# Co-integration Rank and the Lag Order Determination in Heteroskedastic VAR Models

Giuseppe Cavaliere*
University of Bologna, Bologna, Italy - giuseppe.cavaliere@unibo.it

Luca De Angelis
University of Bologna, Bologna, Italy - l.deangelis@unibo.it

Anders Rahbek
University of Copenhagen, Copenhagen, Denmark - anders.rahbek@econ.ku.dk

A.M.Robert Taylor
University of Essex, Colchester, U.K. - rtaylor@essex.ac.uk

**Abstract**

In this paper we investigate the asymptotic and finite sample properties of a number of methods for estimating the co-integration rank and the lag order in integrated vector autoregressive systems 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 co-integration rank and the autoregressive lag order. We also extend the corpus of available large sample theory for the conventional sequential approach of Johansen 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. For the simulations DGPs considered, we find that the two best performing methods are a wild bootstrap implementation of the Johansen procedure implemented with BIC selection of the lag length and an approach which uses the BIC to jointly select the lag order and the co-integration rank.

**Keywords**: Co-integration rank; Information criteria; Wild bootstrap; Lag length.

## 1. Introduction

Determining the co-integration rank is a crucial aspect of modelling vector autoregressive [VAR] systems of variables integrated of order one, I(1). This is often done by using a test-based procedure such as the asymptotic (pseudo-) likelihood ratio [PLR] test of Johansen (1996). However, it is well-known that, even where the autoregressive lag order is assumed known, the sequential procedures based on PLR tests can be significantly over-sized in finite samples, particularly so in cases where the shocks contain strong serial correlation; see, in particular, Johansen, (2002) and the references therein. The PLR tests are also known to display significant upward size distortions in the presence of conditional heteroskedasticity (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. In the latter case these effects also persist asymptotically. As a result, and again assuming a known autoregressive lag length, Cavaliere, Rahbek and Taylor (2012, 2014) [CRT2012, CRT2014 hereafter] develop bootstrap implementations of the PLR tests. CRT2014 show that wild bootstrap PLR tests are asymptotically valid under heteroskedasticity in the shocks.

There is a close link between PLR tests and information criteria-based methods for determining the co-integration rank (and, indeed, the autoregressive lag length). Consequently, the efficacy of information criteria-based methods for determining the co-integrating rank has also been explored by a number of authors. Again for the case where the autoregressive lag length is assumed known, Cavaliere, De Angelis, Rahbek and Taylor (2014) [CDRT, hereafter] show that selecting the co-integrating rank using the Bayesian Information Criterion [BIC] provides a useful complement to the bootstrap sequential procedures. They also investigate the performance of the Akaike information criterion [AIC] and the Hannan-Quinn information criterion [HQC], and find that the BIC gives the best performance of the three overall. Under the assumption of a known lag length, CDRT extend the large sample results by Aznar and Salvador (2002) to the case of heteroskedastic shocks, showing that the BIC and HQC approaches remain consistent under the same conditions on the penalty function as are required in the homoskedastic case, whereas AIC-based approach is inconsistent and the resulting estimate of the co-integration rank displays a severe upward bias.

In practice, the assumption of a known autoregressive lag length in the co-integrated VAR [CVAR] model is challenged, as documented in numerous places. Specifically, simulation studies show that the PLR tests can suffer from serious size distortions, even asymptotically, if a lag order smaller than the true order is used; see, for example, Lütkepohl and Saikkonen (1999). Moreover, fitting a lag order in excess of the true order will necessarily compromise the finite sample power of the PLR tests. Consequently, the practitioner needs to estimate both the lag length and the co-integration rank. This is often done using a two-step procedure, whereby the autoregressive lag length is estimated in the first step, with this chosen lag length then employed in the co-integration rank determination second step. In practice, the first step will often involve using an information criterion. Under the assumption of homoskedastic shocks, Lütkepohl and Saikkonen (1999) show that PLR tests where the lag length is estimated by a consistent information criterion, such as BIC or HQC, have the same asymptotic distributions as PLR tests computed with knowledge of the true lag order.

Important gaps therefore exist in the literature on the theory of co-integration 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. Specifically, for shocks which can display stationary conditional heteroskedasticity and/or non-stationary unconditional heteroskedasticity, we provide sufficient conditions for information criteria based procedures to be consistent as the sample size diverges, when the autoregressive lag length is unknown. We discuss the cases where the lag length and co-integration rank are jointly determined and where a sequential procedure is employed. As part of establishing large sample results for the latter, we also demonstrate that the co-integration rank can be consistently estimated (using standard penalty functions) under heteroskedasticity regardless of the lag length used. 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 information-based method to first select the lag length prior to computing the (bootstrap) PLR tests leads to the same large sample properties as hold when the autoregressive lag length is known.

The second main contribution of this paper is to use Monte Carlo simulation methods to provide a comparison of the relative finite sample performances of the various methods for determining the co-integration rank discussed in the paper. Specifically, and by extending the Monte Carlo design employed in CDRT, our results suggest the following important conclusions: (i) an incorrect choice of the lag length significantly impacts on the efficacy of both PLR-based and information-based co-integration rank determination procedures in finite samples, highlighting the importance in finite samples of being able to accurately estimate the autoregressive lag order; (ii) a joint information-criterion based approach based on the BIC tends to outperform sequential information-criterion based methods; (iii) the joint BIC-based procedure provides a useful complement to the wild bootstrap sequential procedure of CRT2010a,b (implemented with BIC-based lag selection).

## 2. The Heteroskedastic Co-integrated VAR Model

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

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t, \quad t = 1, ..., T \tag{1}$$

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

The innovation process $\varepsilon_t := (\varepsilon_{1t}, ..., \varepsilon_{pt})'$ is assumed to satisfy the following general assumption:

**Assumption $\mathcal{H}$** The innovations $\{\varepsilon_t\}$ are defined as $\varepsilon_t := \sigma_t z_t$, where $\sigma_t$ is non-stochastic and satisfies $\sigma_t := \sigma(t/T)$ for all $t = 1, ..., T$, where $\sigma(\cdot) \in \mathcal{D}_{\mathbb{R}^{p\times p}}[0,1]$ and $\Sigma(u) := \sigma(u)\sigma(u)'$ 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 $T^{-1}\sum_{t=1}^{T} h_t \xrightarrow{p} E(z_t z_t') = I_p$.

Assumption $\mathcal{H}$ implies that $\varepsilon_t$ 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).

## 3. Methods for the Determining the Co-integration Rank when the Lag Length is Unknown

In this section we briefly discuss available methods for co-integration rank determination in the context of (1) where the true autoregressive lag, $k_0$, is unknown to the practitioner. Following Aznar and Salvador (2002), the autoregressive lag order and the co-integration rank can be jointly determined by (jointly) minimising an information criterion of the following generic form $\mathrm{IC}(k,r) := -2\ell_T(k,r) + p_T(k,r)$ over both all possible lag lengths, $k = 1, ..., K$, where $K$ denotes a given maximum lag order, and over all possible co-integration ranks, $r = 0, ...., p$. $\ell_T(k,r)$ denotes the maximised pseudo log-likelihood function associated with (1) under lag order $k$ and co-integration rank $r$, and $p_T(k,r) := c_T \pi(k,r)$ denotes the penalty function which depends on the number of parameters, $\pi(k,r)$. Different values of the coefficient $c_T$ yield different information criteria. The most widely used are AIC, BIC, and HQC where $c_T = 2$, $\log T$, and $2\log\log T$, respectively.

Up to a constant term which does not depend on $k$ or $r$, we therefore have from Johansen (1996) 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 (1) under rank $r$ for lag length $k$. As demonstrated in Johansen (1996), $|\hat{\Sigma}_{k,r}| := |S_{00}^{(k)}|\prod_{i=1}^{r}(1-\hat{\lambda}_i^{(k)})$, 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 $\hat{\lambda}_1^{(k)} > ... > \hat{\lambda}_p^{(k)}$ are the $p$ largest solutions to the eigenvalue problem, $|\lambda S_{11}^{(k)} - S_{10}^{(k)} S_{00}^{(k)-1} S_{01}^{(k)}| = 0$, where $S_{ij}^{(k)} := T^{-1}\sum_{t=1}^{T} R_{it}^{(k)} R_{jt}^{(k)\prime}$, $i,j = 0,1$, with $R_{0t}^{(k)}$ and $R_{1t}^{(k)}$ respectively denoting $\Delta X_t$ and $X_{t-1}'$, corrected (by OLS) for $\Delta X_{t-1}, ..., \Delta X_{t-k+1}$. We therefore have that $\mathrm{IC}(k,r) = T\log|S_{00}^{(k)}| + T\sum_{i=1}^{r}\log(1-\hat{\lambda}_i^{(k)}) + c_T\pi(k,r)$ where, in the case of no deterministic component, $\pi(k,r) = r(2p-r) + p(p+1) + p^2(k-1)$. Any information criterion as $\mathrm{IC}(k,r)$ can be used to jointly determine $k$ and $r$. In particular, the co-integration rank and lag order estimator for the joint procedure, denoted by $\tilde{k}_{\mathrm{IC}}$ and $\tilde{r}_{\mathrm{IC}}$, respectively, is then given, in generic form, by $(\tilde{k}_{\mathrm{IC}}, \tilde{r}_{\mathrm{IC}}) := \arg\min_{r=0,...,p;\ k=1,...,K} \mathrm{IC}(k,r)$; see Aznar and Salvador (2002). We show that provided the penalty term coefficient, $c_T$, is such that $c_T/T + 1/c_T \to 0$ as $T \to \infty$ then so $(\tilde{k}_{\mathrm{IC}}, \tilde{r}_{\mathrm{IC}})$ will provide consistent estimates for $k_0$ and $r_0$ under Assumption $\mathcal{H}$.

An alternative to the joint determination strategy is to use a two-step sequential procedure whereby one first estimates the lag length prior to the estimation the co-integration rank based on that estimated lag length in the second step. In relation to the first step, we show that the co-integration rank can be consistently estimated (using standard penalty functions where an information-based criterion is used) regardless of the lag length used. The generic information criterion is given by $\mathrm{IC}(k,p) = T\log|\hat{\Sigma}_{k,p}| + c_T\pi(k,p)$ 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 $\pi(k,p) := p(pk)$ when no deterministic component is involved. The resulting information-based lag length estimator is then given, in generic form, by $\hat{k}_{\mathrm{IC}}(p) := \arg\min_{k=1,...,K} \mathrm{IC}(k,p)$. Based on the lag order $\hat{k}_{\mathrm{IC}}(p)$ obtained in the first step, we then determine the co-integration 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 (1996) or the bootstrap equivalents of CRT2012 (i.i.d. re-sampling) and CRT2014 (wild bootstrap). A sequential information criteria-based approach then estimates the co-integration rank in the second step via $\hat{r}_{\mathrm{IC}}(\hat{k}_{\mathrm{IC}}(p)) := \arg\min_{r=0,...,p} \mathrm{IC}(\hat{k}_{\mathrm{IC}}(p), r)$ with $\mathrm{IC}(\hat{k}_{\mathrm{IC}}(p), r) = T\log|\hat{\Sigma}_{\hat{k}_{\mathrm{IC}}(p),r}| + c_T\pi(\hat{k}_{\mathrm{IC}}(p), r)$

where $\hat{\Sigma}_{\hat{k}_{\mathrm{IC}}(p),r}$ denotes the estimate of the residual covariance matrix obtained from the conventional reduced rank regression estimation of (1) under rank $r$ for lag length $\hat{k}_{\mathrm{IC}}(p)$, while the penalty component is $\pi(\hat{k}_{\mathrm{IC}}(p), r) = r(2p - r) + p(p+1) + p^2(\hat{k}_{\mathrm{IC}}(p) - 1)$, in the case of no deterministics. We show that if $\{X_t\}$ is generated as in (1) with the parameters satisfying the I(1, $r_0$) conditions and the penalty term coefficient, $c_T$, satisfies the usual conditions, i.e., $c_T/T + 1/c_T \to 0$ as $T \to \infty$, then $\hat{k}_{\mathrm{IC}}(p) \xrightarrow{p} k_0$ and $\hat{r}_{\mathrm{IC}}(\hat{k}_{\mathrm{IC}}(p)) \xrightarrow{p} r_0$, under Assumption $\mathcal{H}$.

The final two-step procedure we consider is the one that is probably the most widely used in empirical work. Here, once $\hat{k}_{\mathrm{IC}}(p)$ has been obtained in the first step, the second step then uses the well-known PLR test-based sequential procedure of Johansen (1996) to determine the co-integration rank. This procedure is based around the sequential application the PLR test of the null hypothesis that the co-integration rank is (less than or equal to) $r$. The PLR test rejects for large values of the trace statistic $Q_{r,\hat{k}_{\mathrm{IC}}(p),T} := -T\sum_{i=r+1}^{p}\log(1 - \hat{\lambda}_i^{(\hat{k}_{\mathrm{IC}}(p))})$, where $\hat{\lambda}_1^{(\hat{k}_{\mathrm{IC}}(p))} > \ldots > \hat{\lambda}_p^{(\hat{k}_{\mathrm{IC}}(p))}$ are the largest $p$ solutions to the eigenvalue problem reported above but with $k$ replaced by $\hat{k}_{\mathrm{IC}}(p)$. The analogous bootstrap procedure for determining the co-integrating rank is proposed in CRT2012 (i.i.d. re-sampling) and CRT2014 (wild bootstrap). Again based on the lag length $\hat{k}_{\mathrm{IC}}(p)$ estimated in the first step. For the case where the lag length is known, CRT2014 show that only the wild bootstrap PLR test is correctly sized asymptotically under both conditional and unconditional heteroskedasticity. We extend this result to the case of unknown lag order. Specifically, where the wild bootstrap version is used, the (wild) bootstrap PLR-based sequential estimator of the co-integration rank, $\hat{r}^*_{\mathrm{PLR}}(\hat{k}_{\mathrm{IC}}(p))$, is consistent under Assumption $\mathcal{H}$, provided $\hat{k}_{\mathrm{IC}}(p) \xrightarrow{p} k_0$, as $T \to \infty$.

## 4. Numerical Results

Using Monte Carlo simulation methods we now investigate the finite sample performances of the methods for estimating the co-integration rank outlined in section 3 under both homoskedastic models and models with conditional or unconditional heteroskedasticity as implied by Assumption $\mathcal{H}$.

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 $p = 4$: $\Delta X_t = \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \varepsilon_t$, where the long-run parameter vectors are set to $\beta' := [1\,0\,0\,0]$ and $\alpha' := [a\,0\,0\,0]$. We consider for the true co-integrating rank $r_0 = 0$, i.e., $a = 0$ and $r_0 = 1$, i.e., $a = -0.4$. Furthermore, we set $\Gamma_1 := \gamma I_4$ with $\gamma \in \{0.0, 0.1, 0.2, 0.3, 0.5, 0.8, 0.9\}$, so that the $\mathrm{I}(1, r)$ conditions are met, as $|\gamma| < 1$. For the individual components of $\varepsilon_t$ we consider the univariate innovation processes and parameter configurations as used in CDRT:

- **Case A**. $\varepsilon_{it}$, $i = 1, ..., p$, is an independent sequence of $N(0, 1)$ variates.
- **Case B**. $\varepsilon_{it}$ 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 $(\xi_{it}, v_{it})' \sim$ i.i.d. $N(0, \mathrm{diag}(\sigma_\xi^2, 1))$, independent across $i = 1, ..., p$. We consider $\lambda = 0.951$, $\sigma_\xi = 0.314$.
- **Case C**. $\varepsilon_{it}$ is a non-stationary, heteroskedastic independent sequence of $N(0, \sigma_{it}^2)$ variates, where $\sigma_{it}^2 = 1$ for $t \le \lfloor T\tau \rfloor$ and $\sigma_{it}^2 = \kappa$ for $t > \lfloor T\tau \rfloor$, all $i = 1, ..., p$. We consider $\tau = 2/3$ and $\kappa = 3$ (late positive variance shift).

Cases A-C are all nested within the general Assumption $\mathcal{H}$. In particular, Case A denotes the homoskedastic case (i.i.d. shocks), which is obtained from Assumption $\mathcal{H}$ by removing both conditional and unconditional heteroskedasticity. Case B considers a well-known conditionally heteroskedastic model for the innovations, thus ruling out unconditional heteroskedasticy. Finally, Case C involves a single, permanent shift in the innovation variance, thus leading to error sequences which are unconditionally heteroskedastic. In our simulation analysis we consider $K = 6$ as maximum lag length and the results are obtained using 10,000 Monte Carlo replications and 399 bootstrap replications.

We first investigate the impact that the value of the lag length used has on the ability of information-based and PLR based procedures to determine the co-integration rank. In particular, we evaluate the performance of the PLR- and BIC-based procedures for determining the co-integration rank, fixing the lag length used at each of $k = 1, ..., 6$. The results 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, the best results are obtained when the lag length used is equal to the true lag length, $k_0 = 2$. Interestingly, both under-specifying and over-specifying the lag length can significantly reduce the efficacy of both procedures. Specifically, the performance is worst when the lag length is either under-specified ($k = 1$) or heavily over-specified ($k = 5, 6$). We then briefly examine the behavior of information criteria $\mathrm{IC}(k, p)$ in determining the true lag length, $k_0 = 2$. Notice that this is used as the first step in the sequential procedures, that is the selected lag order, $\hat{k}_{\mathrm{IC}}(p)$, will then be adopted in the second step where the co-integration rank is determined. The results suggest that the presence of either conditional (Case 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. Moreover, and as might be expected, the value of the co-integration rank also does not appear to have a strong influence on the performance of these procedures. From the 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 $\gamma = 0.1$, 0.2, and 0.3, $\mathrm{BIC}(k, p)$ performs poorly and the AIC-based and HQC-based approaches appear preferable. The results also confirm some well-known properties of the information criteria considered in our analysis; most notably, the tendency of $\mathrm{AIC}(k, p)$ to over-estimate $k$, and the result that, in general, $\hat{k}_{\mathrm{BIC}} \leq \hat{k}_{\mathrm{HQC}} \leq \hat{k}_{\mathrm{AIC}}$, in line with the respective penalty functions used in these three criteria, noting that $c_T$ satisfies $\log T < 2 \log \log T < 2$ when $T > 15$.

Next we consider the results pertaining to co-integration rank determination. Overall, the results show that 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 first consider the results for the IC-based procedure. First note that the sequential determination of $k$ and $r$ based on the same information criterion in each step, e.g., $\mathrm{BIC}(\hat{k}_{\mathrm{BIC}}(p), r)$ behaves very similarly to the corresponding joint procedure, as might be expected. In line with previous studies (see, e.g., CDRT) the AIC-based procedures is shown to be completely unreliable as they tend to over-estimate the co-integration rank for all values of $\gamma$ and for all of Cases A-C. In contrast, the joint BIC-based procedure, $\mathrm{BIC}(k, r)$, generally performs well when the sample size is relatively large, i.e., $T \geq 200$, even under either conditional heteroskedasticity (Case B) or unconditional heteroskedasticity (Case C). However, the performance of $\mathrm{BIC}(k, r)$ is affected by the value of $\gamma$; specifically, performance is significantly better when $\gamma$ is large ($\gamma \geq 0.5$), especially when the sample size is small ($T < 200$). In general, when $\gamma \leq 0.3$ and the sample size is small, $T \leq 100$, the HQC-based procedures seem preferable to the BIC-based procedures, although this is largely an artefact of the greater tendency of the former to over-estimate both the lag order and the co-integration rank in such cases.

The results for the sequential procedure based on the wild bootstrap, $Q^*_{r,\hat{k}_{\mathrm{IC}}(p),T}$ show that, in line with the results in CRT2014 and CDRT, the wild bootstrap procedure delivers very decent finite sample performance throughout. Overall, the performance of $Q^*_{r,\hat{k}_{\mathrm{IC}}(p),T}$ is only marginally affected by either the presence of heteroskedasticity. In all cases when $T = 400$, the empirical frequency with which $Q^*_{r,\hat{k}_{\mathrm{IC}},T}$ selects the true co-integration rank generally exceeds 90%, even in the presence of either conditional heteroskedasticity (Case B) or unconditional heteroskedasticity (Case C). However, the performance of the sequential procedures based on wild bootstrap PLR tests deteriorates in the smaller sample sizes considered, especially when the parameter $\gamma$ is small. Although the results 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 co-integration rank, the $Q^*_{r,\hat{k}_{\mathrm{BIC}}(p),T}$ does marginally outperform the other procedures in most cases. However, we note that $Q^*_{r,\hat{k}_{\mathrm{IC}}(p),T}$ tends to lose power when $\hat{k}_{\mathrm{IC}}(p) < k_0$ in small samples. Finally, we also note that the joint BIC-based procedure and $Q^*_{r,\hat{k}_{\mathrm{BIC}}(p),T}$ tend to complement one another when the sample size is small.

## 5. Conclusions

In this paper we have analyzed the asymptotic and finite sample properties of various methods for determining the co-integration rank in the context of a heteroskedastic VAR model whose autoregressive lag length is unknown. The VAR model considered allows for both conditional and unconditional heteroskedasticity of a very general form. 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 (1996) based on

PLR or bootstrap PLR tests, but using a lag length determined by an information criterion. We established the conditions required on the penalty functions such that the joint and sequential information criterion-based methods deliver consistent estimates of both the lag length and co-integration rank. This condition was shown to coincide with that required in the homoskedastic case. Consequently, the BIC and HQC penalties satisfy this condition while the AIC penalty does not. The same requirement on the penalty function was 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 (1996) sequential procedure which is commonly used by practitioners, and shown that these methods hold under a very general class of heteroskedastic innovations.

We also investigated the relative finite sample properties of the various methods in a Monte Carlo simulation study using a VAR(2) DGP model driven by either homoskedastic or various forms of heteroskedastic shocks. Our results suggested that the best performing methods were the Johansen (1996)-type sequential procedure based on wild bootstrap PLR tests and BIC selection of the lag length, and the approach which uses the BIC to jointly select the lag order and the co-integration rank, and that these two methods provide a very useful complement to one another.

**References**


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.

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

Cavaliere, G., A. Rahbek and A.M.R. Taylor (2010*b*). 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, forthcoming in *Econometric Reviews*.

Cavaliere, G., L. De Angelis, A. Rahbek and A.M.R. Taylor (2014). A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic VAR models, forthcoming in *Oxford Bulletin of Economics and Statistics*.

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

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*, R.F. Engle and H. White (eds.), Oxford, England, Oxford University Press, 168–199.