) and Aznar and Salvador (2002), the autoregressive lag order and the cointegration rank can be jointly determined by (jointly) minimising an information criterion

¹Notice that we only consider the case where the lag length is estimated in the first step. One could consider a reversed procedure where one first estimates the cointegration rank for some chosen value of the lag length, then in the second step estimates the lag length conditional on that estimate of the rank. The simulation results in Table 1 clearly suggest that this approach can be highly unreliable in practice. We will therefore not discuss this approach further.

of the following generic form

$$\text{IC}(k, r) := -2\ell_T(k, r) + p_T(k, r) \quad (3.1)$$

over both all possible lag lengths, $k = 1, \dots, K$, where K denotes a given maximum lag order which, unless otherwise stated, we assume in the following to be at least as large as the true lag order, k_0 , and over all possible cointegration ranks, $r = 0, \dots, p$. In the context of (3.1), $\ell_T(k, r)$ denotes the maximised pseudo log-likelihood function associated with (2.1) (under the assumption that ε_t is i.i.d. Gaussian with mean vector zero and constant variance matrix Σ) under lag order k and cointegration rank r , and where $p_T(k, r) := c_T\pi(k, r)$ denotes the penalty function which depends on the number of parameters, $\pi(k, r)$, and on the term c_T which may depend on the sample size T .

Up to a constant term which does not depend on k or r , we therefore have from Johansen (1995) that $\ell_T(k, r) = -\frac{T}{2} \log |\hat{\Sigma}_{k,r}|$, where $\hat{\Sigma}_{k,r}$ is the residual covariance matrix estimate from the conventional reduced rank regression estimation of (2.1) under rank r for lag length k , and the notation $|A|$ is used to denote the determinant of the matrix A . As demonstrated in Johansen (1995), $|\hat{\Sigma}_{k,r}| := |S_{00}^{(k)}| \prod_{i=1}^r (1 - \hat{\lambda}_i^{(k)})$, and so $\ell_T(k, r) = -\frac{T}{2} \log |S_{00}^{(k)}| - \frac{T}{2} \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(k)})$, where $\hat{\lambda}_1^{(k)} > \dots > \hat{\lambda}_p^{(k)}$ are the p largest solutions to the eigenvalue problem,

$$\left| \lambda S_{11}^{(k)} - S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \right| = 0, \quad (3.2)$$

where $S_{ij}^{(k)} := T^{-1} \sum_{t=1}^T R_{it}^{(k)} R_{jt}^{(k)'}$, $i, j = 0, 1$, with $R_{0t}^{(k)}$ and $R_{1t}^{(k)}$ respectively denoting ΔX_t and $(X'_{t-1}, D_t)'$, corrected (by OLS) for $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$ and d_t .

On substitution into (3.1), we therefore have that

$$\text{IC}(k, r) = T \log |S_{00}^{(k)}| + T \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(k)}) + c_T \pi(k, r) \quad (3.3)$$

where: (i) in the case of no deterministic component ($D_t = 0$, $d_t = 0$ in 2.1), $\pi(k, r) = r(2p - r) + p(p + 1)/2 + p^2(k - 1)$; (ii) for the restricted constant case ($D_t = 1$, $d_t = 0$ in 2.1), $\pi(k, r) = r(2p - r + 1) + p(p + 1)/2 + p^2(k - 1)$, and (iii) for the case of a restricted trend ($D_t = 1$, $d_t = 1$ in 2.1), $\pi(k, r) = r(2p - r + 1) + p(p + 2)/2 + p^2(k - 1)$. Different values of the coefficient c_T yield different information criteria through the resulting penalty function, p_T . The most widely used are AIC, BIC, and HQC where $c_T = 2$, $\log T$, and $2 \log \log T$, respectively, thereby yielding:

$$\text{AIC}(k, r) := T \log |S_{00}^{(k)}| + T \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(k)}) + 2\pi(k, r), \quad (3.4)$$

$$\text{BIC}(k, r) := T \log |S_{00}^{(k)}| + T \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(k)}) + (\log T)\pi(k, r), \quad (3.5)$$

$$\text{HQC}(k, r) := T \log |S_{00}^{(k)}| + T \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(k)}) + 2(\log \log T)\pi(k, r). \quad (3.6)$$

Any information criterion as in (3.3) can be used to jointly determine k and r .² In particular, the cointegration rank and lag order estimator for the joint procedure, denoted by \tilde{k}_{IC} and \tilde{r}_{IC} , respectively, is then given, in generic form, by

$$(\tilde{k}_{\text{IC}}, \tilde{r}_{\text{IC}}) := \arg \min_{r=0, \dots, p; k=1, \dots, K} \text{IC}(k, r); \quad (3.7)$$

see Aznar and Salvador (2002).³ The specific estimators resulting from the BIC (first proposed in Phillips and McFarland, 1997), AIC and HQC penalty functions will be denoted $(\tilde{k}_{\text{BIC}}, \tilde{r}_{\text{BIC}})$, $(\tilde{k}_{\text{AIC}}, \tilde{r}_{\text{AIC}})$ and $(\tilde{k}_{\text{HQC}}, \tilde{r}_{\text{HQC}})$, respectively. We will use the same convention in what follows for obtaining the penalty-term specific versions of generic information criterion based terms.

Notice that the global joint minimiser $(\tilde{k}_{\text{IC}}, \tilde{r}_{\text{IC}})$ in (3.7) also solves

$$\begin{cases} \tilde{k}_{\text{IC}} = \arg \min_{k=1, \dots, K} \text{IC}(k, \hat{r}_{\text{IC}}(k)) \\ \tilde{r}_{\text{IC}} = \hat{r}_{\text{IC}}(\tilde{k}_{\text{IC}}). \end{cases} \quad (3.8)$$

This follows because, for any chosen $k = 1, \dots, K$, $\text{IC}(k, r)$ is minimised at $(k, \hat{r}_{\text{IC}}(k))$. Similarly, $(\tilde{k}_{\text{IC}}, \tilde{r}_{\text{IC}})$ solves:

$$\begin{cases} \tilde{r}_{\text{IC}} = \arg \min_{r=0, \dots, p} \text{IC}(\hat{k}_{\text{IC}}(r), r) \\ \tilde{k}_{\text{IC}} = \hat{k}_{\text{IC}}(\tilde{r}_{\text{IC}}). \end{cases} \quad (3.9)$$

In section 4.2 we will show that provided the penalty term coefficient, c_T , is such that $c_T/T + 1/c_T \rightarrow 0$ as $T \rightarrow \infty$ then so (3.7) will provide consistent estimates for k_0 and r_0 under Assumption \mathcal{H} .

3.2 Sequential Determination of the Lag Length and Cointegration Rank

Nielsen (2006), *inter alia*, shows that the autoregressive lag length in non-stationary VAR models can be consistently estimated from the levels of the data using either an information criterion or a sequence of likelihood ratio tests. The first of these methods is also discussed in section 5 of Lütkepohl and Saikkonen (1999) and will also be our focus in this paper. To that end, observe that this obtains as a special case of the general information criterion in (3.3) where we do not impose a reduced rank structure on $\Pi := \alpha\beta'$ in (2.1); that is, where $r = p$. Consequently the generic information criterion is given by

$$\text{IC}(k, p) = T \log \left| \hat{\Sigma}_{k,p} \right| + c_T \pi(k, p) \quad (3.10)$$

where $\hat{\Sigma}_{k,p}$ is an estimate of the residual covariance matrix obtained by estimating an unrestricted VAR model of order k in the levels of X_t , and where $\pi(k, p) := p(pk + i) + p(p + 1)/2$ with $i = 0$ when no deterministic component is involved, $i = 1$ in the case of restricted constant, and $i = 2$ for the

²Although not considered here, the use of IC to jointly select k and r could be extended to jointly determine these along with the form of the deterministic component in (2.1); cf. Phillips (1996). The authors are currently working on this development.

³Notice that the joint PIC-based estimator of Phillips (1996) and Chao and Phillips (1999) does not fall within the generic form given in (3.7). Nor indeed do estimators based on information criteria which adopt stochastic penalty terms, such as those proposed in Takeuchi (1976). We will therefore not consider such estimators further here.

restricted trend. The resulting information-based lag length estimator is then given, in generic form, by

$$\hat{k}_{\text{IC}}(p) := \arg \min_{k=1, \dots, K} \text{IC}(k, p). \quad (3.11)$$

Based on the lag order $\hat{k}_{\text{IC}}(p)$ of (3.11) obtained in the first step, we then determine the cointegration rank in the second step. This can be done in various ways, the two most obvious either by minimising an appropriate information criterion or by use of the sequential PLR test-based procedure of Johansen (1995) or the bootstrap equivalents of CRT2012 (i.i.d. re-sampling) and CRT2014 (wild bootstrap). We now outline some leading cases of how the two-step procedure could be implemented.

Two-Step Information Criteria-Based Procedure

Using the lag length $\hat{k}_{\text{IC}}(p)$ of (3.11) estimated in the first step, a sequential information criteria-based approach then estimates the cointegration rank in the second step via, again written in generic form,

$$\hat{r}_{\text{IC}}(\hat{k}_{\text{IC}}(p)) := \arg \min_{r=0, \dots, p} \text{IC}(\hat{k}_{\text{IC}}(p), r). \quad (3.12)$$

where

$$\text{IC}(\hat{k}_{\text{IC}}(p), r) = T \log |\hat{\Sigma}_{\hat{k}_{\text{IC}}(p), r}| + \tilde{c}_T \pi(\hat{k}_{\text{IC}}(p), r). \quad (3.13)$$

In (3.13), $\hat{\Sigma}_{\hat{k}_{\text{IC}}(p), r}$ denotes the estimate of the residual covariance matrix obtained from the conventional reduced rank regression estimation of (2.1) under rank r for lag length $\hat{k}_{\text{IC}}(p)$, while the penalty component, $\pi(\hat{k}_{\text{IC}}(p), r)$, is defined as for the definition of $\pi(k, r)$ just below (3.3) but with k replaced by $\hat{k}_{\text{IC}}(p)$. Notice that the coefficient on the penalty term is labelled \tilde{c}_T in (3.13) to highlight the fact that it need not be the same coefficient as used in (3.10) in step one.

In section 4.2 we will provide sufficient conditions on the penalty terms c_T and \tilde{c}_T which ensure the consistency of (3.11) and (3.12) under Assumption \mathcal{H} .

The Cheng and Phillips (2009, 2012) Procedure

In this approach, due to Cheng and Phillips (2009, 2012), about which we have already provided some initial simulation results in Table 1, the researcher remains agnostic about the short-run dynamics in the first step by imposing a single lag in the VAR model specification. This can be seen as a special case of (3.3) where k is fixed to 1. Specifically, the cointegration rank estimator of Cheng and Phillips (2009, 2012) is given, in generic form, by

$$\hat{r}_{\text{IC}}(1) = \arg \min_{r=0, \dots, p} \text{IC}(1, r). \quad (3.14)$$

In section 4.2 we will show that provided the penalty term c_T satisfies the same condition, $c_T/T + 1/c_T \rightarrow 0$ as $T \rightarrow \infty$, as in Cheng and Phillips (2009, 2012) then so (3.14) is consistent under Assumption \mathcal{H} . Indeed we show that this result holds when any lag length greater than or equal to one is used in the first step.

PLR and Bootstrap PLR Test-Based Procedures

The final two-step procedure we consider is the one that is probably the most widely used in empirical work. Here, once $\hat{k}_{\text{IC}}(p)$ has been obtained in the first step, the second step then uses the well-known PLR test-based sequential procedure of Johansen (1995) to determine the cointegration rank. The PLR test for the null hypothesis that the cointegration rank is (less than or equal to) r , denoted $H(r)$, against $H(p)$ rejects for large values of the trace statistic

$$Q_{r, \hat{k}_{\text{IC}}(p), T} := -T \sum_{i=r+1}^p \log(1 - \hat{\lambda}_i^{(\hat{k}_{\text{IC}}(p))}), \quad (3.15)$$

where $\hat{\lambda}_1^{(\hat{k}_{\text{IC}}(p))} > \dots > \hat{\lambda}_p^{(\hat{k}_{\text{IC}}(p))}$ are the largest p solutions to the eigenvalue problem in (3.2) but with k replaced by $\hat{k}_{\text{IC}}(p)$. Johansen's sequential procedure consists of, starting with $r = 0$, testing iteratively $H(r)$ against $H(p)$, for $r = 0, \dots, p-1$, until, for a given value of r , the asymptotic p -value associated with $Q_{r, \hat{k}_{\text{IC}}(p), T}$, exceeds a chosen (marginal) significance level. In what follows we will denote this estimator as $\hat{r}_{\text{PLR}}(\hat{k}_{\text{IC}}(p))$.

The analogous bootstrap procedure for determining the cointegrating rank is proposed in CRT2012 (i.i.d. re-sampling) and CRT2014 (wild bootstrap).⁴ Again based on the lag length $\hat{k}_{\text{IC}}(p)$ estimated in the first step, the corresponding bootstrap algorithm iterates over the following steps:

Algorithm 1:

- (i) Initiating the Algorithm at $r = 0$, estimate model (2.1) for lag length $\hat{k}_{\text{IC}}(p)$ using the conventional reduced rank regression of rank r , yielding the estimates $\hat{\beta}^{(r)}, \hat{\alpha}^{(r)}, \hat{\rho}^{(r)}, \hat{\Gamma}_1^{(r)}, \dots, \hat{\Gamma}_{\hat{k}_{\text{IC}}(p)-1}^{(r)}$ and $\hat{\phi}^{(r)}$, together with the corresponding residuals, $\hat{\varepsilon}_{r,t}$.⁵
- (ii) Construct the bootstrap sample recursively from

$$\Delta X_{r,t}^* = \hat{\alpha}^{(r)} \hat{\beta}^{(r)'} X_{r,t-1}^* + \sum_{i=1}^{\hat{k}_{\text{IC}}(p)-1} \hat{\Gamma}_i^{(r)} \Delta X_{r,t-i}^* + \hat{\alpha}^{(r)} \hat{\rho}^{(r)'} D_t + \hat{\phi}^{(r)} d_t + \varepsilon_{r,t}^*, \quad t = 1, \dots, T$$

initialised at $X_{r,j}^* = X_j$, $j = 1 - \hat{k}_{\text{IC}}(p), \dots, 0$, and with the T bootstrap errors $\varepsilon_{r,t}^*$ generated using the re-centred residuals, $\hat{\varepsilon}_{r,t}^c := \hat{\varepsilon}_{r,t} - T^{-1} \sum_{i=1}^T \hat{\varepsilon}_{r,i}$, for either:

⁴It should be noted that the algorithms proposed in CRT2012 and CRT2014, and indeed in Swensen (2006), include a step which checks that the bootstrap data generated in step (ii) of Algorithm 1 below satisfy the $I(1, r)$ conditions. In the context of the setting here this entails checking that the equation $|\hat{A}^{(r)}(z)| = 0$, where $\hat{A}^{(r)}(z) := (1 - z)I_p - \hat{\alpha}^{(r)} \hat{\beta}^{(r)'} z - \sum_{i=1}^{\hat{k}_{\text{IC}}(p)-1} \hat{\Gamma}_i^{(r)} z(1 - z)$, has roots either equal to 1 or outside the unit circle. Where this condition is not satisfied the bootstrap samples generated in step (ii) 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. 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 cointegration rank considered in CRT2012 and CRT2014, violations of the root check can be safely ignored in practice, as we choose to do here.

⁵The estimates $\hat{\beta}^{(r)}$ are here defined as $\hat{\beta}^{(r)} := \hat{V} K_p^{(r)}$, where $K_p^{(r)} := (I_r, 0_{r \times (p-r)})'$ is a selection matrix indexed by r and p , and, when deterministic terms are included, $\hat{\beta}^{\#(r)} := (\hat{\beta}^{(r)'}, \hat{\rho}^{(r)'})' = \hat{V} K_{p+1}^{(r)}$. Here $\hat{V} = [\hat{v}_1, \dots, \hat{v}_p]$, normalised by $\hat{V}' S_{11}^{(\hat{k}_{\text{IC}}(p))} \hat{V} = I_p$, denotes the eigenvectors corresponding to the ordered eigenvalues $\hat{\lambda}_1^{(\hat{k}_{\text{IC}}(p))} > \dots > \hat{\lambda}_p^{(\hat{k}_{\text{IC}}(p))}$ which solve the determinantal equation in (3.2). See CRT2012 for further details.

- (a) the i.i.d. bootstrap, such that $\varepsilon_{r,t}^* := \hat{\varepsilon}_{r,\mathcal{U}_t}^c$, where \mathcal{U}_t , $t = 1, \dots, T$ is an i.i.d. sequence of discrete uniform distributions on $\{1, 2, \dots, T\}$, or
- (b) a wild bootstrap, where for each $t = 1, \dots, T$, $\varepsilon_{r,t}^* := \hat{\varepsilon}_{r,t}^c w_t$, where w_t , $t = 1, \dots, T$, is an i.i.d. $N(0,1)$ sequence.
- (iii) Using the bootstrap sample, $\{X_{r,t}^*\}$, and denoting by $\hat{\lambda}_1^* > \dots > \hat{\lambda}_p^*$ the ordered solutions to the bootstrap analogue of the eigenvalue problem in (3.2), compute the bootstrap LR statistic $Q_{r,\hat{k}_{IC}(p),T}^* := -T \sum_{i=r+1}^p \log(1 - \hat{\lambda}_i^*)$. Define the corresponding p -value as $p_{r,\hat{k}_{IC}(p),T}^* := 1 - G_{r,\hat{k}_{IC}(p),T}^*(Q_{r,\hat{k}_{IC}(p),T}^*)$, $G_{r,\hat{k}_{IC}(p),T}^*(\cdot)$ denoting the conditional (on the original data) cdf of $Q_{r,\hat{k}_{IC}(p),T}^*$.
- (iv) If $p_{r,\hat{k}_{IC}(p),T}^*$ exceeds the significance level, η , set $\hat{r}_{PLR}^*(\hat{k}_{IC}(p)) = r$, otherwise repeat steps (i)–(iv) testing the null of rank $(r + 1)$ against rank p if $r + 1 < p$, or set $\hat{r}_{PLR}^*(\hat{k}_{IC}(p)) = p$ if $r + 1 = p$.

For the case where the lag length is known, CRT2014 show that in the presence of unconditional and/or conditional heteroskedasticity in ε_t the wild bootstrap procedure significantly outperforms the corresponding procedures based on either the asymptotic PLR test or the i.i.d. bootstrap PLR test. Moreover, they show that only the wild bootstrap PLR test is correctly sized asymptotically under unconditional heteroskedasticity. Consequently, we will only report simulation results for the wild bootstrap PLR-test based procedure; that is, option (b) of step (ii) of Algorithm 1. We will analyse the large sample properties of the PLR-based two-step procedures in section 4.2.

4 Asymptotic Analysis

In this section we detail the large sample properties of the various cointegration rank determination methods discussed in section 3. In doing so, we extend the asymptotic analysis in CRT2012 and CRT2014 for (standard and bootstrap) sequential PLR testing by providing new results on the large sample properties of PLR tests in the case where the lag length of the CVAR model in (2.1) is estimated from the data, rather than assumed known. We similarly generalise the large sample results presented in CDRT for using information criteria to determine the cointegration rank to the case where the lag length is unknown. Finally, our results also establish the consistency of information-criteria based methods for selecting the lag length in non-stationary VAR processes driven by heteroskedastic shocks of the form considered in Assumption \mathcal{H} .

4.1 Joint Determination of the Lag Length and Cointegration Rank

In the case where ε_t in (2.1) is i.i.d. with mean vector zero, variance matrix Σ and finite fourth moments, Aznar and Salvador (2002) outline a strategy to prove the consistency of the joint estimator of the lag length and cointegration rank in (3.11), which minimises (3.3) jointly over $r = 0, \dots, p$ and $k = 1, \dots, K$, which is as follows. First they provide proofs of the results that: (i) r can be consistently

determined independently of the lag length adopted, and (ii) that conditional on the true cointegration rank being used, the lag order can then be consistently determined, in each case provided $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$. Unfortunately the proof they provide of the result in (i) contains errors for the case where $k < k_0$, that is where the lag length in (3.11) is set smaller than the true lag length. They subsequently claim, but do not provide a formal proof of this conjecture, that these two results together imply that $(\tilde{k}_{IC}, \tilde{r}_{IC}) \xrightarrow{P} (k_0, r_0)$. In Lemma 1 and Lemma 2, respectively, we generalise these first two results from Aznar and Salvador (2002) to the case where ε_t satisfies Assumption \mathcal{H} , also correcting the error in the proof of the first result (see Remark A.1 for further comment on this). In Theorem 1 we then provide a formal proof that taken together these two results are indeed sufficient to prove the consistency of (3.11).

Lemma 1 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions and let Assumption \mathcal{H} hold. Then, for any $0 < k \leq K$, it holds that, as $T \rightarrow \infty$:*

- (i) *for $r > r_0$, $\text{Prob}(IC(k, r) > IC(k, r_0)) \rightarrow 1$, provided $c_T \rightarrow \infty$;*
- (ii) *for $r < r_0$, $\text{Prob}(IC(k, r) > IC(k, r_0)) \rightarrow 1$, provided $c_T/T \rightarrow 0$.*

Remark 2: The results in Lemma 1 therefore imply that, provided $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$, i.e. such that the coefficient c_T in the penalty function diverges at a slower rate than the sample size, then it will hold for any $k = 1, \dots, K$ that $\hat{r}_{IC}(k) \xrightarrow{P} r_0$. Consequently, the use of either the BIC or HQC, but not the AIC, penalty will yield a weakly consistent estimate of r_0 under Assumption \mathcal{H} , regardless of the lag length used in estimating (2.1).

Remark 3: Lemma 1 generalises the results given in Theorem 1 of CDRT which establish consistent estimation of r_0 in the case when the lag order is known and fixed to the true value; that is, they show that $\hat{r}_{IC}(k_0) \xrightarrow{P} r_0$ under the same conditions on the penalty term. Lemma 1 also extends the large sample results given in Cheng and Phillips (2009, 2012) for $k = 1$ to allow for conditional heteroskedasticity in ε_t through Assumption \mathcal{H} .

Lemma 2 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions and let Assumption \mathcal{H} hold. Then it holds that, as $T \rightarrow \infty$:*

- (i) *for any k such that $k_0 < k \leq K$, $\text{Prob}(IC(k, r_0) > IC(k_0, r_0)) \rightarrow 1$, provided $c_T \rightarrow \infty$;*
- (ii) *for any k such that $0 < k < k_0$, $\text{Prob}(IC(k, r_0) > IC(k_0, r_0)) \rightarrow 1$, provided $c_T/T \rightarrow 0$.*

Remark 4: The results in Lemma 2 imply that, again provided $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$, then so an information criterion-based estimator of the lag length, calculated for the true cointegration rank r_0 in (2.1), that is $\hat{k}_{IC}(r_0) := \arg \min_{k=1, \dots, K} IC(k, r_0)$, will be such that $\hat{k}_{IC}(r_0) \xrightarrow{P} k_0$ under Assumption \mathcal{H} . Again, therefore the BIC and HQC penalties satisfy this condition and, hence, will deliver a consistent

estimate of the lag length in the case where (2.1) is estimated under the true cointegration rank, whereas the AIC penalty does not. Lemma 2 extends the corresponding results given in Theorem 4.3 of Aznar and Salvador (2002) to the case where ε_t is heteroskedastic, satisfying Assumption \mathcal{H} , although it should be noted that in doing so we have adopted a different method of proof from that used in Aznar and Salvador (2002).

Using the results in Lemmas 1 and 2, we are now in a position to state the following theorem which establishes the weak consistency of the joint procedure.

Theorem 1 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions and let Assumption \mathcal{H} hold. Then it holds that $(\tilde{k}_{IC}, \tilde{r}_{IC}) \xrightarrow{P} (k_0, r_0)$, provided c_T in (3.3) satisfies the condition that $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$.*

Remark 5: An immediate corollary of the result in Theorem 1 is that the resulting BIC-based and HQC-based estimators, $(\tilde{k}_{BIC}, \tilde{r}_{BIC})$ and $(\tilde{k}_{HQC}, \tilde{r}_{HQC})$ respectively, are weakly consistent for the cointegration rank and lag length, but the corresponding AIC-based estimator is not.

4.2 Sequential Determination of Lag Length and Cointegration Rank

We first provide the sufficient conditions on c_T which ensure the consistency of an approach based on minimising the criteria in (3.10) and (3.13) under Assumption \mathcal{H} . In doing so, we use again Lemma 1 and derive, in Lemma 3, the conditions under which minimising an information criterion of the form given in (3.10) consistently selects the true lag order, k_0 , in the first step when we do not impose a reduced rank structure on the matrix Π . To the best of our knowledge, the consistency of information criteria-based lag length estimation in non-stationary VAR models has not previously been established in the literature for the case where the shocks can display unconditional heteroskedasticity of the form given in Assumption \mathcal{H} . Using Lemma 1, we then demonstrate that the information criterion-based estimator, (3.12), is consistent for the cointegration rank under the same conditions as are required on c_T in the first step.

Lemma 3 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions and let Assumption \mathcal{H} hold. Then, for any $0 < k \leq K$, it holds that, as $T \rightarrow \infty$:*

(i) *for $k > k_0$, $\text{Prob}(IC(k, p) > IC(k_0, p)) \rightarrow 1$, provided $c_T \rightarrow \infty$;*

(ii) *for $k < k_0$, $\text{Prob}(IC(k, p) > IC(k_0, p)) \rightarrow 1$, provided $c_T/T \rightarrow 0$.*

Remark 6: Lemma 3 implies that, again provided $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$, then $\hat{k}_{IC}(p) \xrightarrow{P} k_0$. Again this implies that the use of the BIC or HQC penalty will yield a consistent estimate of the true lag order while AIC will not. Lemma 3 generalises the results obtained for the case where the shocks are unconditionally homoskedastic in, among others, Paulsen (1984) and Nielsen (2006), to the case where ε_t can be unconditionally heteroskedastic as in Assumption \mathcal{H} .

Using the results in Lemma 1 and Lemma 3, we are now in a position to establish the consistency of the two-step information criteria-based procedure for determining the lag length and cointegration rank.

Theorem 2 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions and let Assumption \mathcal{H} hold. Then it holds that $\hat{k}_{IC}(p) \xrightarrow{p} k_0$ and $\hat{r}_{IC}(\hat{k}_{IC}(p)) \xrightarrow{p} r_0$, provided that c_T in (3.10) satisfies the condition that $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$, and that the analogous condition holds on \tilde{c}_T in (3.13).*

Remark 7: It is straightforward to show that the condition placed on c_T in Theorem 2 is not required if we only desire a consistent estimate of the cointegration rank. Indeed, any fixed lag length k will also suffice in that case, as was shown in Lemma 1. Consequently, for example, the second step rank estimator, $\hat{r}_{IC}(\hat{k}_{AIC}(p))$, which uses the lag length determined by the AIC penalty in (3.10) in the first step, will be consistent for the cointegration rank, provided the penalty function, c_T^* , satisfies the condition in Theorem 2. That condition, of course, rules out using the AIC penalty function in (3.13) in the second step of the procedure.

Remark 8: Notice that the consistency results given in Theorems 1 and 2 imply that the information criteria satisfying the conditions on the term c_T , by construction, cannot detect local-to- r_0 alternatives. Consider, e.g., $\Delta X_t = \Pi_T X_{t-1} + \sum_{i=1}^{k_0-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$, with $\Pi_T = \alpha\beta' + T^{-1}\alpha_1\beta_1'$ where α and β are $p \times r_0$ full rank matrices whereas α_1 and β_1 are of dimension $p \times 1$ and β_1 is not in the span of β . That is, for fixed T the cointegration rank is $r_0 + 1$ but it becomes r_0 as $T \rightarrow \infty$. As in Johansen (1995, Theorem 14.4), in our heteroskedastic case (under Assumption \mathcal{H}), $T \log(1 - \hat{\lambda}_{r_0+1}^{(k_0)})$ is still of order $O_p(1)$ and, thus, in all the cases where c_T diverges (e.g., BIC and HQC), $P(\text{IC}(k_0, r_0 + 1) - \text{IC}(k_0, r_0) > 0) \rightarrow 1$ in the limit. This result obtains as $\text{IC}(k_0, r_0 + 1) - \text{IC}(k_0, r_0) = T \log(1 - \hat{\lambda}_{r_0+1}^{(k_0)}) + c_T(2p - 1)$ will thus diverge to positive infinity as $T \rightarrow \infty$. Furthermore, using the results in Lemma 1, the latter also holds for any fixed value of k , with $k = 1, \dots, K$. Similar arguments can be used for local-to- k_0 alternatives, that is, for, e.g., $\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k_0-1} \Gamma_i \Delta X_{t-i} + T^{-1/2} \Gamma_{k_0} \Delta X_{t-k_0} + \varepsilon_t$, so that for fixed T the autoregressive lag length is $k_0 + 1$ but it becomes k_0 as $T \rightarrow \infty$. In this case, results (i) in Lemmas 2 and 3 imply that consistent information criterion-based estimators of the lag length cannot detect local-to- k_0 alternatives.

We now turn to establishing the large sample properties of the asymptotic PLR-based and bootstrap PLR-based sequential estimators of the cointegration rank from section 3.2; that is, $\hat{r}_{\text{PLR}}(\hat{k}_{IC}(p))$ and $\hat{r}_{\text{PLR}}^*(\hat{k}_{IC}(p))$ respectively.

It is known from Johansen (1995) and CRT2010a that $\hat{r}_{\text{PLR}}(k_0)$, the rank estimator obtained using the sequential procedure based on the asymptotic PLR test with knowledge of the true lag length, k_0 , is asymptotically accurately capped⁶ when ε_t is i.i.d. with finite fourth moments. CRT2010a show that this also holds when the shocks are conditionally heteroskedastic; that is, under Assumption

⁶ The sequential rank determination scheme of Johansen (1995) is *asymptotically accurately capped* in that if each PLR (or bootstrap PLR) test in the sequence is run with nominal (asymptotic) significance level η , then the limiting

\mathcal{H} restricted such that $\sigma(\cdot) = \sigma$. However, CRT2010b show that $\hat{r}_{\text{PLR}}(k_0)$ is not asymptotically accurately capped if a fixed significant level η is used (although it still never selects a rank smaller than the true rank in large samples, it will not select the true rank with probability $1 - \eta$ in the limit) when the shocks are unconditionally heteroskedastic because of the dependence of the limiting null distribution of the PLR statistic, $Q_{r_0, k_0, T}$, on nuisance parameters arising from the heteroskedasticity. The foregoing results also hold for the corresponding bootstrap PLR procedure based on i.i.d. re-sampling. However, where the wild bootstrap PLR test is used, CRT2014 show that $\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}(p))$ is asymptotically accurately capped (in the sense defined in footnote 5) under Assumption \mathcal{H} .

Using the large sample results above from CRT2010a, CRT2012 and CRT2014, coupled with Lemma 3, we now establish in Theorem 3 that the wild bootstrap $\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}(p))$ is an asymptotically accurately capped estimator of the cointegration rank under Assumption \mathcal{H} .

Theorem 3 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions and let Assumption \mathcal{H} hold. Then for the wild bootstrap PLR-tests: (i) $\lim_{T \rightarrow \infty} P(\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}(p)) = r) = 0$ for all $r = 0, 1, \dots, r_0 - 1$; (ii) $\lim_{T \rightarrow \infty} P(\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}(p)) = r_0) = 1 - \eta \cdot \mathbb{I}(r_0 < p)$, where $\mathbb{I}(\cdot)$ denotes the indicator function, and $\lim_{T \rightarrow \infty} \sup_{r \in \{r_0+1, \dots, p\}} P(\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}(p)) = r) \leq \eta$, provided that c_T in (3.10) is such that $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$.*

Remark 9: If the conditions of Theorem 3 are strengthened to include the further restriction that $\sigma(\cdot) = \sigma$ in Assumption \mathcal{H} , so that only conditional heteroskedasticity is permitted in ε_t , then a similar proof strategy can be used to show that the results given in Theorem 3 will also hold for the rank estimators from the analogous procedures based on either the asymptotic PLR tests or the i.i.d bootstrap PLR tests. A proof of this result can be found in the accompanying working paper, Cavaliere *et al.* (2014).

Remark 10: As can be inferred straightforwardly from, for example the proof of Theorem 3 in the Appendix, the results in Theorem 3 and Remark 9 will also hold (under the same conditions) for any consistent lag length estimator obtained in the first step. It is also straightforward to show that these large sample results hold for any fixed lag length k used in place of $\hat{k}_{\text{IC}}(p)$ in Algorithm 1, provided $k \geq k_0$. Moreover, it can be shown that these results also hold even if the penalty function used in (3.10) does not satisfy the condition that $c_T \rightarrow \infty$ as $T \rightarrow \infty$; consequently, using the lag length estimator $\hat{k}_{\text{AIC}}(p)$, based on the AIC penalty function, in Algorithm 1 will deliver a rank estimator with the same large sample properties as stated in Theorem 3. That said, as the results in Table 1 above and in Boswijk and Franses (1992) demonstrate, in finite samples over-specification of k tends to negatively impact on the accuracy of the rank selection method.

probability of selecting a rank smaller than, equal to, and greater than the true rank will be 0, $1 - \eta$ and η , respectively, when $r_0 < p$ and 0, 1 and 0, respectively, when $r_0 = p$. Consistency in the same sense as previously used for the information-based procedures would require the nominal significance level used to be some sequence η_T which converges to zero as $T \rightarrow \infty$.

Remark 11: In the context of a triangular error-correction DGP driven by shocks which form an infinite order moving average (subject to a summability condition) with innovations which are i.i.d. with finite fourth moments, Lütkepohl and Saikkonen (1999, Theorem 5.1) demonstrate that first estimating the lag length by minimising an information criterion, such as that given in (3.10), has no impact on the asymptotic null distributions of the PLR tests for the cointegration rank based on that estimated lag length. In this sense, the results in Theorem 3 (and in Remarks 8 and 9) extend their results to show that the same can also be achieved when the shocks are heteroskedastic, albeit in the context of a pure VAR model.

5 Numerical Results

Using Monte Carlo simulation methods we now investigate the finite sample performances of the methods for estimating the cointegration rank outlined in section 3 in the context of models with conditional or unconditional heteroskedasticity. The results from these Monte Carlo experiments are reported in Tables 2, 3.1, 3.2, 3.3, 4.1, 4.2 and 4.3 which can be found in the accompanying online Supplementary Appendix.

For our simulation DGP we will consider an extended version of the simulation DGP used in CDRT, given by the following VAR(2) process of dimension $p = 4$:

$$\Delta X_t = \alpha\beta'X_{t-1} + \Gamma_1\Delta X_{t-1} + \varepsilon_t, \quad t = 1 - K, \dots, T \quad (5.1)$$

with ε_t specified below, K denotes the maximum lag order, $X_{-K} = \Delta X_{-K} = 0$, and for sample sizes $T \in \{50, 100, 200, 400\}$. Notice that, with respect to CDRT, the true lag order, $k_0 = 2$, is assumed unknown to the practitioner throughout the Monte Carlo simulation study and is thus estimated using an information criterion following either a joint procedure or a sequential procedure, as outlined in sections 3.1 and 3.2, respectively. In our simulation analysis we consider $K = 6$ as maximum lag length. In order to allow for a true cointegration rank, r_0 , of 0, 1 or 2, the long-run parameter vectors are set to

$$\beta' := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \alpha' := \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \end{bmatrix}. \quad (5.2)$$

Furthermore, we set $\Gamma_1 := \gamma I_4$. Consequently, for the $I(1, r)$ conditions to be met, it suffices that $(a, b, \gamma) \in (-2, 0]^2 \times [0, 1)$. The specific values of a and b to be used when $r_0 > 0$ will be given in section 5.3. In the following, results are reported for $\gamma \in \{0.0, 0.3, 0.5, 0.9\}$. Notice that the system in (5.1) approaches the $I(2)$ boundary as γ approaches unity.⁷

In the context of (5.1) for the individual components of ε_t we consider a selection of the univariate innovation processes and parameter configurations as used in CDRT, CRT2014, Gonçalves and Kilian

⁷Notice that the stationary roots for the case of $r_0 = 0$ are equal to the short-run parameter γ . The stationary roots for the case of $r_0 = 1$ and $r_0 = 2$ are $\{0, 0.6\}$, $\{0.3, 0.45 \pm 0.3122i\}$, $\{0.5, 0.55 \pm 0.4444i\}$ and $\{0.9, 0.75 \pm 0.5809i\}$, where $i := \sqrt{-1}$, for $\gamma = 0, 0.3, 0.5$ and 0.9 , respectively.

(2004), and CRT2010b. These are as follows:⁸

- **Case A.** ε_{it} is a GJR-GARCH(1,1) process driven by skewed Student- t innovations of the form $\varepsilon_{it} = h_{it}^{1/2} v_{it}$, $i = 1, \dots, p$, where v_{it} is distributed as a skewed t with ν degrees of freedom and skewness parameter δ , independent across i , and $h_{it} = \omega + d_0 \varepsilon_{it-1}^2 + d_1 \mathbb{I}(\varepsilon_{it-1} < 0) \varepsilon_{it-1}^2 + d_2 h_{it-1}$, $t = 0, \dots, T$. Results are reported for $d_0 = 0.03$, $d_1 = 0.04$, $d_2 = 0.92$, $\nu = 5$, and $\delta = -0.1$.
- **Case B.** ε_{it} is the first-order AR stochastic volatility [SV] model: $\varepsilon_{it} = v_{it} \exp(h_{it})$, $h_{it} = \lambda h_{it-1} + 0.5 \xi_{it}$, with $(\xi_{it}, v_{it})' \sim \text{i.i.d. } N(0, \text{diag}(\sigma_\xi^2, 1))$, independent across $i = 1, \dots, p$. Results are reported for $\lambda = 0.951$, $\sigma_\xi = 0.314$.
- **Case C.** ε_{it} is a non-stationary, heteroskedastic independent sequence of $N(0, \sigma_{it}^2)$ variates, where $\sigma_{it}^2 = 1$ for $t \leq \lfloor T\tau \rfloor$ and $\sigma_{it}^2 = \kappa$ for $t > \lfloor T\tau \rfloor$, all $i = 1, \dots, p$. Results are reported for $\tau = 2/3$ and $\kappa = 3$ (late positive variance shift).

Cases A-C are all nested within the general Assumption \mathcal{H} . Cases A and B constitute two well-known conditionally heteroskedastic models for the innovations, both of which rule out unconditional heteroskedasticity. Case C involves a single, permanent shift in the innovation variance, thereby generating unconditionally heteroskedastic error sequences. Further details on these three Cases can be found in section 5 of CDRT and in CRT2014.

All reported experiments are run over 10,000 Monte Carlo replications and were programmed using MATLAB. These programs are available on request. For the reasons stated at the end of section 3.2, results for the two-step PLR-based approach will only be reported in connection with the wild bootstrap PLR-tests, $Q_{r,\hat{k},T}^*$. The number of replications used in the wild bootstrap algorithms is set to 399. The wild bootstrap PLR tests were all conducted at the nominal 0.05 significance level.⁹

We first consider the determination of the lag length using information criteria in section 5.1. Then, in section 5.2, we consider the case of no cointegration and we investigate both the joint and the sequential determination of k and r using the methods discussed in section 3. Analogously, section 5.3 discusses the cases of a single cointegration vector, $r_0 = 1$. In view of length considerations, we will not report results for the case of two linearly independent cointegrating relations here because these add rather little to the results for $r_0 = 1$. Moreover, and in line with results reported in previous

⁸We also considered a homoskedastic case where ε_{it} , $i = 1, \dots, p$, is an independent sequence of $N(0, 1)$ variates and the further conditionally heteroskedastic case where ε_{it} , $i = 1, \dots, p$, are standard GARCH(1,1) processes driven by standard normal innovations. The results for these two cases were very similar to one another and also to the results reported here for Case A and, hence, are omitted in the interests of space. They can, however, be found in the accompanying working paper, Cavaliere *et al.* (2014).

⁹Violations of the root check condition associated with the bootstrap sequential algorithm outlined in footnote 3 do occur, most notably when volatility is persistent (Cases B and C) and the sample size is small ($T = 50$ and 100), but the frequency with which these occur decreases rapidly as the sample size increases, other things equal. For example, when $T \geq 200$, the root check condition is failed (for at least one of the values of r considered) in Algorithm 1 under Case C for $r_0 = 0$ ($r_0 = 1$) no more than 1.0% (1.3%) and 0.8% (1.1%) of the time when k is determined in the first step using $\text{HQC}(k, p)$ and $\text{BIC}(k, p)$, respectively.

studies (see, e.g., Kapetanios, 2004; CDRT), we also found that the joint $AIC(k, r)$ procedure, the semi-parametric $AIC(1, r)$ approach and the sequential $AIC(\hat{k}_{IC}(p), r)$ procedures all displayed very much poorer performance than the other methods considered for determining cointegrating rank and, hence, the results for these procedures are also not reported. They can, however, be found, along with the results for $r_0 = 2$, in the accompanying working paper, Cavaliere *et al.* (2014). Throughout a restricted constant was fitted for the deterministic component.¹⁰

5.1 Lag Length Determination

First we briefly examine the behaviour of information criteria $IC(k, p)$ defined in (3.10) in determining the true lag length, $k_0 = 2$. Notice that this is used as the first step in the sequential procedures outlined in section 3.2, that is the selected lag order, $\hat{k}_{IC}(p)$ of (3.11), will then be adopted in the second step where the cointegration rank is determined. Corresponding results for the lag length estimators \tilde{k}_{IC} of (3.7), obtained from the joint information criteria-based procedures, were very similar and, hence, are not reported here in the interests of space; these results are available on request. For a maximum value of the lag length set to $K = 6$, Table 2 reports the results for the determination of the lag order k using information criteria in the case where the true cointegration rank is equal to 1.¹¹ From these results it is seen that the BIC-based approach provides the best performance either when $\gamma \geq 0.5$ or when $\gamma = 0$, the latter being the case where the true VAR order is 1. However, and for each of Cases A-C, when $\gamma = 0.3$ BIC(k, p) performs poorly and the HQC-based approach appears preferable.

The results in Table 2 suggest that the presence of either conditional (Cases A and B) or unconditional heteroskedasticity (Case C) in the innovation process has only limited impact on the performance of the information criteria-based lag length estimation procedures.

The results in Table 2 and in the accompanying working paper, Cavaliere *et al.* (2014), also confirm some well-known properties of the information criteria considered in our analysis; most notably, the tendency of $AIC(k, p)$ to over-estimate k , and the result that, in general, $\hat{k}_{BIC} \leq \hat{k}_{HQC} \leq \hat{k}_{AIC}$, in line with the respective penalty functions used in these three criteria, noting that c_T in (3.10) satisfies $\log T < 2 \log \log T < 2$ when $T > 15$.

We now report the results pertaining to cointegration rank determination. In the interests of space, in what follows we omit results for the sequential determination of k and r which are based on the same criterion in each step, e.g., $BIC(\hat{k}_{BIC}(p), r)$, because these behaved very similarly to the corresponding joint procedure, as might be expected.

¹⁰We also considered the restricted trend case. Overall, and as might be expected, these results were qualitatively similar to those for the restricted constant case, but with an overall deterioration observed in the performance of all of the procedures the more so the smaller the sample size, other things equal. These results are available on request.

¹¹The results for the determination of k in the cases where $r_0 = 0$ and $r_0 = 2$ are very similar to those reported in Table 2. Indeed, as might be expected, the value of the cointegration rank does not appear to have a strong influence on the performance of these procedures, albeit they perhaps fair marginally better overall where cointegration holds ($r_0 > 0$) vis-à-vis the no cointegration case, $r_0 = 0$. These results are therefore not reported here but can be found in the accompanying working paper, Cavaliere *et al.* (2014).

5.2 The No Cointegration Case ($r_0 = 0$)

No cointegration ($r_0 = 0$) obtains by setting $a = b = 0$ in the long-run parameter vector α in (5.2). In this case (5.1) reduces to the VAR(1) in first differences, $\Delta X_t = \Gamma_1 \Delta X_{t-1} + \varepsilon_t$, $t = 1, \dots, T$.

The frequencies of cointegration rank determination for the methods considered in section 3 are reported in Tables 3.1, 3.2 and 3.3 for Cases A, B and C, respectively. The first point worth noting is that the results in Table 3.1 are very similar overall to those which are obtained when the innovations are i.i.d. Gaussian (the latter results are reported in the accompanying working paper, Cavaliere *et al.*, 2014). In contrast, and consonant with the simulation results in CDRT, the finite sample performances of all the methods considered deteriorate significantly under both the autoregressive SV model (Case B) and under a single, permanent shift in volatility (Case C).

We now consider the results in more detail. We first consider the results for the joint IC-based procedure reported in the upper-left panels of Tables 3.1-3.3. The joint BIC-based procedure, $\text{BIC}(k, r)$, generally performs well, indeed it selects $r = 0$ at least 90% of the time when $T = 200$ and exceeds 96% when $T = 400$, for all values of γ and for all of Cases A-C. However, $\text{BIC}(k, r)$ does display a significant tendency to over-estimate the cointegration rank when the sample size is small ($T \leq 100$), especially for larger values of γ . The joint HQC-based procedure is dominated by $\text{BIC}(k, r)$ throughout the results in the upper-left panels of Tables 3.1-3.3 but is clearly preferred to the $\text{AIC}(k, r)$ procedure.

Consider next the results for the sequential IC-based procedure reported in the upper-right panels of Tables 3.1-3.3. The results in Tables 3.1-3.3 show that $\text{BIC}(\hat{k}_{\text{HQC}}(p), r)$ leads to similar results as for the joint $\text{BIC}(k, r)$, and consequently to the sequential $\text{BIC}(\hat{k}_{\text{BIC}}(p), r)$, when $T \geq 200$. This is due to the fact that BIC and HQC tend to select the same number of lags in larger samples (see the results in Table 2). However, when $T = 50$ and 100 the joint BIC-based approach outperforms $\text{BIC}(\hat{k}_{\text{HQC}}(p), r)$ for all values of γ and for all of Cases A-C. Moreover, although the sequential $\text{BIC}(\hat{k}_{\text{AIC}}(p), r)$ procedure performs well when $T \geq 200$ ($T = 400$ for Case C) it is dominated by both $\text{BIC}(\hat{k}_{\text{HQC}}(p), r)$ and $\text{BIC}(k, r)$ when the sample size is small. This is due to the tendency of the $\text{AIC}(k, p)$ to over-estimate k , especially in small samples, which can be observed in the accompanying working paper, Cavaliere *et al.*, 2014). The sequential HQC-based approaches for estimating r perform very similarly to the corresponding joint HQC-based procedure. However, the joint BIC-based approach and $\text{BIC}(\hat{k}_{\text{HQC}}(p), r)$ appear preferable in each of Cases A-C.

The results for the semi-parametric approach by Cheng and Phillips (2009), based on $\text{IC}(1, r)$, reported in the lower-left panels of Tables 3.1-3.3, display significantly poorer performance than the corresponding joint criteria $\text{IC}(k, r)$. This is somewhat to be expected given that these criteria are based on fitting a (misspecified) VAR(1) to the data and in the light of the results already reported in Table 1 of section 3. In particular, when $\gamma = 0.9$, the frequency of selecting the true cointegration rank is almost zero for all the $\text{IC}(1, r)$, regardless of how large the sample size is. These results highlight that the semi-parametric procedure based on a reduced rank regression using a simple VAR(1) model is considerably affected by the misspecification of the lag length. The only cases where this procedure works well is where $\gamma = 0$, i.e., when the simulation DGP is in fact a VAR(1) process, where the

IC(1, r) based procedures perform similarly to the joint IC(k, r) procedures (upper-left panel) and generally outperform its sequential IC($\hat{k}_{IC}(p), r$) analogues (upper-right panel).

The lower-right panels of Tables 3.1-3.3 report the corresponding results for the sequential procedure based on the wild bootstrap, $Q_{r, \hat{k}_{IC}(p), T}^*$. Since the tests are run at the (asymptotic) 5% significance level, $Q_{r, \hat{k}_{IC}(p), T}^*$ should (in the limit) select $r = 0$ with probability 95% and $r > 0$ with probability 5%. In line with the results in CRT2014 and CDRT, the results in Tables 3.1-3.3 show that the wild bootstrap procedure delivers very decent finite sample performance throughout. In all cases when $T = 400$, the empirical frequency with which $Q_{r, \hat{k}_{IC}, T}^*$ selects the true cointegration rank in general exceeds 90%, even in the presence of either conditional heteroskedasticity (Cases A and B) or unconditional heteroskedasticity (Case C). Although the results in Tables 3.1-3.3 suggest that the information criterion used to select the lag length in the first step does not have a particularly large impact on the ability of the wild bootstrap rank test to determine the cointegration rank, the $Q_{r, \hat{k}_{BIC}(p), T}^*$ does marginally outperform the other procedures in most cases. Some exceptions occur for $\gamma = 0.3$ where BIC(k, p) has a tendency to select $\hat{k}_{BIC}(p) = 1$ and, hence, underfit the autoregressive lag order; cf. Table 2. Overall, in the case of no cointegration, the $Q_{r, \hat{k}_{IC}(p), T}^*$ procedures appear to be more reliable than the corresponding IC(k, r) procedures, especially in small samples. For example, comparing the results in the lower-right and upper-left panels of Table 3.3 for non-stationary unconditional heteroskedasticity (Case C), we observe that, when $T = 50$ and $\gamma = 0.9$, $Q_{r, \hat{k}_{BIC}(p), T}^*$ detects no cointegration 72.3% of the time¹², while BIC(k, r) only does so 29.3% of the time. The only exceptions are for small values of γ , namely when $\gamma \leq 0.5$, and GJR-GARCH shocks (Case A, cf. Table 3.1) where BIC(k, r) in general marginally outperforms $Q_{r, \hat{k}_{IC}(p), T}^*$, especially when T is small. For example, the results in the lower-right and upper-left panels of Table 3.1 for Case A (GJR-GARCH shocks) show that the joint BIC-based approach selects no cointegration 95.7% of the time when $T = 100$ and $\gamma = 0.3$ compared to 83.2% and 93.0% for $Q_{r, \hat{k}_{BIC}(p), T}^*$ and $Q_{r, \hat{k}_{AIC}(p), T}^*$, respectively.

5.3 The Cointegrated Case ($r_0 = 1$)

We now consider the case of a single cointegration vector. In this case, as in CDRT we consider the VAR(2) in (5.1) where we set $a = -0.4$ and $b = 0$ in the long-run parameter vector α in (5.2), thus obtaining $\beta' = (1, 0, 0, 0)$ and $\alpha' = (-0.4, 0, 0, 0)$. The speed of adjustment parameter is therefore -0.4 .

Tables 4.1-4.3 report the results for the GJR-GARCH(1,1), autoregressive SV, and single volatility break cases, respectively. First consider the joint procedures, IC(k, r), reported in the upper-left panels of Tables 4.1-4.3. The results show that the joint BIC-based approach correctly determines the true cointegration rank, $r_0 = 1$, more than 95% of the time when $T = 400$ in all the cases considered. These results confirm and extend the simulation findings of CDRT to the case where the lag order is estimated (and not set to the true value). However, as was seen for the $r_0 = 0$ case, its performance

¹²It should be stressed though that this still implies an empirical size for the wild bootstrap $Q_{0, \hat{k}_{BIC}(p), T}^*$ test of 26.7%, considerably in excess of the nominal 5% level.

is affected by the value of γ ; specifically, performance is significantly better when γ is large ($\gamma \geq 0.5$), although this property in itself is largely an artefact of the tendency of the joint BIC-based approach to over-estimate the cointegration rank in such cases. When $\gamma = 0.3$ and the sample size is small, $T \leq 100$, the joint HQC-based procedure seems preferable to the BIC-based procedure, although again this is largely an artefact of the greater tendency of the former to over-estimate both the lag order (see Table 2) and the cointegration rank in such cases when $r_0 = 1$.

In the case of one cointegration vector ($r_0 = 1$), the results in the upper-right panels of Tables 4.1-4.3 show that the sequential BIC-based procedure, $\text{BIC}(\hat{k}_{\text{AIC}}(p), r)$ and $\text{BIC}(\hat{k}_{\text{HQC}}(p), r)$, performs satisfactorily when the sample size is large, i.e., when $T = 400$ and dominates the sequential HQC-based approaches when $T \geq 200$ in all the cases considered. Conversely, $\text{BIC}(\hat{k}_{\text{IC}}(p), r)$ performs poorly for small values of γ , namely when $\gamma = 0.3$, and small sample sizes ($T = 50$ and 100). For these combinations of γ and T , the sequential approach based on $\text{HQC}(\hat{k}_{\text{BIC}}(p), r)$ is generally preferable.

The results for the semi-parametric information criteria, $\text{IC}(1, r)$, reported in the lower-left panels of Tables 4.1-4.3 show that, as in the case of $r_0 = 1$ discussed in the previous sub-section, they appear to deliver markedly worse performance in general than those based on an information criterion where the lag order is estimated in a prior step. Again these results show that the ability of the semi-parametric information criteria to determine the true cointegration rank is affected by the value of γ : the bigger is γ , the worse their performance, other things equal. In particular, their performance does not seem to improve as T increases when γ is large. This again highlights the importance of obtaining a good estimate of the autoregressive lag length and that not doing so can very adversely affect the ability of information criteria-based methods to select the true cointegration rank.

As was also observed in section 5.2 for the non-cointegrated DGP, the sequential procedures based on wild bootstrap PLR tests perform very well in general for the larger of the sample sizes considered ($T = 200$ and 400). However, their performance deteriorates in the smaller sample sizes considered, especially when the parameter γ is small. Overall, when $T = 400$, $Q_{r, \hat{k}_{\text{BIC}}(p), T}^*$ performs well for any value of γ and for all the heteroskedastic cases considered. Moreover, when the sample size is small and $\gamma \geq 0.5$, $Q_{r, \hat{k}_{\text{BIC}}(p), T}^*$ generally outperforms the other $Q_{r, \hat{k}_{\text{IC}}(p), T}^*$ procedures. Conversely, when $T \leq 200$ and $\gamma = 0.3$, $Q_{r, \hat{k}_{\text{BIC}}(p), T}^*$ performs poorly and either $Q_{r, \hat{k}_{\text{HQC}}(p), T}^*$ or $Q_{r, \hat{k}_{\text{AIC}}(p), T}^*$ do better. This is due to the well-known tendency of $\text{BIC}(k, p)$ to underestimate the lag length (cf. Table 2) which in turn affects the cointegration rank determination at the second step of the sequential procedure, as discussed in section 3 (see Table 1).¹³ In particular, we note that $Q_{r, \hat{k}_{\text{IC}}(p), T}^*$ tends to lose power when $\hat{k}_{\text{IC}}(p) < k_0$ in small samples. Overall, the performance of $Q_{r, \hat{k}_{\text{IC}}(p), T}^*$ is only marginally affected by either the presence of heteroskedasticity or the information criterion used to select the lag length in the first step of the sequential procedure. This highlights the fact that the sequential wild bootstrap method for cointegration rank determination is a robust procedure, especially when the sample size is large. Moreover, comparing the results in Tables 4.1-4.3, we note that $\text{BIC}(k, r)$ and $Q_{r, \hat{k}_{\text{BIC}}(p), T}^*$

¹³These results are in line with the simulation results obtained for standard Johansen's PLR test by Cheung and Lai (1993), Yap and Reinsel (1995), Haug (1996), and Saikkonen and Luukkonen (1997) which show that under-specifying the lag order leads to massive size distortions in the PLR test, while over-specifying the VAR model order is less problematic.

tend to complement one another when the sample size is small. For example, when $T = 100$ or 200 and $\gamma \leq 0.3$, $Q_{r, \hat{k}_{\text{BIC}(p), T}}^*$ delivers better results than the joint BIC-based approach when the shocks are GJR-GARCH(1,1) (Case A); but, on the other hand, $\text{BIC}(k, r)$ seems preferable for either autoregressive SV shocks (Case B) or a single volatility break (Case C). When $\gamma = 0.9$, i.e., the system is near-I(2), the wild bootstrap approach clearly outperforms $\text{BIC}(k, r)$ when T is small for all the cases considered.¹⁴

To conclude this section, it is interesting to compare the simulation results given in this paper with those given in CDRT for the case where the autoregressive lag length is known to the practitioner and, hence, does not need to be estimated from the data. Such a comparison reveals that in most cases rather little is lost in terms of finite sample efficacy to select the correct cointegration rank relative to the oracle (the known lag length case considered in CDRT) by estimating the autoregressive lag length as part of either a joint or sequential procedure. The greatest differences occur when both the sample size is small and the autoregressive coefficient is small but non-zero, such that lag selection methods have a tendency to underfit the lag length. These results are therefore very encouraging given that the lag length cannot be assumed known in practice.

6 Conclusions

In this paper we have analysed the asymptotic and finite sample properties of various methods for determining the cointegration rank in the context of a heteroskedastic VAR model whose autoregressive lag length is unknown. Joint and sequential estimation procedures based on the most widely used information criteria were considered, along with a two-step variant of the sequential approach of Johansen (1995) based on PLR or bootstrap PLR tests, using a lag length determined by an information criterion. The semi-parametric information criteria proposed by Cheng and Phillips (2009, 2012), which does not attempt to estimate the lag length, was also considered. We have established the conditions required to hold on the penalty functions such that the joint and sequential information criterion-based methods deliver consistent estimates of both the lag length and cointegration rank, showing that these coincide with the conditions required in the homoskedastic case. While the BIC and HQC penalties satisfy this condition, the AIC penalty does not. The same requirement on the penalty function was also shown to ensure that the two-step sequential approach consistently estimates the rank, provided wild bootstrap PLR tests are used in the sequential procedure when unconditional heteroskedasticity is present. Our results have therefore provided a formal justification for the two-step variant of Johansen's (1995) sequential procedure which is commonly used by practitioners, and shown that these methods hold under a very general class of heteroskedastic innovations.

¹⁴The condition $i_2 := |\alpha'_\perp \Gamma \beta_\perp| \neq 0$, which is required for X_t to be I(1), also provides a measure of the distance that the system is from being I(2). In our case, we have that $i_2 = (1 - \gamma)^{p-r} |\alpha'_\perp \beta_\perp|$. Thus, values of γ close to one represent near-I(2) systems which is a known problem case for cointegration rank tests (see, e.g., Johansen, 2002; CRT2012). The results in Tables 3.1-3.3 and 4.1-4.3 show that the wild bootstrap approach provides very decent finite sample performances in such cases.

We have also investigated the relative finite sample properties of the various methods outlined in the paper via a Monte Carlo simulation study allowing for both homoskedastic and heteroskedastic shocks. For our chosen simulation DGP, a 4-dimensional VAR(2) DGP model with adjustment speed of -0.4 , we observed a deterioration in the performance of all of the methods in the presence of either autoregressive stochastic volatility or a single volatility break in the innovation process, relative to their performance under stationary GARCH-type processes. These results suggest, at least within the context of the simulation DGP considered, the following conclusions: (i) an incorrect choice of the lag length significantly impacts on the efficacy of both PLR-based and information-based cointegration rank determination procedures in finite samples, highlighting the importance in finite samples of being able to accurately estimate the autoregressive lag order; (ii) semi-parametric criteria, based on fitting a VAR(1) to the data, can be substantially biased when the process is of higher-autoregressive order, especially so when the process is close to being integrated of order two, $I(2)$; (iii) a joint information-criterion based approach based on the BIC tends to outperform sequential information-criterion based methods and provides a useful complement to the wild bootstrap sequential procedure of CRT2010a,b (implemented with BIC-based lag selection). Moreover, the losses in finite sample efficacy from selecting the cointegration rank using either the joint BIC-based procedure or the Johansen (1995)-type sequential procedure based on wild bootstrap PLR tests and BIC selection of the lag length, as compared with the corresponding procedure based on knowledge of the true autoregressive lag length, appeared to be relatively small.

Supplementary material to this article is provided in “Online Supplement to ‘Determining the Cointegration Rank in Heteroskedastic VAR Models of Unknown Order’”, which is available at Cambridge Journals Online (journals.cambridge.org/ect).

References

- Akaike, H. (1974): A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19, 716–723.
- Aznar, A. and M. Salvador (2002): Selecting the rank of the cointegration space and the form of the intercept using an information criterion, *Econometric Theory* 18, 926–947.
- Boswijk, H.P., Cavaliere, G., A. Rahbek and A.M.R. Taylor (2016): Inference on co-integration parameters in heteroskedastic vector autoregressions, *Journal of Econometrics* 192, 64–85.
- Boswijk, H.P. and P.H. Franses (1992): Dynamic specification and cointegration, *Oxford Bulletin of Economics and Statistics* 54, 369–381.
- Camba-Mendez, G., G. Kapetanios, R.J. Smith and M.R. Weale (2003): Tests of rank in reduced rank regression models, *Journal of Business & Economic Statistics* 21, 145–155.

- Cavaliere, G., L. De Angelis, A. Rahbek and A.M.R. Taylor (2014): Determining the co-integration rank in heteroskedastic VAR models of unknown order, available at <http://www2.stat.unibo.it/deangelis/IC2-WP.pdf>.
- Cavaliere, G., L. De Angelis, A. Rahbek and A.M.R. Taylor (2015): A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic VAR models, *Oxford Bulletin of Economics and Statistics* 77, 106–128.
- 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 VAR 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 Bulletin of Economics and Statistics* 77, 740–759.
- Cheng, X. and P.C.B. Phillips (2009): Semiparametric cointegrating rank selection, *Econometrica Journal* 12, 83–104.
- Cheng, X. and P.C.B. Phillips (2012): Cointegrating rank selection in models with time-varying variance, *Journal of Econometrics* 169, 155–165.
- Cheung, Y.W. and K.S. Lai (1993): Finite-sample sizes of Johansen’s likelihood ratio tests for cointegration, *Oxford Bulletin of Economics and Statistics* 55, 313–328.
- Chao, J.C. and P.C.B. Phillips (1999): Model selection in partially nonstationary vector autoregressive processes with reduced rank structure, *Journal of Econometrics* 91, 227–272.
- Gonçalves, S. and L. Kilian (2004): Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* 123, 89–120.
- Hannan, E.J. and B.G. Quinn (1979): The determination of the order of an autoregression, *Journal of the Royal Statistical Society, Series B* 41, 190–195.
- Haug, A.A. (1996): Test for cointegration: A Monte Carlo comparison, *Journal of Econometrics* 71, 89–115.

- Johansen, S. (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.
- Johansen, S. (2002): A small sample correction of the test for cointegrating rank in the vector autoregressive model, *Econometrica* 70, 1929–1961.
- Kapetanios, G. (2000): Information criteria, model selection uncertainty and the determination of cointegration rank, NIESR Discussion Papers 166, National Institute of Economic and Social Research.
- Kapetanios, G. (2004): The asymptotic distribution of the cointegration rank estimator under the Akaike information criterion, *Econometric Theory* 20, 735–742.
- Kascha, C. and C. Trenkler (2011): Bootstrapping the likelihood ratio cointegration test in error correction models with unknown lag order, *Computational Statistics and Data Analysis* 55, 1008–1017.
- Lee, T.-H. and Y. Tse (1996): Cointegration tests with conditional heteroskedasticity, *Journal of Econometrics* 73, 401–410.
- Lütkepohl, H. and D. S. Poskitt (1998): Consistent estimation of the number of cointegration relations in a vector autoregressive model, in *Econometrics in Theory and Practice*, Galata, R. and H. Küchenhoff, (eds.), Heidelberg: Physica-Verlag, 87–100.
- Lütkepohl, H. and P. Saikkonen (1999): Order selection in testing for the cointegrating rank of a VAR process, In: *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Engle, R.F. and H. White (eds.), Oxford: Oxford University Press, 168–199
- Nielsen, B. (2006): Order determination in general vector autoregressions, *IMS Lecture Notes - Monograph Series* 52, 93–112.
- Paulsen, J. (1984): Order determination of multivariate autoregressive time series with unit roots, *Journal of Time Series Analysis* 5, 115–127.
- Pesaran, M.H., Y. Shin and R.J. Smith (2000): Structural analysis of vector error correction models with exogenous I(1) variables, *Journal of Econometrics* 97, 293–343.
- Pesaran, M.H., Y. Shin and R.J. Smith (2001): Bounds testing approaches to the analysis of long-run relationships, *Journal of Applied Econometrics* 16, 289–326.
- Phillips, P.C.B. (1996): Econometric model determination, *Econometrica* 64, 763–812.
- Phillips, P.C.B. and J.W. McFarland (1997): Forward exchange market unbiasedness: the case of the Australian dollar since 1984, *Journal of International Money and Finance* 16, 885–907.

- Phillips, P.C.B. and W. Ploberger (1996): An asymptotic theory of Bayesian inference for time series, *Econometrica* 64, 381–412.
- Rissanen, J. (1978): Modeling by shortest data description, *Automatica* 14, 465–471.
- J-M. Robin and R.J. Smith (2000): Tests of rank, *Econometric Theory* 16, 151–175.
- Saikkonen, P. and R. Luukkonen (1997): Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics* 81, 93–126
- Schwarz, G. (1978): Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.
- Swensen, A.R. (2006): Bootstrap algorithms for testing and determining the cointegration rank in VAR models, *Econometrica* 74, 1699–1714.
- Takeuchi, K. (1976): Distribution of informational statistics and a criterion of model fitting, *Suri-Kagaku* (in Japanese) 153, 12–18.
- van Giersbergen, N. P. A. (1996): Bootstrapping the trace statistic in VAR models: Monte Carlo results and applications, *Oxford Bulletin of Economics and Statistics* 58, 391–408.
- Yap, S.F. and G.C. Reinsel (1995): Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model, *Journal of the American Statistical Association* 90, 253–267.

A Appendix

A.1 Definitions and Preliminary Results

Throughout the appendix we will make use of the following notation: Weak convergence of A_T to B as $T \rightarrow \infty$ is denoted either by $A_T \xrightarrow{w} B$ or $B = \text{wlim } A_T$, $B = \text{plim } A_T$ means that A_T converges in probability to B as $T \rightarrow \infty$ and $0_{j \times k}$ is used to denote the $j \times k$ matrix of zeroes. If a is of full column rank $n < m$, then $\bar{a} := a(a'a)^{-1}$ and a_{\perp} is an $m \times (m - n)$ full column rank matrix satisfying $a'_{\perp}a = 0$; for any square matrix, A , $\|A\|$ denotes the norm $\|A\|^2 := \text{tr}\{A'A\}$.

Without loss of generality, in what follows we assume that there is no deterministic component (case (i) in (2.1)), i.e., $D_t = 0$ and $d_t = 0$. Furthermore, we will repeatedly refer to results in Boswijk, Cavaliere, Rahbek and Taylor (2016) [BCRT, hereafter]. We first need the following lemma where we extend the results in Theorem 11.1 in Johansen (1995) and Lemma 3.2 in Cheng and Phillips (2009) to the case of a general $k < k_0$ and to the case of heteroskedastic innovations. In particular, we establish the asymptotic behaviour of the eigenvalues in the determinantal problem (3.2) when the number of lags included is fewer than the true lag length. The case where $k \geq k_0$ is discussed in Remark A.1 below.

As in Lemma 2 in BCRT, we define, in terms of X_t given by (2.1),

$$\begin{aligned}\Omega_{\beta\beta} &:= \text{plim} \frac{1}{T} \sum_{t=1}^T \beta' X_{t-1} X'_{t-1} \beta, \quad \Omega_{00} := \text{plim} \frac{1}{T} \sum_{t=1}^T \Delta X_t \Delta X'_t \\ \Omega_{\beta 0} &:= \text{plim} \frac{1}{T} \sum_{t=1}^T \beta' X_{t-1} \Delta X'_t =: \Omega'_{0\beta}, \quad \Omega_{\beta 2}^{(k)} := \text{plim} \frac{1}{T} \sum_{t=1}^T \beta' X_{t-1} Z_{2t}^{(k)'} =: \Omega_{2\beta}^{(k)'} \\ \Omega_{02}^{(k)} &:= \text{plim} \frac{1}{T} \sum_{t=1}^T \Delta X_t Z_{2t}^{(k)'} =: \Omega_{20}^{(k)'}, \quad \Omega_{22}^{(k)} := \text{plim} \frac{1}{T} \sum_{t=1}^T Z_{2t} Z_{2t}^{(k)'}\end{aligned}$$

where $Z_{2t}^{(k)} := (\Delta X'_{t-1} \dots \Delta X'_{t-k+1})'$. Moreover, let $\bar{\Sigma}_{ij}^{(k)} := \Omega_{ij} - \Omega_{i2}^{(k)} \Omega_{22}^{(k)-1} \Omega_{2j}^{(k)}$, for $i, j = 0, \beta$.

Furthermore, let $G(\cdot) := \beta'_\perp CM(\cdot)$ be a $(p-r)$ -dimensional process where, under Assumption \mathcal{H} , $M(\cdot) := \int_0^\cdot \sigma(s) dW(s) = \text{wlim} T^{-1/2} \sum_{t=1}^{\lfloor T \cdot \rfloor} \varepsilon_t$ is a p -variate continuous martingale (see Lemma 1 in BCRT) and $C := \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$ (here $\Gamma := I_p - \sum_{j=1}^{k_0-1} \Gamma_j$). Finally, let $Z_t = \Delta X_t - \Omega_{02}^{(k)} \Omega_{22}^{(k)-1} Z_{2t}^{(k)}$ be a linear process of the form $Z_t = C^*(L) \varepsilon_t$, where $C^*(L)$ is defined in equation (A.7) below.

Lemma A.1 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions. Then, for any $0 < k < k_0$, as $T \rightarrow \infty$, the r_0 largest solutions to (3.2), denoted by $\hat{\lambda}_i^{(k)}$, $i = 1, \dots, r_0$, converge to the nonzero roots of $|\lambda \bar{\Sigma}_{\beta\beta}^{(k)} - \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)}| = 0$, with $\bar{\Sigma}_{\beta\beta}^{(k)} > 0$. The remaining $p - r_0$ roots, denoted by $\hat{\lambda}_i^{(k)}$, for $i = r_0 + 1, \dots, p$, are of order $O_p(T^{-1})$ as $T \hat{\lambda}_i^{(k)}$, for $i = r_0 + 1, \dots, p$, converge to the roots of the equation*

$$\left| \rho \int_0^1 G G' du - \left(\int_0^1 G dM' C^*(1)' + \Lambda_{\beta_\perp z} \right) N \left(\Lambda'_{\beta_\perp z} + C^*(1) \int_0^1 dM G' \right) \right| = 0 \quad (\text{A.1})$$

where $\Lambda_{\beta_\perp z} := \sum_{h=1}^\infty \text{Cov}(\beta'_\perp \Delta X_t, Z_{t+h})$, $N := \bar{\Sigma}_{00}^{(k)-1} - \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)} (\bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)})^{-1} \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1}$ and $C^*(1) \neq 0$.

Remark A.1: Observe that, unlike in the equation given in (A.1), it seems that the results (provided only for the special case where ε_t in (2.1) is i.i.d.) by Aznar and Salvador (2002, p. 938), six lines above equation (A.2) in the proof of their Lemma 4.1, do not appear to take account of the fact that the CVAR model associated with (3.2) is misspecified whenever $k < k_0$.

Proof of Lemma A.1: Let $S^{(k)}(\lambda) := \lambda S_{11}^{(k)} - S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)}$ so that the determinantal problem in (3.2) is $|S^{(k)}(\lambda)| = 0$, where $S_{ij}^{(k)} = M_{ij} - M_{i2}^{(k)} M_{22}^{(k)-1} M_{2j}^{(k)}$, for $i, j = 0, 1$, with

$$\begin{aligned}M_{00} &= \frac{1}{T} \sum_{t=1}^T \Delta X_t \Delta X'_t, & M_{01} &= \frac{1}{T} \sum_{t=1}^T \Delta X_t X'_{t-1}, & M_{11} &= \frac{1}{T} \sum_{t=1}^T X_{t-1} X'_{t-1}, \\ M_{22}^{(k)} &= \frac{1}{T} \sum_{t=1}^T Z_{2t}^{(k)} Z_{2t}^{(k)'}, & M_{02}^{(k)} &= \frac{1}{T} \sum_{t=1}^T \Delta X_t Z_{2t}^{(k)'}, & M_{12}^{(k)} &= \frac{1}{T} \sum_{t=1}^T X_{t-1} Z_{2t}^{(k)'}. \end{aligned}$$

We first show that the r_0 largest solution to $|S^{(k)}(\lambda)| = 0$ converge to the roots of $|\lambda \bar{\Sigma}_{\beta\beta}^{(k)} - \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)}| = 0$. Let $A_T := (\beta, \beta_\perp / \sqrt{T})$. We have that

$$|A'_T S^{(k)}(\lambda) A_T| =$$

$$\left| \begin{pmatrix} \lambda \beta' S_{11}^{(k)} \beta & \frac{1}{\sqrt{T}} \lambda \beta' S_{11}^{(k)} \beta_{\perp} \\ \frac{1}{\sqrt{T}} \lambda \beta'_{\perp} S_{11}^{(k)} \beta & \frac{1}{T} \lambda \beta'_{\perp} S_{11}^{(k)} \beta_{\perp} \end{pmatrix} - \begin{pmatrix} \beta' S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta & \frac{1}{\sqrt{T}} \beta' S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta_{\perp} \\ \frac{1}{\sqrt{T}} \beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta & \frac{1}{T} \beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta_{\perp} \end{pmatrix} \right|. \quad (\text{A.2})$$

We show that (A.2) converges to

$$\left| \begin{pmatrix} \lambda \bar{\Sigma}_{\beta\beta}^{(k)} & 0 \\ 0 & \lambda \int_0^1 GG' du \end{pmatrix} - \begin{pmatrix} \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)} & 0 \\ 0 & 0 \end{pmatrix} \right| = \left| \lambda \bar{\Sigma}_{\beta\beta}^{(k)} - \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)} \right| \left| \lambda \int_0^1 GG' du \right|$$

where $\bar{\Sigma}_{\beta\beta}^{(k)} > 0$. Therefore, in the limit the determinantal equation has $p - r_0$ roots equal to zero and r_0 positive roots, the latter given by the solutions of $\left| \lambda \bar{\Sigma}_{\beta\beta}^{(k)} - \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)} \right| = 0$.

Specifically, since X_t satisfies the $I(1, r_0)$ conditions with $\beta' X_t$, ΔX_t and $Z_{2t}^{(k)}$ being $I(0)$, we have that

$$\beta' S_{11}^{(k)} \beta = \beta' M_{11} \beta - \beta' M_{12}^{(k)} M_{22}^{(k)-1} M_{21}^{(k)} \beta \xrightarrow{p} \Omega_{\beta\beta} - \Omega_{\beta 2}^{(k)} \Omega_{22}^{(k)-1} \Omega_{2\beta}^{(k)} =: \bar{\Sigma}_{\beta\beta}^{(k)} \quad (\text{A.3})$$

and

$$\beta' S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta \xrightarrow{p} \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)}. \quad (\text{A.4})$$

Moreover, since by the Granger representation theorem and Lemma 1 in BCRT

$$\frac{1}{\sqrt{T}} \beta'_{\perp} X_{[T \cdot]} \xrightarrow{w} \beta'_{\perp} CM(\cdot) = G(\cdot),$$

we have that

$$\frac{1}{T} \beta'_{\perp} S_{11}^{(k)} \beta_{\perp} \xrightarrow{w} \beta'_{\perp} C \left(\int_0^1 MM' du \right) C' \beta_{\perp} = \int_0^1 GG' du. \quad (\text{A.5})$$

Finally, we show that $\frac{1}{\sqrt{T}} \beta'_{\perp} S_{11}^{(k)} \beta$, $\frac{1}{\sqrt{T}} \beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta$, and $\frac{1}{T} \beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta$ converge in probability to zero. In particular, we first show that

$$\beta'_{\perp} S_{11}^{(k)} \beta = \beta'_{\perp} M_{11} \beta - \beta'_{\perp} M_{12}^{(k)} M_{22}^{(k)-1} M_{21}^{(k)} \beta = O_p(1). \quad (\text{A.6})$$

By the Granger representation theorem (see Lemma 1 in CRT2010b), we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T \cdot]} \beta' X_t \xrightarrow{w} -\beta' P(\mathbb{B}' \mathbb{A})^{-1} \mathbb{B}' E_1 M(\cdot) = H' M(\cdot)$$

where

$$\mathbb{A} := \begin{bmatrix} \alpha & \Gamma_1 & \dots & \Gamma_{k_0-1} \\ 0 & I_p & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_p \end{bmatrix}, \quad \mathbb{B} := \begin{bmatrix} \beta & I_p & 0 & \dots & 0 \\ 0 & -I_p & I_p & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & I_p \\ 0 & 0 & 0 & \dots & -I_p \end{bmatrix},$$

with $P := (I_p, 0, \dots, 0)' \mathbb{A} (\mathbb{B}' \mathbb{A})^{-1}$, $E_1 := (I_p, 0, \dots, 0)'$, and $H := -(\beta' P(\mathbb{B}' \mathbb{A})^{-1} \mathbb{B}' E_1)'$. Therefore,

$$\beta'_{\perp} M_{11} \beta = \frac{1}{T} \sum_{t=1}^T \beta'_{\perp} X_{t-1} X'_{t-1} \beta = \frac{1}{T} \sum_{t=1}^T \beta'_{\perp} \left(\sum_{i=1}^{t-1} \Delta X_i \right) X'_{t-1} \beta \xrightarrow{w} \int_0^1 G(dM)' H + \Lambda_{\beta_{\perp} \beta}$$

where $\Lambda_{\beta_{\perp}\beta} := \sum_{h=0}^{\infty} \text{Cov}(\beta'_{\perp} \Delta X_t, \beta' X_{t+h})$, which shows that $\beta'_{\perp} M_{11} \beta$ is $O_p(1)$. Moreover, since $M_{22}^{(k)-1} M_{21}^{(k)} \beta \xrightarrow{p} \Omega_{22}^{(k)-1} \Omega_{2\beta}^{(k)} =: V$ and

$$\beta'_{\perp} M_{12} = \frac{1}{T} \sum_{t=1}^T \beta'_{\perp} X_{t-1} \Delta X'_t = \frac{1}{T} \sum_{t=1}^T \beta'_{\perp} \left(\sum_{i=1}^{t-1} \Delta X_i \right) \Delta X'_t \xrightarrow{w} \int_0^1 G(dM)' + \Lambda_{\beta_{\perp}\Delta}$$

where $\Lambda_{\beta_{\perp}\Delta} := \sum_{h=0}^{\infty} \text{Cov}(\beta'_{\perp} \Delta X_t, \Delta X_{t+h})$, we have that $\beta'_{\perp} M_{12}^{(k)} M_{22}^{(k)-1} M_{21}^{(k)} \beta$ converges in distribution to $\int_0^1 G(dM)' V + \Lambda_{\beta_{\perp}\Delta} V$. Consequently, (A.6) holds. Analogously, we show that $\beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta = O_p(1)$. Specifically,

$$\begin{aligned} \beta'_{\perp} S_{10}^{(k)} &= \beta'_{\perp} M_{10} - \beta'_{\perp} M_{12}^{(k)} M_{22}^{(k)-1} M_{20}^{(k)} \\ &= \frac{1}{T} \sum_{t=1}^T \beta'_{\perp} X_{t-1} \Delta X'_t - \frac{1}{T} \sum_{t=1}^T \beta'_{\perp} X_{t-1} Z_{2t}^{(k)'} \left(\frac{1}{T} \sum_{t=1}^T Z_{2t}^{(k)} Z_{2t}^{(k)'} \right)^{-1} \frac{1}{T} \sum_{t=1}^T Z_{2t}^{(k)} \Delta X'_t \\ &= \frac{1}{T} \sum_{t=1}^T \beta'_{\perp} X_{t-1} \left(\Delta X'_t - Z_{2t}^{(k)'} \Omega_{22}^{(k)-1} \Omega_{20}^{(k)} \right) + o_p(1) \end{aligned}$$

where, under the $I(1, r_0)$ conditions, $Z_t := \Delta X_t - \Omega_{02}^{(k)} \Omega_{22}^{(k)-1} Z_{2t}^{(k)}$ is a linear process of the form $Z_t = C^*(L) \varepsilon_t$. In particular, let $D := \Omega_{02}^{(k)} \Omega_{22}^{(k)-1}$ with $D = (D_1, \dots, D_{k-1})$ where D_i , for $i = 1, \dots, k-1$, is a $p \times p$ matrix. Then,

$$\begin{aligned} Z_t &= \Delta X_t - D Z_{2t}^{(k)} = (I_p - D_1 L - \dots - D_{k-1} L^{k-1}) \Delta X_t \\ &= D(L) \Delta X_t = D(L) C(L) \varepsilon_t =: C^*(L) \varepsilon_t \end{aligned} \quad (\text{A.7})$$

where $D(L) := (I_p - D_1 L - \dots - D_{k-1} L^{k-1})$. Therefore,

$$\beta'_{\perp} S_{10}^{(k)} \xrightarrow{w} \int_0^1 G dM' C^*(1)' + \Lambda_{\beta_{\perp}z} \quad (\text{A.8})$$

where $C^*(1) = D(1)C(1)$, $\Lambda_{\beta_{\perp}z} := \sum_{h=1}^{\infty} \text{Cov}(\beta'_{\perp} \Delta X_t, Z_{t+h})$ and $C(1) \neq 0$ follows from ΔX_t being $I(0)$. Moreover, since $S_{00}^{(k)-1} S_{01}^{(k)} \beta \xrightarrow{p} \bar{S}_{00}^{(k)-1} \bar{S}_{0\beta}^{(k)} =: J$, we have that $\beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta$ converges in distribution to $\int_0^1 G dM' C^*(1)' J + \Lambda_{\beta_{\perp}z} J$ and, consequently, is of order $O_p(1)$. By the same arguments, we also have that $\beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta_{\perp} = O_p(1)$.

We now show that the remaining $p - r_0$ roots are of order $O_p(T^{-1})$. Let $B := (\beta, \beta_{\perp})$. We have that

$$\begin{aligned} |B' S^{(k)}(\lambda) B| &= \left| \begin{pmatrix} \beta' S^{(k)}(\lambda) \beta & \beta' S^{(k)}(\lambda) \beta_{\perp} \\ \beta'_{\perp} S^{(k)}(\lambda) \beta & \beta'_{\perp} S^{(k)}(\lambda) \beta_{\perp} \end{pmatrix} \right| \\ &= \left| \beta' S^{(k)}(\lambda) \beta \right| \left| \beta'_{\perp} \left\{ S^{(k)}(\lambda) - S^{(k)}(\lambda) \beta \left[\beta' S^{(k)}(\lambda) \beta \right]^{-1} \beta' S^{(k)}(\lambda) \right\} \beta_{\perp} \right|. \end{aligned} \quad (\text{A.9})$$

As in Theorem 11.1 in Johansen (1995), let $T \rightarrow \infty$ with $\rho := T\lambda$ fixed. The first term in (A.9) is

$$\beta' S^{(k)}(\lambda) \beta = \frac{\rho}{T} \beta' S_{11}^{(k)} \beta - \beta' S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta = -\bar{S}_{\beta 0}^{(k)} \bar{S}_{00}^{(k)-1} \bar{S}_{0\beta}^{(k)} + o_p(1) \quad (\text{A.10})$$

since $T^{-1} \beta' S_{11}^{(k)} \beta \xrightarrow{p} 0$ (see (A.3) and (A.4)). Moreover,

$$\beta'_{\perp} S^{(k)}(\lambda) \beta_{\perp} = \frac{\rho}{T} \beta'_{\perp} S_{11}^{(k)} \beta_{\perp} - \beta'_{\perp} S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)} \beta_{\perp} \quad (\text{A.11})$$

and

$$\beta'_\perp S^{(k)}(\lambda)\beta = \frac{\rho}{T}\beta'_\perp S_{11}^{(k)}\beta - \beta'_\perp S_{10}^{(k)}S_{00}^{(k)-1}S_{01}^{(k)}\beta = -\beta'_\perp S_{10}^{(k)}S_{00}^{(k)-1}S_{01}^{(k)}\beta + o_p(1) \quad (\text{A.12})$$

since $\beta'_\perp S_{11}^{(k)}\beta = O_p(1)$ (see (A.6)). By substituting (A.11) and (A.12) into the second factor in (A.9) we obtain

$$\beta'_\perp \left\{ S^{(k)}(\lambda) - S^{(k)}(\lambda)\beta \left[\beta' S^{(k)}(\lambda)\beta \right]^{-1} \beta' S^{(k)}(\lambda) \right\} \beta_\perp = \frac{\rho}{T}\beta'_\perp S_{11}^{(k)}\beta_\perp - \beta'_\perp S_{10}^{(k)}N_T S_{01}^{(k)}\beta_\perp + o_p(1)$$

where $N_T = S_{00}^{(k)-1} - S_{00}^{(k)-1}S_{01}^{(k)}\beta(\beta' S_{10}^{(k)}S_{00}^{(k)-1}S_{01}^{(k)}\beta)^{-1}\beta' S_{10}^{(k)}S_{00}^{(k)-1}$. The distribution of the $p - r_0$ smallest solutions of (3.2) can be derived using the result in (A.5) and noting that, by the result in (A.4), $N_T = N + o_p(1)$, where $N := \bar{\Sigma}_{00}^{(k)-1} - \bar{\Sigma}_{00}^{(k)-1}\bar{\Sigma}_{0\beta}^{(k)}(\bar{\Sigma}_{\beta 0}^{(k)}\bar{\Sigma}_{00}^{(k)-1}\bar{\Sigma}_{0\beta}^{(k)})^{-1}\bar{\Sigma}_{\beta 0}^{(k)}\bar{\Sigma}_{00}^{(k)-1}$. Therefore, we have that $\beta'_\perp S_{10}^{(k)}N_T S_{01}^{(k)}\beta_\perp = \beta'_\perp S_{10}^{(k)}N S_{01}^{(k)}\beta_\perp + o_p(1)$, where $\beta'_\perp S_{10}^{(k)} \xrightarrow{w} \int_0^1 G dM' C^*(1)' + \Lambda_{\beta_\perp z}$ (see (A.8)).

Taken together, these results imply that the $p - r_0$ smallest solutions of (3.2), normalised by T , converge to those of the equation

$$\left| \rho \int_0^1 G G' du - \left(\int_0^1 G dM' C^*(1)' + \Lambda_{\beta_\perp z} \right) N \left(\Lambda'_{\beta_\perp z} + C^*(1) \int_0^1 dM G' \right) \right| = 0$$

and hence that $\hat{\lambda}_i^{(k)}$, $i = 1, \dots, p - r_0$, are all of order $O_p(T^{-1})$. \square

Remark A.2: For the case of $k_0 < k < K$, note that asymptotic theory for the estimated eigenvalues $\hat{\lambda}_i$, as well as the parameter estimators of α , β and Γ_i ($i = 1, \dots, k - 1$), is standard as the parameter point at which we evaluate the asymptotic behaviour satisfies the $I(1, r_0)$ conditions for the k -lag model.

A.2 Proof of Theorem 1 and Related Lemmas

Proof of Lemma 1: We now show that, for any fixed value of k , with $k = 1, \dots, K$, where K denotes a given maximum number of lags, $\text{IC}(k, r)$ is weakly consistent under Assumption \mathcal{H} provided that $c_T \rightarrow \infty$ and $c_T/T \rightarrow 0$, as $T \rightarrow \infty$.

(i) For any $r > r_0$, we have that

$$\text{IC}(k, r) - \text{IC}(k, r_0) = T \sum_{i=r_0+1}^r \log(1 - \hat{\lambda}_i^{(k)}) + c_T(2p - r - r_0)(r - r_0) \quad (\text{A.13})$$

where, by Lemma A.1 and Remark A.1 for the cases of $0 < k < k_0$ and $k_0 \leq k \leq K$, respectively, the first term on the right hand side of (A.13) satisfies $T \sum_{i=r_0+1}^r \log(1 - \hat{\lambda}_i^{(k)}) = O_p(1)$. As $(2p - r - r_0)(r - r_0) > 0$ and $c_T \rightarrow \infty$, we conclude that $P(\text{IC}(k, r) - \text{IC}(k, r_0) > 0) \rightarrow 1$ as $T \rightarrow \infty$.

(ii) For any $r < r_0$ (and $r_0 \neq 0$), we have that

$$\text{IC}(k, r) - \text{IC}(k, r_0) = -T \sum_{i=r+1}^{r_0} \log(1 - \hat{\lambda}_i^{(k)}) + c_T(2p - r - r_0)(r - r_0). \quad (\text{A.14})$$

Observe that, by Lemma A.1 and Remark A.1, the r_0 largest sample eigenvalues $\hat{\lambda}_i^{(k)}$, $i = 1, \dots, r_0$ converge to the roots $\lambda_i^{(k)}$ of the equation $\left| \lambda \bar{\Sigma}_{\beta\beta}^{(k)} - \bar{\Sigma}_{\beta 0}^{(k)} \bar{\Sigma}_{00}^{(k)-1} \bar{\Sigma}_{0\beta}^{(k)} \right| = 0$. As $\bar{\Sigma}_{\beta\beta}^{(k)}, \bar{\Sigma}_{00}^{(k)} > 0$, and all the

$\lambda_i^{(k)}$ are positive and, by definition, smaller than one, we have that $-T \sum_{i=r+1}^{r_0} \log(1 - \hat{\lambda}_i^{(k)})$ diverges at rate T . Moreover, since by assumption $c_T/T \rightarrow 0$ as $T \rightarrow \infty$, we have that the first term on the right hand side of (A.14) dominates and, hence, $P(\text{IC}(k, r) - \text{IC}(k_0, r_0) > 0) \rightarrow 1$. \square

Proof of Lemma 2: (i) The result for $k > k_0$ follows by showing that, for any $k_0 < k \leq K$,

$$\hat{\Sigma}_{k,r_0} - \hat{\Sigma}_{k_0,r_0} = O_p(T^{-1}). \quad (\text{A.15})$$

In particular, (A.15) implies that, when $c_T \rightarrow \infty$,

$$\begin{aligned} \text{IC}(k, r_0) - \text{IC}(k_0, r_0) &= T \log \det(\hat{\Sigma}_{k_0,r_0}^{-1} \hat{\Sigma}_{k,r_0}) + c_T(\pi(k, r_0) - \pi(k_0, r_0)) \\ &= T \log \det(I_p - \hat{\Sigma}_{k_0,r_0}^{-1} (\hat{\Sigma}_{k_0,r_0} - \hat{\Sigma}_{k,r_0})) + c_T(\pi(k, r_0) - \pi(k_0, r_0)) \\ &= T \log \det(I_p + O_p(T^{-1})) + c_T(\pi(k, r_0) - \pi(k_0, r_0)) \\ &= O_p(1) + c_T(\pi(k, r_0) - \pi(k_0, r_0)) \xrightarrow{p} \infty \end{aligned}$$

since $\pi(k, r_0) - \pi(k_0, r_0) > 0$.

In order to prove (A.15), for any $k_0 \leq k \leq K$, let $\theta_0^{(k_0)} := (\alpha, \beta, \Gamma_1, \dots, \Gamma_{k_0-1})$ and $\theta_0^{(k)} := (\alpha, \beta, \Gamma_1, \dots, \Gamma_{k_0-1}, 0, \dots, 0)$ be the true parameter of the $\text{VAR}(k_0)$ model and the $\text{VAR}(k)$ model, respectively, and let $\hat{\theta}^{(k_0)} = (\hat{\alpha}^{(k_0)}, \hat{\beta}^{(k_0)}, \hat{\Gamma}_1^{(k_0)}, \dots, \hat{\Gamma}_{k_0-1}^{(k_0)})$ and $\hat{\theta}^{(k)} = (\hat{\alpha}^{(k)}, \hat{\beta}^{(k)}, \hat{\Gamma}_1^{(k)}, \dots, \hat{\Gamma}_{k_0-1}^{(k)}, \dots, \hat{\Gamma}_{k-1}^{(k)})$ denote the corresponding estimated parameters. For the estimators in the $\text{VAR}(k_0)$ -based reduced rank regression, it holds that $\sqrt{T}((\hat{\alpha}^{(k_0)}, \hat{\Gamma}_1^{(k_0)}, \dots, \hat{\Gamma}_{k_0-1}^{(k_0)}) - (\alpha, \Gamma_1, \dots, \Gamma_{k_0-1}))$ is asymptotically normal distributed, while $(\hat{\beta}^{(k_0)} - \beta)$, suitably normalised, is asymptotically mixed Gaussian (see Theorem 1 in BCRT). Moreover, under the $\text{I}(1, r_0)$ conditions, we also have for the $\text{VAR}(k)$ -based reduced rank regression estimators that, as $T \rightarrow \infty$, $\sqrt{T}((\hat{\alpha}^{(k_0)}, \hat{\Gamma}_1^{(k_0)}, \dots, \hat{\Gamma}_{k_0-1}^{(k_0)}, \dots, \hat{\Gamma}_{k-1}^{(k_0)}) - (\alpha, \Gamma_1, \dots, \Gamma_{k_0-1}, 0, \dots, 0))$ is normal and $(\hat{\beta}^{(k)} - \beta)$, suitably normalised, is mixed Gaussian. Now consider the likelihood ratio (LR) statistic

$$Q_T = T \log \det(\hat{\Sigma}_{k_0,r_0}^{-1} \hat{\Sigma}_{k,r_0})$$

which may be written as

$$Q_T = Q_T^{(k_0)} - Q_T^{(k)}$$

where $Q_T^{(k_0)}$ and $Q_T^{(k)}$ are the LR statistics for the simple hypothesis $\theta^{(k_0)} = \theta_0^{(k_0)}$ and $\theta^{(k)} = \theta_0^{(k)}$, respectively. For each we have,

$$\begin{aligned} Q_T^{(k_0)} &= T \log \det(I_p - \hat{\Sigma}_{k_0,r_0}^{-1} (\hat{\Sigma}_{k_0,r_0} - \tilde{\Sigma}_{k_0,r_0})) \\ Q_T^{(k)} &= T \log \det(I_p - \hat{\Sigma}_{k,r_0}^{-1} (\hat{\Sigma}_{k,r_0} - \tilde{\Sigma}_{k,r_0})) \end{aligned}$$

where $\tilde{\Sigma}_{k_0,r_0} = \tilde{\Sigma}_{k,r_0} = \frac{1}{T} \sum_{t=1}^T (\Delta X_t - \theta_0^{(k_0)} (X'_{t-1}, Z_{2t}^{(k_0)})')' (\Delta X_t - \theta_0^{(k_0)} (X'_{t-1}, Z_{2t}^{(k_0)})')'$ is the sample variance with the true parameters inserted. Because of the aforementioned convergence rates for the estimators $\hat{\theta}^{(k_0)}$ and $\hat{\theta}^{(k)}$, it holds that $T(\hat{\Sigma}_{k_0,r_0} - \tilde{\Sigma}_{k_0,r_0}) = U'U + o_p(1)$ and $T(\hat{\Sigma}_{k,r_0} - \tilde{\Sigma}_{k,r_0}) = U'MU + o_p(1)$, where U is a suitable Gaussian vector and M is a projection matrix, so that $Q_T^{(k_0)}$ and $Q_T^{(k)}$ are of order $O_p(1)$ and, consequently, Q_T is $O_p(1)$, which proves (A.15).

(ii) As in Paulsen (1984), the result for $0 < k < k_0$ follows by showing that $\hat{\Sigma}_{k,r_0} - \hat{\Sigma}_{k_0,r_0}$ converges to a positive definite matrix. In particular, as in Theorem 1 in BCRT, we have that

$$\hat{\Sigma}_{k_0,r_0} \xrightarrow{p} \bar{\Sigma}_{00}^{(k_0)} - \bar{\Sigma}_{0\beta}^{(k_0)} \bar{\Sigma}_{\beta\beta}^{(k_0)-1} \bar{\Sigma}_{\beta 0}^{(k_0)} =: \bar{\Sigma}^{(k_0)} \quad (\text{A.16})$$

and

$$\hat{\Sigma}_{k,r_0} = S_{00}^{(k)} - S_{01}^{(k)} \beta (\beta' S_{11}^{(k)} \beta)^{-1} \beta' S_{10}^{(k)} \xrightarrow{p} \bar{\Sigma}_{00}^{(k)} - \bar{\Sigma}_{0\beta}^{(k)} \bar{\Sigma}_{\beta\beta}^{(k)-1} \bar{\Sigma}_{\beta 0}^{(k)} \quad (\text{A.17})$$

where $\bar{\Sigma}_{ij}^{(k)}$, for $i, j = 0, \beta$, are defined above. Thus, (A.16) and (A.17) imply that

$$\hat{\Sigma}_{k,r_0} - \hat{\Sigma}_{k_0,r_0} \xrightarrow{p} (\bar{\Sigma}_{00}^{(k)} - \bar{\Sigma}_{0\beta}^{(k)} \bar{\Sigma}_{\beta\beta}^{(k)-1} \bar{\Sigma}_{\beta 0}^{(k)}) - (\bar{\Sigma}_{00}^{(k_0)} - \bar{\Sigma}_{0\beta}^{(k_0)} \bar{\Sigma}_{\beta\beta}^{(k_0)-1} \bar{\Sigma}_{\beta 0}^{(k_0)})$$

which is positive definite since $\bar{\Sigma}_{00}^{(k)} - \bar{\Sigma}_{0\beta}^{(k)} \bar{\Sigma}_{\beta\beta}^{(k)-1} \bar{\Sigma}_{\beta 0}^{(k)} > \bar{\Sigma}^{(k_0)}$. Therefore that, when $c_T/T \rightarrow 0$,

$$\begin{aligned} \text{IC}(k, r_0) - \text{IC}(k_0, r_0) &= T \log \det \left(\hat{\Sigma}_{k,r_0} \hat{\Sigma}_{k_0,r_0}^{-1} \right) + c_T (\pi(k, r_0) - \pi(k_0, r_0)) \\ &= T \log \det \left(I_p + (\hat{\Sigma}_{k,r_0} - \hat{\Sigma}_{k_0,r_0}) \hat{\Sigma}_{k_0,r_0}^{-1} \right) + c_T (\pi(k, r_0) - \pi(k_0, r_0)) \xrightarrow{p} \infty \end{aligned}$$

since the first term on the right side diverges and dominates the second term. \square

Remark A.3. Note that the proof of equation (A.15) reflects the well-known result that the (LR) test of the hypothesis that $k - k_0$ lags equal to zero in a VAR(k) model with known cointegration rank r_0 is asymptotically distributed as a χ^2 with $k - k_0$ degrees of freedom.

Proof of Theorem 1: In Lemmas 1 and 2, respectively, we have shown that

$$\text{P}(\hat{r}_{\text{IC}}(k) = r_0) \rightarrow 1, \text{ for any } k = 1, \dots, K \quad (\text{A.18})$$

and

$$\text{P}(\hat{k}_{\text{IC}}(r_0) = k_0) \rightarrow 1. \quad (\text{A.19})$$

Consider next the fact that

$$\text{P}(\tilde{k}_{\text{IC}} = k_0, \tilde{r}_{\text{IC}} = r_0) = \text{P}(\tilde{r}_{\text{IC}} = r_0) \text{P}(\tilde{k}_{\text{IC}} = k_0 \mid \tilde{r}_{\text{IC}} = r_0), \quad (\text{A.20})$$

where, for T large enough, $\text{P}(\tilde{r}_{\text{IC}} = r_0) \neq 0$. First, we prove that

$$\text{P}(\tilde{r}_{\text{IC}} = r_0) \rightarrow 1.$$

Because $\tilde{k}_{\text{IC}} \in \{1, \dots, K\}$, observe that (3.8) implies that

$$\text{P}(\tilde{r}_{\text{IC}} = r_0) = \text{P}(\hat{r}_{\text{IC}}(\tilde{k}_{\text{IC}}) = r_0) \geq \min_{k=1, \dots, K} \text{P}(\hat{r}_{\text{IC}}(k) = r_0)$$

where, from (A.18), we have that $\text{P}(\hat{r}_{\text{IC}}(k) = r_0) \rightarrow 1$ for any $k = 1, \dots, K$. Therefore, $\text{P}(\tilde{r}_{\text{IC}} = r_0) = \text{P}(\hat{r}_{\text{IC}}(\tilde{k}_{\text{IC}}) = r_0) \rightarrow 1$.

To complete the proof we now show that the second term in (A.20) tends to 1, i.e.,

$$\text{P}(\tilde{k}_{\text{IC}} = k_0 \mid \tilde{r}_{\text{IC}} = r_0) \rightarrow 1.$$

To prove this fact, notice first that (3.9) implies $P(\tilde{k}_{IC} = k_0 \mid \tilde{r}_{IC} = r_0) = P(\hat{k}_{IC}(\tilde{r}_{IC}) = k_0 \mid \tilde{r}_{IC} = r_0) = P(\hat{k}_{IC}(r_0) = k_0 \mid \tilde{r}_{IC} = r_0)$. Then, by using the law of total probability as follows

$$\begin{aligned} P(\hat{k}_{IC}(r_0) = k_0) &= P(\hat{k}_{IC}(r_0) = k_0 \mid \tilde{r}_{IC} = r_0)P(\tilde{r}_{IC} = r_0) + \\ &+ P(\hat{k}_{IC}(r_0) = k_0 \mid \tilde{r}_{IC} \neq r_0)P(\tilde{r}_{IC} \neq r_0), \end{aligned}$$

we obtain

$$P(\hat{k}_{IC}(r_0) = k_0 \mid \tilde{r}_{IC} = r_0) = \frac{P(\hat{k}_{IC}(r_0) = k_0)}{P(\tilde{r}_{IC} = r_0)} - \frac{P(\hat{k}_{IC}(r_0) = k_0 \mid \tilde{r}_{IC} \neq r_0)P(\tilde{r}_{IC} \neq r_0)}{P(\tilde{r}_{IC} = r_0)}. \quad (\text{A.21})$$

As $T \rightarrow \infty$, (A.18) and (A.19) imply that the first term on the right side of (A.21) tends to 1, while the second term tends to 0, since $P(\tilde{r}_{IC} \neq r_0) \rightarrow 0$. Therefore, $P(\hat{k}_{IC}(r_0) = k_0 \mid \tilde{r}_{IC} = r_0) = P(\tilde{k}_{IC} = k_0 \mid \tilde{r} = r_0) \rightarrow 1$. Together with $P(\tilde{r}_{IC} = r_0) \rightarrow 1$, this proves the result that $P(\tilde{k}_{IC} = k_0, \tilde{r}_{IC} = r_0) \rightarrow 1$. \square

A.3 Proofs of Theorems 2, 3 and Related Lemmas

Proof of Lemma 3: The proof of this lemma (which is based on fitting unrestricted VAR models, i.e. with cointegration rank $r = p$) follows by similar arguments as in the proof of Lemma 2 (which is based on fitting VAR models with cointegration rank $r = r_0$) and is therefore omitted. \square

Proof of Theorem 2: We want to prove that $(\hat{k}_{IC}, \hat{r}_{IC}(\hat{k}_{IC})) \xrightarrow{p} (k_0, r_0)$. In Theorem 1 in CDRT, we have that $\hat{r}_{IC}(k_0) \xrightarrow{p} r_0$, and, since \hat{r}_{IC} is a discrete valued random variable, $P(\hat{r}_{IC}(k_0) = r_0) \rightarrow 1$, as $T \rightarrow \infty$. From Lemma 3, we have that $\hat{k}_{IC}(p) \xrightarrow{p} k_0$, i.e., $P(\hat{k}_{IC} = k_0) \rightarrow 1$. Here we need to show that

$$P(\hat{k}_{IC} = k_0) \rightarrow 1 \quad \text{and} \quad P(\hat{r}_{IC}(k_0) = r_0) \rightarrow 1 \quad (\text{A.22})$$

imply $(\hat{k}_{IC}, \hat{r}_{IC}(\hat{k}_{IC})) \xrightarrow{p} (k_0, r_0)$. This result is established by proceeding along the same lines as in the proof of Theorem 1. \square

Proof of Theorem 3: We want to prove that $(\hat{k}_{IC}, \hat{r}_{PLR}^*(\hat{k}_{IC}))$ is consistent for (k_0, r_0) , in the sense that $\lim_{T \rightarrow \infty} P(\hat{k}_{IC} = k_0, \hat{r}_{PLR}^*(\hat{k}_{IC}) = r_0) = 1 - \alpha$ and $\lim_{T \rightarrow \infty} P(\hat{k}_{IC} = k_0, \hat{r}_{PLR}^*(\hat{k}_{IC}) < r_0) = 0$.

We have that, for a chosen significance level α ,

$$\begin{cases} P(\hat{r}_{PLR}^*(k_0) < r_0) \rightarrow 0 \\ P(\hat{r}_{PLR}^*(k_0) = r_0) \rightarrow 1 - \alpha \end{cases} \quad (\text{A.23})$$

under either i.i.d. innovations (CRT2012), conditional heteroskedasticity (CRT2010a), or unconditional heteroskedastic innovations (CRT2010b), which are combined in Assumption \mathcal{H} . Moreover, from Lemma 3, we have that $\hat{k}_{IC} := \hat{k}_{IC}(p) \xrightarrow{p} k_0$, i.e., $P(\hat{k}_{IC} = k_0) \rightarrow 1$ (see (A.22)). We therefore need to show that (A.22) and (A.23) imply the result that $\lim_{T \rightarrow \infty} P(\hat{k}_{IC} = k_0, \hat{r}_{PLR}^*(\hat{k}_{IC}) = r_0) = 1 - \alpha$.

To show this result we start from the decomposition

$$P(\hat{k}_{IC} = k_0, \hat{r}_{PLR}^*(\hat{k}_{IC}) = r_0) = P(\hat{k}_{IC} = k_0)P(\hat{r}_{PLR}^*(\hat{k}_{IC}) = r_0 \mid \hat{k}_{IC} = k_0),$$

where, from (A.22), $P(\hat{k}_{\text{IC}} = k_0) \rightarrow 1$. Hence, to complete the proof it suffices to show that $P(\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}) = r_0 \mid \hat{k}_{\text{IC}} = k_0) \rightarrow 1 - \alpha$. To establish this, we use the law of total probability as follows

$$P(\hat{r}_{\text{PLR}}^*(k_0) = r_0) = P(\hat{r}_{\text{PLR}}^*(k_0) = r_0 \mid \hat{k}_{\text{IC}} = k_0)P(\hat{k}_{\text{IC}} = k_0) + P(\hat{r}_{\text{PLR}}^*(k_0) = r_0 \mid \hat{k}_{\text{IC}} \neq k_0)P(\hat{k}_{\text{IC}} \neq k_0)$$

or, equivalently,

$$P(\hat{r}_{\text{PLR}}^*(k_0) = r_0 \mid \hat{k}_{\text{IC}} = k_0) = \frac{P(\hat{r}_{\text{PLR}}^*(k_0) = r_0)}{P(\hat{k}_{\text{IC}} = k_0)} - \frac{P(\hat{r}_{\text{PLR}}^*(k_0) = r_0 \mid \hat{k}_{\text{IC}} \neq k_0)P(\hat{k}_{\text{IC}} \neq k_0)}{P(\hat{k}_{\text{IC}} = k_0)} \quad (\text{A.24})$$

where, for T large enough, $P(\hat{k}_{\text{IC}} = k_0) > 0$. As $T \rightarrow \infty$, (A.22) and the second result in (A.23) imply that the first term on the right side of (A.24) tends to $1 - \alpha$, whereas the second term in (A.24) tends to 0, since $P(\hat{k}_{\text{IC}} \neq k_0) \rightarrow 0$. Therefore, $P(\hat{r}_{\text{PLR}}^*(k_0) = r_0 \mid \hat{k}_{\text{IC}} = k_0) \rightarrow 1 - \alpha$. Furthermore, since $P(\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}) = r_0 \mid \hat{k}_{\text{IC}} = k_0) = P(\hat{r}_{\text{PLR}}^*(k_0) = r_0 \mid \hat{k}_{\text{IC}} = k_0)$, we have that $P(\hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}) = r_0 \mid \hat{k}_{\text{IC}} = k_0) \rightarrow 1 - \alpha$. Similarly, we can prove that $P(\hat{k}_{\text{IC}} = k_0, \hat{r}_{\text{PLR}}^*(\hat{k}_{\text{IC}}) < r_0) \rightarrow 0$, using the first result in (A.23). \square