

Adaptive information-based methods for determining the co-integration rank in heteroskedastic VAR models

H.Peter Boswijk^a, Giuseppe Cavaliere^b, Luca De Angelis^b, and A.M.Robert Taylor^c

^a Tinbergen Institute & Amsterdam School of Economics, University of Amsterdam

^b Department of Economics, University of Bologna

^c Essex Business School, University of Essex

January 11, 2019

Abstract

Standard methods, such as sequential procedures based on Johansen's (pseudo-)likelihood ratio (PLR) test, for determining the co-integration rank of a vector autoregressive (VAR) system of variables integrated of order one can be significantly affected, even asymptotically, by unconditional heteroskedasticity (non-stationary volatility) in the data. Known solutions to this problem include wild bootstrap implementations of the PLR test or the use of an information criterion, such as the BIC, to select the co-integration rank. Although asymptotically valid in the presence of heteroskedasticity, these methods can display very low finite sample power under some patterns of non-stationary volatility. In particular, they do not exploit potential efficiency gains that could be realised in the presence of non-stationary volatility by using adaptive inference methods. Under the assumption of a known autoregressive lag length, Boswijk and Zu (2016) develop adaptive PLR test based methods using a non-parameteric estimate of the covariance matrix process. It is well-known, however, that selecting an incorrect lag length can significantly impact on the efficacy of both information criteria and bootstrap PLR tests to determine co-integration rank in finite samples. We show that adaptive information criteria-based approaches can be used to estimate the autoregressive lag order to use in connection with bootstrap adaptive PLR tests, or to jointly determine the co-integration rank and the VAR lag length and that in both cases they are weakly consistent for these parameters in the presence of non-stationary volatility provided standard conditions hold on the penalty term. Monte Carlo simulations are used to demonstrate the potential gains from using adaptive methods and an empirical application to the U.S. term structure is provided.

Keywords: Co-integration rank; Adaptive estimation; Information criteria; Autoregressive lag length; Non-stationary volatility.

J.E.L. Classifications: C32, C14.

1 Introduction

It is well-known that standard methods for determining the co-integration rank of vector autoregressive (VAR) systems of variables integrated of order one are affected by the presence of heteroskedasticity. In particular, sequential procedures based on (pseudo-) likelihood ratio [PLR] test as developed by Johansen (1996) can be significantly over-sized, even in large samples, when the volatility process displays non-stationary variation (so called *non-stationary unconditional volatility*) and, moreover, the finite sample power of these tests can vary enormously depending on the pattern of heteroskedasticity present; see, in particular, Cavaliere, Rahbek and Taylor (2010). This is an important issue in practice because time-varying behaviour in unconditional volatility appears to be a common feature in many key macroeconomic and financial time series; see, among many others, McConnell and Perez Quiros (2000), Sensier and van Dijk (2004), and Cavaliere and Taylor (2008).

In a series of recent papers, Cavaliere, Rahbek and Taylor (2010, 2014) show that a solution to the size problems induced by non-stationary volatility is obtained by using wild bootstrap based implementations of the standard PLR tests. In particular, Cavaliere *et al.* (2010) show that the sequential procedure based on wild bootstrap PLR tests leads to consistent co-integration rank determination in the presence of non-stationary unconditional volatility. As alternative solution to the use of wild bootstrap PLR tests is considered by Cavaliere, De Angelis, Rahbek and Taylor (2015, 2018) who show that methods based on information criteria can also be used to consistently determine the co-integration rank in the presence of non-stationary volatility. In particular, they show that popular information criteria such as the Bayesian information criterion [BIC] (Schwarz, 1978) and the Hannan-Quinn information criterion [HQC] (Hannan and Quinn, 1979) provide a useful complement to the wild bootstrap sequential procedures.

The wild bootstrap PLR tests are correctly sized in the presence of non-stationary volatility and attain the same asymptotic local power functions as infeasible size-corrected versions of the standard PLR tests. As such they can therefore display very low power properties for some patterns of non-stationary volatility. Indeed, other things equal, their asymptotic local power functions are reduced, relative to the unconditionally homoskedastic case, under non-stationary volatility. Similarly, the ability of the standard information criteria-based methods discussed above to select the correct co-integration rank can also be greatly reduced under non-stationary volatility. In particular, none of these methods exploits the potential efficiency gains that could be provided by using inference methods which adapt to the volatility process. Adaptive methods, where the covariance matrix process is estimated non-parametrically, have the potential to be particularly useful in this context.

Under the assumption of a known autoregressive lag length, Boswijk and Zu (2016) develop an procedure based on adaptive PLR tests for determining the co-integration rank in possibly heteroskedastic VAR models. Specifically, they propose a procedure where the volatility process is estimated using a non-parametric kernel estimator, with this estimate then used in the adaptive PLR test procedure. Under suitable conditions, they establish that the non-parametric volatility estimator is consistent and that the resulting adaptive PLR co-integration rank tests have the same asymptotic local power functions as for infeasible tests based on the assumption that the volatility process is known. The asymptotic null distribution of their proposed statistics are, however, non-standard and depend on the realisation of the volatility process. As such, asymptotic p -values for the adaptive PLR tests need to be obtained using bootstrap methods.

The assumption of a known of autoregressive lag order is problematic in practice. It is well-known that an incorrect lag length choice can significantly impact on the efficacy of both information criteria and PLR tests, in particular where a lag order smaller than the true order is used; see, among others, Boswijk and Franses (1992), Cheung and Lai (1993), Haug (1996), Lütkepohl and Saikkonen (1999), and Cavaliere *et al.* (2018). In practice the autoregressive lag length will need to be estimated along with the co-integration rank. To that end, the practitioner can use either a sequential procedure, where the lag length is consistently estimated in a first step and then subsequently employed in the second step in a procedure such as either the adaptive PLR test approach of Boswijk and Zu (2016) or an information criterion for determining the co-integration rank, or a joint information criteria-based approach can be used whereby the lag length and co-integration rank are determined simultaneously. Cavaliere *et al.* (2018) show that both joint and sequential procedures based on standard information criteria consistently determine both the lag length and the co-integration rank in the presence of non-stationary unconditional volatility, provided standard conditions hold on the penalty term. They also show the asymptotic validity of a sequential procedure based on wild bootstrap PLR tests with the autoregressive lag length chosen by an information criterion.

The contribution of this paper is to develop adaptive information criteria methods, based around a (non-parametric) estimation of the volatility process, for jointly selecting the co-integration rank and autoregressive lag order. We show that these adaptive information criteria-based methods are weakly consistent for the co-integration rank and autoregressive lag order under the precisely the same conditions on the penalty function are as required for the consistency of standard (non-adaptive) information criteria under non-stationary volatility of the form considered in this paper. We also establish the asymptotic validity of a sequential procedure selecting the autoregressive lag length by an adaptive information criterion [ALS-IC] in the first step and then determining the co-integration rank using again an ALS-IC in the second step based on the first step estimate of the lag length. Because the co-integration rank

is determined by minimising an adaptive information criterion over all possible values of the co-integration rank from zero up to the dimension of the system, the practitioner does not therefore need to obtain p -values by bootstrap methods, making the procedure considerably less time consuming than the Boswijk and Zu (2016) procedure based on adaptive PLR tests. We also establish the asymptotic validity of a sequential procedure selecting the autoregressive lag length by an ALS-IC in the first step and then using the adaptive PLR test-based approach of Boswijk and Zu (2018) in the second step based on the first step estimate of the lag length.

The remainder of the paper is organized as follows. Section 2 details our reference heteroskedastic co-integrated VAR model. Section 3 outlines adaptive information criteria-based methods for determining the co-integration rank and the autoregressive lag length. The large sample properties of these procedures are detailed in Section 4. Monte Carlo simulation experiments reported in Section 5 are used to explore the finite sample performance of the ALS-IC methods relative to standard methods such as those based on standard information criteria-based procedures. These results highlight the potential gains that can be achieved by using adaptive methods. Section 6 provides an empirical application of the methods discussed in this paper to the term structure of interest rates in the US. Section 7 concludes. Proofs of our main results along with the tabulation of the results from our Monte Carlo simulation experiments are contained in the accompanying online Supplementary Appendix.

2 The Heteroskedastic Co-integrated VAR Model

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

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

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 true co-integration rank of the system (see also Cavaliere, Rahbek and Taylor, 2012); 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 deterministic variables in (2.1) are taken to satisfy one of the following cases (see, e.g., Johansen, 1996): (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).

The innovation process $\varepsilon_t := (\varepsilon_{1t}, \dots, \varepsilon_{pt})'$ in (2.1) is taken to satisfy the following set of conditions collectively labelled Assumption 1.

Assumption 1 *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 $\sigma(u)$ is non-singular for all $u \in [0, 1]$ and continuous in $u \in [0, 1]$; z_t is an i.i.d. sequence with $E(z_t) = 0$ and $E(z_t z_t') = I_p$.*

Remark 1. Assumption 1 implies that $E(\varepsilon_t) = 0$ and that ε_t has the time-varying unconditional variance matrix $\Sigma_t := E(\varepsilon_t \varepsilon_t') = \sigma_t \sigma_t' > 0$. In what follows, σ_t will be referred to as the *volatility matrix* of ε_t . Elements of Assumption 1 have previously been employed by, *inter alia*, Cavaliere *et al.* (2010), Boswijk, Cavaliere, Rahbek and Taylor (2016), Cavaliere *et al.* (2018) and Boswijk and Zu (2016). In particular, Assumption 1 allows for a countable number of discontinuities in $\sigma(\cdot)$ therefore allowing for a wide class of potential models for the time-varying behaviour of the unconditional variance matrix of ε_t . As discussed in Boswijk and Zu (2016), the continuity assumption on $\sigma(\cdot)$ is made so that $\sigma(\cdot)$ can be consistently estimated. This assumption is not restrictive in practice however because one can always approximate discontinuities in $\sigma(\cdot)$ arbitrarily well using smooth transition functions. Moreover, one could relax this assumption by assuming that $\sigma(\cdot)$ is a piecewise Lipschitz-continuous function; see Xu and Phillips (2008). \diamond

Remark 2. In order to simplify our presentation, Assumption 1 rules out the possibility of conditional heteroskedasticity in z_t . We do so because adaptive estimation can only lead to efficiency gains over standard estimation in cases where $\sigma(u)$ varies across u which can only happen where non-stationary volatility is present. Conditional heteroskedasticity of the form considered in Assumption 2(b) of Boswijk, Cavaliere, Rahbek and Taylor (2016), cannot induce time-variation in $\sigma(u)$ and so it is irrelevant so far as adaptive estimation is concerned. It is straightforward, however, to show that the large sample results given in this paper remain valid if we allow for conditional heteroskedasticity in z_t of the form considered in Assumption 2(b) of Boswijk, Cavaliere, Rahbek and Taylor (2016). \diamond

3 Adaptive Information Criteria for Determining the Co-integration Rank and the Autoregressive Lag Length

In this section we discuss adaptive information-based methods for determining the co-integration rank and the autoregressive lag length in the context of (2.1). In particular, we first derive the log-likelihood function in Section 3.1 and the nonparametric estimator of the volatility matrix in Section 3.2. We then outline the adaptive information criterion for the joint determination

of the co-integration rank and the lag length in Section 3.3 and we discuss how to sequentially estimate the lag length and the co-integration rank using adaptive methods in Section 3.4.

3.1 The Likelihood Function

Define $\Psi := [\Gamma_1 : \dots : \Gamma_{k-1}]$ and $Z_t^{(k)} := (\Delta X'_{t-1}, \dots, \Delta X'_{t-k+1})'$, such that the model in (2.1) with no deterministic components (case (i)) can be rewritten more compactly as

$$\Delta X_t = \alpha\beta X_{t-1} + \Psi Z_t^{(k)} + \varepsilon_t. \quad (3.1)$$

Suppose for the present that $\{\sigma_t\}$ is known, and that z_t is Gaussian; i.e., $z_t \sim i.i.d. N(0, I_p)$. Then under Assumption 1 we have that $\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \Sigma_t)$, where $\mathcal{F}_{t-1} := \{X_{t-1}, \dots, X_1, X_0, \dots, X_{1-k}\}$, and the log-likelihood function is given by (see Boswijk and Zu, 2016):

$$\begin{aligned} \ell_T(\alpha, \beta, \Psi) &= -\frac{Tp}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log |\Sigma_t| \\ &\quad - \frac{1}{2} \sum_{t=1}^T (\Delta X_t - \alpha\beta' X_{t-1} - \Psi Z_t^{(k)})' \Sigma_t^{-1} (\Delta X_t - \alpha\beta' X_{t-1} - \Psi Z_t^{(k)}). \end{aligned} \quad (3.2)$$

Maximum likelihood estimation of the parameters (α, β, Ψ) can be achieved by using the so-called *generalised reduced rank regression* procedure (Boswijk, 1995; Hansen, 2002, 2003), which uses a switching algorithm in order to circumvent the issue of the lack of a closed-form expression for the maximum likelihood estimator (MLE). In particular, because the MLE of (α, Ψ) for fixed β and the MLE of β for fixed (α, Ψ) have closed-form expressions, the maximisation of (3.2) can be achieved, starting from an initial guess, by switching between maximisation over (α, Ψ) and β ; see Boswijk and Zu (2016) for further details.

3.2 Volatility Estimation

In this paper we focus on the two-sided smoothing nonparametric estimator of the volatility matrix adopted by Boswijk and Zu (2018). This estimator is a multivariate extension of Hansen (1995)'s nonparametric volatility filter based on leads and lags of the outer product of the residual vector. A similar approach to adaptive estimation has also been considered by Xu and Phillips (2008) and Patilea and Raïssi (2012), among others.

Let $K(\cdot)$ denote some kernel function and define $K_h(x) := K(x/h)/h$ with $h > 0$ a *window width*. The kernel estimator for Σ_t that we will consider is then defined as,

$$\hat{\Sigma}_t := \frac{\sum_{s=1}^T K_h\left(\frac{t-s}{T}\right) \hat{e}_s \hat{e}_s'}{\sum_{s=1}^T K_h\left(\frac{t-s}{T}\right)} \quad (3.3)$$

where \hat{e}_t is the residual vector obtained by estimating an unrestricted VAR model of order K in the levels of X_t , i.e. $\hat{e}_t = X_t - \sum_{i=1}^K \hat{A}_i X_{t-i}$, where A_i , $i = 1, \dots, K$, are $p \times p$ coefficient matrices. The value K denotes the maximum autoregressive lag order we will allow for which, unless otherwise stated, is assumed in the following to be at least as large as the true lag order, k_0 in (2.1).

The kernel function in (3.3) is implemented with two-sided smoothing, so that $\hat{\Sigma}_t$ is based on leads and lags of $\hat{e}_t \hat{e}'_t$, as outlined in Assumption 3 in Boswijk and Zu (2016). In their Lemma 2, Boswijk and Zu (2016) show that the volatility matrix process implied by the T nonparametrically estimated covariance matrices is uniformly consistent over the compact interval $[0, 1]$, which, in turn, implies uniform consistency of the nonparametric estimator $\hat{\Sigma}_t$ in (3.3) over $t = 1, \dots, T$. Therefore, these consistent estimators can be used to replace Σ_t in the log-likelihood function in (3.2), thereby allowing for a feasible version of the generalised reduced rank regression procedure and the computation of the adaptive information criteria and the adaptive bootstrap PLR tests.

In implementing the nonparametric estimator of Σ_t in (3.3), we will select the window width h by minimising the quantity

$$\tilde{C}_T(h) := \sum_{t=1}^T \|\hat{\Sigma}_t^{-t}(h) - \hat{e}_t \hat{e}'_t\|^2,$$

where $\|\cdot\|$ denotes the Euclidean matrix norm, and where $\hat{\Sigma}_t^{-t}(h)$ is given by (3.3), but with $K(0)$ replaced by 0, so that $\hat{e}_t \hat{e}'_t$ does not enter the expression for $\hat{\Sigma}_t^{-t}(h)$. This leave-one-out cross-validation technique is implemented in Boswijk and Zu (2016) and Patilea and Raïssi (2012), and satisfies the requirement that h decreases with the sample size at a certain rate; see Lemma 2 of Boswijk and Zu (2016) and Section 4 below.

3.3 Joint Determination of the Lag Length and Co-integration Rank

The maximised pseudo log-likelihood function (3.2) associated with (3.1) under lag order k and co-integration rank r , say $\hat{\ell}_T^{(k,r)}(\alpha, \beta, \Psi)$, in conjunction with the volatility estimator in (3.3) substituted for Σ_t in (3.2) can then be used to construct a feasible *adaptive* information criterion of the following generic form

$$\text{ALS-IC}(k, r) := -2\hat{\ell}_T^{(k,r)}(\alpha, \beta, \Psi) + c_T \pi(k, r) \quad (3.4)$$

where the term c_T may depend on the sample size T (see below) and where $\pi(k, r)$ denotes the number of parameters in the estimated model.¹ The autoregressive lag order and the co-integration rank can then be jointly estimated by minimising the information criterion in (3.4)

¹The number of parameters which defines the penalty term in (3.4) depends on the deterministic components included in the model (2.1) as follows: (i) in the case of no deterministic component ($D_t = 0$, $d_t = 0$ in 2.1),

jointly over both all possible lag lengths, $k = 1, \dots, K$, and over all possible co-integration ranks, $r = 0, \dots, p$; that is,

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

Different values of the coefficient c_T yield different adaptive information criteria. In the standard (non-adaptive) case, which can be obtained as a special case of the adaptive information criterion in (3.4) by restricting $\Sigma_t = I_p$ in the likelihood function (3.2), the most widely used information criteria are the Akaike information criterion [AIC] (Akaike, 1974), the Bayes information criterion [BIC] (Schwarz, 1978), and the Hannan-Quinn information criterion [HQC] (Hannan and Quinn, 1979), which obtain setting $c_T = 2$, $\log T$, and $2 \log \log T$, respectively. We will denote the generic standard information criterion in this case as $\text{IC}(k, r)$ and the resulting estimate in (3.3) as $(\tilde{k}_{\text{IC}}, \tilde{r}_{\text{IC}})$. In the context of (3.4), we will refer to the adaptive information criteria based on the AIC, BIC and HQ choices of c_T as ALS-AIC, ALS-BIC, and ALS-HQC, respectively.

3.4 Sequential Determination of the Lag Length and Co-integration Rank

Because the lag length k in (2.1) is in general unknown and needs to be estimated prior to estimating the co-integration rank, practitioners often use a two-step procedure, whereby the autoregressive lag length is estimated in the first step and then subsequently employed as if it were the known lag length in a second step for determining the co-integration rank, such as a sequential procedure based on PLR tests or an information criterion. In particular, Lütkepohl and Saikkonen (1999) and Nielsen (2006), *inter alia*, show that the lag length in nonstationary VAR models can be consistently estimated from the levels of the data using an information criterion. Therefore, the lag length could be selected in the first step of the sequential procedure according to a (standard) information criterion where we do not impose a reduced rank structure on $\Pi := \alpha\beta'$ in (2.1), that is by imposing $r = p$; see, among others, Cavaliere *et al.* (2018).

As for the joint determination of the lag length and the co-integration rank considered in section 3.3, an adaptive version of the information criterion for determining the lag length can also be considered. In particular, the lag length may be selected using an adaptive information criterion of the generic form

$$\text{ALS-IC}(k, p) := -2\hat{\ell}_T^{(k,p)}(\Pi, I_p, \Psi) + c_T\pi(k, p) \quad (3.5)$$

$\pi(k, r) = r(2p-r) + p^2(k-1)$; (ii) for the restricted constant case ($D_t = 1$, $d_t = 0$ in 2.1), $\pi(k, r) = r(2p-r+1) + p^2(k-1)$, and (iii) for the case of a restricted trend ($D_t = 1$, $d_t = 1$ in 2.1), $\pi(k, r) = r(2p-r+1) + p + p^2(k-1)$.

where $\hat{\ell}_T^{(k,p)}(\Pi, I_p, \Psi)$ is the maximised pseudo likelihood (3.2) associated with (3.1) where we do not impose a reduced rank structure on $\Pi = \alpha\beta'$ under lag length k and Σ_t in (3.2) is substituted with the volatility estimator in (3.3). Again, the choice of the c_T term identifies different information criteria as outlined above and, in this case, $\pi(k, p) = p(pk + i)$ with $i = 0$ when no deterministic component is involved, $i = 1$ in the case of restricted constant, and $i = 2$ for the restricted trend. The resulting adaptive information criterion-based lag length estimator is then given by

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

We note again that the generic standard information criterion, which we will denote by $\text{IC}(k, p)$, can be obtained as a special case of the adaptive information criterion in (3.5), by restricting $\Sigma_t = I_p$ in the likelihood function (3.2), with the resulting lag length estimator denoted by \hat{k}_{IC} . In the simulation experiments discussed in Section 5, we will consider both the standard and the adaptive versions of the information criterion for determining the lag length in the first step of the two step sequential procedure. The selected lag length, either \hat{k}_{IC} or $\hat{k}_{\text{ALS-IC}}$ generically denoted by \hat{k} for the remainder of this section, is then used as if it were the true lag length in the second step for determining the co-integration rank. The second step could be based on either the sequential procedure of Boswijk and Zu (2016) based on adaptive bootstrap PLR tests or an adaptive information criterion for selecting the co-integration rank. We now outline these two possibilities.

The adaptive PLR test-based procedure of Boswijk and Zu (2016). Boswijk and Zu (2016) introduce the adaptive PLR statistic for testing the null hypothesis that the true co-integration rank is (no more than) r , $0 \leq r \leq p - 1$,

$$Q_{r,k,T} := -2 \left[\hat{\ell}_T^{(k,r)}(\alpha, \beta, \Psi) - \hat{\ell}_T^{(k,p)}(\Pi, I_p, \Psi) \right] = \sum_{t=1}^T \left(\hat{\varepsilon}'_{r,t} \hat{\Sigma}_t^{-1} \hat{\varepsilon}_{r,t} - \hat{\varepsilon}'_{p,t} \hat{\Sigma}_t^{-1} \hat{\varepsilon}_{p,t} \right) \quad (3.6)$$

where $\hat{\varepsilon}_{r,t}$ and $\hat{\varepsilon}_{p,t}$ denote the residuals from the restricted and unrestricted VAR model in (2.1), respectively. For the case where the autoregressive lag length is known ($k = k_0$), they demonstrate that the limiting distribution of (3.6) depends on the unknown volatility process. Consequently, bootstrap methods are required to approximate the critical values from this distribution. In order to do so, a bootstrap sample $\{X_{r,t}^*\}_{t=1}^T$ is generated recursively from

$$\Delta X_{r,t}^* = \hat{\alpha}^{(r)} \hat{\beta}^{(r)'} X_{r,t-1}^* + \sum_{i=1}^{k-1} \hat{\Gamma}_i^{(r)} \Delta X_{r,t-i}^* + \varepsilon_{r,t}^*, \quad t = 1, \dots, T \quad (3.7)$$

initialised at $X_{r,j}^* = X_j$, for $j = 1-k, \dots, 0$, where $\hat{\alpha}^{(r)}$, $\hat{\beta}^{(r)}$, and $\hat{\Gamma}_i^{(r)}$ are the estimated parameter matrices from the model (2.1) obtained using conventional reduced rank regression under the

rank r imposed by the null hypothesis. The adaptive PLR test statistic based on the bootstrap sample is then computed as

$$Q_{r,k,T}^* := \sum_{t=1}^T \left(\hat{\varepsilon}_{r,t}^{*/\prime} \hat{\Sigma}_t^{-1} \hat{\varepsilon}_{r,t}^* - \hat{\varepsilon}_{p,t}^{*/\prime} \hat{\Sigma}_t^{-1} \hat{\varepsilon}_{p,t}^* \right) \quad (3.8)$$

where $\hat{\varepsilon}_{r,t}^*$ and $\hat{\varepsilon}_{p,t}^*$ denote the (bootstrap) residuals from the restricted and unrestricted models, respectively. Following Boswijk and Zu (2016), we consider the following two bootstrap implementations: (i) the variance bootstrap, $\varepsilon_{r,t}^* := \hat{\Sigma}_t^{1/2} z_t^*$, where $\hat{\Sigma}_t^{1/2}$ is any square root of $\hat{\Sigma}_t$ and $z_t^* \sim i.i.d.N(0, I_p)$; (ii) the wild bootstrap, $\varepsilon_{r,t}^* := \hat{\varepsilon}_{r,t} w_t^*$, where w_t^* is a scalar i.i.d. $N(0,1)$ sequence; see Section 4.2 of Boswijk and Zu (2016) for more details. As is typically done in practice, the unknown lag length k in (3.7) and (3.8) is replaced by the lag length estimated in the first step of the sequential procedure, say \hat{k} , in order to compute the bootstrap statistic $Q_{r,\hat{k},T}^*$ using the bootstrap sample in (3.7) based on \hat{k} . The corresponding p -value is then computed as $p_{r,\hat{k},T}^* := 1 - G_{r,\hat{k},T}^*(Q_{r,\hat{k},T}^*)$, where $G_{r,\hat{k},T}^*(\cdot)$ denotes the conditional (on the original data) cdf of $Q_{r,\hat{k},T}^*$. Starting from $r = 0$, the bootstrap algorithm is repeated as long as $p_{r,\hat{k},T}^*$ exceeds the significance level η , thus yielding $\hat{r}^*(\hat{k}) = r$. If the null is not rejected for $r = p - 1$, then $\hat{r}^*(\hat{k}) = p$.

The asymptotic validity of the two bootstrap procedures outlined above is established in Theorem 3 of Boswijk and Zu (2016) with the implication that, for the case where the autoregressive lag length is known ($k = k_0$), the variance and wild bootstrap adaptive PLR test-based procedures, $\hat{r}^*(k_0)$, are asymptotically accurately capped estimator of the co-integration rank r_0 .² In Section 4 we will generalise these results to the case where the lag length is unknown and estimated in the first step of the sequential procedure.

The adaptive IC-based procedure. Alternatively, to determine the co-integration rank in the second step of a sequential procedure based on ALS-IC, k in the generic form (3.4) can be replaced by the lag length estimated in the first step, thus yielding

$$\text{ALS-IC}(\hat{k}, r) := -2\hat{\ell}_T^{(\hat{k}, r)}(\alpha, \beta, \Psi) + c_T \pi(\hat{k}, p). \quad (3.9)$$

The resulting adaptive information criterion-based co-integration rank estimator is then given by

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

²The sequential rank determination procedure of Johansen (1996) 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 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 $r_0 = p$.

4 Asymptotic Analysis

In this section we establish the large sample properties of the adaptive methods for determining the co-integration rank and autoregressive lag length outlined in sections 3.3 and 3.4.

Lemma 2 of Boswijk and Zu (2016) establishes that the nonparameteric estimate of the volatility matrix process defined as $\hat{\Sigma}_T(u) := \sum_{t=1}^T \hat{\Sigma}_t 1_{[(t-1)/T, t/T]}(u)$ is uniformly consistent over the compact interval $[0, 1]$. This result is a basic building block needed to demonstrate weak consistency³ for the adaptive information criteria in (3.4) and (3.5) and so for completeness we first reproduce that result below as Result 1.

Result 1 *Let $\{X_t\}$ be generated as in (2.1) with the parameters satisfying the $I(1, r_0)$ conditions and let Assumption 1 hold, and let K be a bounded non-negative function defined on \mathcal{R} which satisfies $\int_{-\infty}^{\infty} K(x)dx = 1$, $0 < \int_{-\infty}^0 K(x)dx < 1$ and $0 < \int_0^{\infty} K(x)dx < 1$. Then, if $T \rightarrow \infty$, $h \rightarrow 0$ and $Th^2 \rightarrow \infty$, it holds that*

$$\sup_{u \in [0, 1]} \|\hat{\Sigma}_T(u) - \Sigma(u)\| \xrightarrow{P} 0$$

where $\Sigma(u) := \sigma(u)\sigma(u)'$ is the true variance matrix process.

Using Result 1, we first show in Lemma 1 that the adaptive information criterion in (3.4) is weakly consistent for the co-integration rank, regardless of the autoregressive lag length used, provided standard conditions hold on the penalty term, c_T . Then second in Lemma 2 we show that for the true co-integration rank, r_0 , the adaptive information criterion in (3.4) is weakly consistent for the autoregressive lag length.

Lemma 1 *Let the conditions of Result 1 hold. Then, for any $0 < k \leq K$, it holds that, as $T \rightarrow \infty$:*

(i) *for $r > r_0$, $\text{Prob}(\text{ALS-IC}(k, r) > \text{ALS-IC}(k, r_0)) \rightarrow 1$, provided $c_T \rightarrow \infty$;*

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

Lemma 2 *Let the conditions of Result 1 hold. Then it holds that, as $T \rightarrow \infty$:*

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

³An estimator T_n is defined to be *weakly* consistent if it converges in *probability* to the true value of the unknown parameter θ ; that is, $T_n \xrightarrow{P} \theta$.

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

Remark 3. The results in Lemma 1 imply that, provided the standard condition that $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$, as $T \rightarrow \infty$, holds on the penalty term, c_T , then for any lag length $k = 1, \dots, K$, the adaptive information criterion-based estimator of the co-integration rank is weakly consistent for the true co-integration rank, r_0 . The results in Lemma 2 imply that, under the same conditions on c_T , the adaptive information criterion-based estimator of the lag length, computed by imposing the true co-integration rank, i.e. $r = r_0$ in (2.1), is a weakly consistent estimator for the true lag order k_0 . Consequently, in each case, the use of either the ALS-BIC or ALS-HQC, but not the ALS-AIC penalty, will yield weakly consistent estimates. Cavaliere *et al.* (2018) demonstrate that analogous results hold, with the same condition on c_T , for the corresponding non-adaptive information criterion-based estimators. \diamond

Using the results in Lemmas 1 and 2, we are now in a position to establish the weak consistency of the joint procedure. This is now given in Theorem 1.

Theorem 1 *Let the conditions of Result 1 hold. Then it holds that $(\tilde{k}_{\text{ALS-IC}}, \tilde{r}_{\text{ALS-IC}}) \xrightarrow{p} (k_0, r_0)$, provided c_T in (3.4) satisfies the condition that $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$.*

Remark 4. An immediate consequence of the result in Theorem 1 is that the resulting ALS-BIC-based and ALS-HQC-based estimators are weakly consistent for both the co-integration rank and autoregressive lag length, but that the corresponding ALS-AIC-based estimator is not. \diamond

To conclude this section we now detail the large sample behaviour of the two-step sequential procedures outlined in section 3.4 where in the first step we select the autoregressive lag and then in the second step an adaptive procedure based on this estimated lag length is used to determine the co-integration rank.

First, in Lemma 3, we generalise the results in Lemma 3 of Cavaliere *et al.* (2018), which show the sufficient conditions on the term c_T that ensure weak consistency for an information criterion of the form given in (3.5), to the case of its adaptive analogue, $\text{ALS-IC}(k, p)$. In particular, we derive the conditions under which minimising an adaptive information criterion consistently selects the true lag order, k_0 , in the first step when we do not impose a reduced rank structure, so that we set $r = p$.

Lemma 3 *Let the conditions of Result 1 hold. Then, for any $0 < k \leq K$, it holds that, as $T \rightarrow \infty$:*

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

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

The results in Lemma 3 imply that $\hat{k}_{\text{ALS-IC}} \xrightarrow{p} k_0$, again provided $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$, as $T \rightarrow \infty$. Using the results in Lemmas 1 and 3, we are now in a position in Theorem 2 to establish the large sample properties of the bootstrap adaptive PLR test-based estimator of the co-integration rank using the lag length estimated by an information criterion as in (3.5) at the first step, $\hat{r}^*(\hat{k}_{\text{ALS-IC}})$.

Theorem 2 *Let the conditions in Result 1 hold. Then, provided c_T in (3.5) is such that $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$, the variance and the wild bootstrap PLR-tests satisfy:*

(i) $\lim_{T \rightarrow \infty} P(\hat{r}^*(\hat{k}_{\text{ALS-IC}}) = r) = 0$ for all $r = 0, 1, \dots, r_0 - 1$;

(ii) $\lim_{T \rightarrow \infty} P(\hat{r}^*(\hat{k}_{\text{ALS-IC}}) = r_0) = 1 - \eta \cdot \mathbb{I}(r_0 < p)$, and $\lim_{T \rightarrow \infty} \sup_{r \in \{r_0+1, \dots, p\}} P(\hat{r}^*(\hat{k}_{\text{ALS-IC}}) = r) \leq \eta$.

Remark 5. The results in Theorem 2 show that, provided the information criterion used in the first step of the sequential procedure is a consistent lag length estimator, that is $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$, as $T \rightarrow \infty$, the bootstrap adaptive PLR test-based procedure is an asymptotically accurately capped estimator of the true co-integration rank, r_0 . \diamond

Remark 6. The results in Theorem 2 can also be shown to hold (under the same conditions) for any consistent lag length estimator obtained in the first step. Therefore, the consistency result in Theorem 2 will also hold for variance and wild bootstrap adaptive PLR tests when a standard information criterion, such either $\text{BIC}(k, p)$ or $\text{HQC}(k, p)$, is used to select the lag length at the first step. \diamond

Finally, in Theorem 3 we generalise the results in Theorem 2 of Cavaliere *et al.* (2018) by establishing the large sample properties of the adaptive IC-based estimator of the co-integration rank as in (3.9) using the lag length estimated by an information criterion as in (3.5) at the first step, $\hat{r}_{\text{ALS-IC}}(\hat{k}_{\text{ALS-IC}})$.

Theorem 3 *Let the conditions in Result 1 hold. Then it holds that $\hat{r}_{\text{ALS-IC}}(\hat{k}_{\text{ALS-IC}}) \xrightarrow{p} r_0$, provided c_T in (3.5) and (3.9) satisfies the condition that $\frac{c_T}{T} + \frac{1}{c_T} \rightarrow 0$ as $T \rightarrow \infty$.*

Remark 7. It is easy to show that the condition on c_T in Theorem 3 is not required if our purpose is to consistently estimate the co-integration rank. Indeed, as was shown in Lemma 1,

any fixed lag length k will also suffice in that case. However, as shown *inter alia* in Cavaliere *et al.* (2018), the finite sample performance of the information criteria for determining the co-integration rank can deteriorate badly if a fixed lag length is used which is not equal to the true lag length, k_0 , and particularly so where it is smaller than k_0 . \diamond

5 Numerical results

In this section we use Monte Carlo simulation methods to investigate the finite sample performance of the joint and sequential adaptive methods for determining the co-integration rank and autoregressive lag length outlined in sections 3.3 and 3.4 and compare these with their standard (non-adaptive) counterparts. The results from these Monte Carlo experiments are reported in Tables A-F which can be found in the accompanying on-line supplementary appendix.

We will consider the following second-order VAR model of dimension $p = 2$ as our simulation DGP:

$$\Delta X_t = \alpha \beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \varepsilon_t, \quad \alpha := \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \quad \beta := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (5.1)$$

with $t = 1 - K, \dots, T$, $X_{-K} = \Delta X_{-K} = 0$, where K denotes the maximum lag order. In order to allow for true co-integration ranks, r_0 , of 0, 1 or 2, we set the parameters a and b in the long-run parameter vector α in (5.1) as follows: $a = b = 0$ for $r_0 = 0$, $a = -0.4$ and $b = 0$ for $r_0 = 1$, and $a = b = -0.4$ for $r_0 = 2$ (full rank). Moreover, we set $\Gamma_1 := \gamma I_2$ with $\gamma \in \{0, 0.1, 0.5, 0.9\}$.⁴

We will consider three cases for the the innovation vector, ε_t in (5.1). The first case is that $\varepsilon_t \sim i.i.d. N(0, I_2)$ so that ε_t is homoskedastic. This case will provide a useful benchmark to investigate the effects of using adaptive methods when they are not needed. The second case considers conditionally heteroskedastic innovation processes, where the individual components of ε_t follow the first-order AR stochastic volatility [SV] model sets as $\varepsilon_{it} = v_{it} \exp(h_{it})$, $h_{it} = \lambda h_{it-1} + 0.5 \xi_{it}$, with $(\xi_{it}, v_{it})' \sim i.i.d. N(0, \text{diag}(\sigma_\xi^2, 1))$, independent across $i = 1, 2$. Results are reported for $\lambda = 0.951$, $\sigma_\xi = 0.314$. This case constitutes a well-known conditionally heteroskedastic model for the innovations which has been used with the same parameter configuration in many other Monte Carlo experiments such as Gonçalves and Kilian (2004), Cavaliere *et al.* (2010), and Cavaliere *et al.* (2015, 2018). The third case we consider sets ε_t to be a non-stationary, unconditionally heteroskedastic independent sequence of Gaussian variates, characterised by a late positive variance shift. Specifically,

$$\varepsilon_t \sim i.i.d. N(0, \sigma_t^2 I_2), \quad \text{with } \sigma_t := \begin{cases} 1 & \text{for } t \leq \lfloor 2T/3 \rfloor \\ 3 & \text{for } t > \lfloor 2T/3 \rfloor \end{cases}.$$

⁴For the simulation DGP in (5.1), it suffices that $(a, b, \gamma) \in (-2, 0]^2 \times [0, 1)$ in order to satisfy the I(1, r) conditions.

In order to evaluate the behaviour of the adaptive and corresponding standard procedures in practically relevant sample sizes we report results for $T = 50$ and 100 . All experiments are run over 1,000 Monte Carlo replications and were programmed using MATLAB. Our experiments are based on the no deterministic component case. In all of our simulation experiments we set $K = 4$ as the maximum lag length considered. Results for the joint information-based estimates of the co-integration rank and lag length from section 3.3 are reported first in Table A, while results relating to the sequential procedures from section 3.4 are reported in Tables B and C for the IC-based approaches in the case of SV innovations and single volatility break, respectively, and in Tables D and E for the sequential bootstrap-based procedures, again for the SV and single volatility break cases, respectively. Finally, for comparison purposes, Table F reports the results for the joint information-based approaches in the homoskedastic case.

Consider first Table A which reports results for determining the co-integration rank r (left two panels of Table A) and the lag order k (right two panels of Table A) using the joint ALS-IC-based procedures detailed in section 3.3 together with their corresponding standard information criteria-based counterparts. In particular, Table A reports the empirical frequencies with which \tilde{r} and \tilde{k} from the joint information-based estimator defined in (3.3) select the values $r = 0, 1, 2$ and $k = 1, 2, 3, 4$, respectively, for each of the adaptive criteria ALS-HQC and ALS-BIC, and the corresponding standard criteria, HQC and BIC.⁵

A number of observations can be made from the results reported in Table A. Consider first the estimators of the co-integration rank.

- In the case of autoregressive SV innovations reported in the upper portion of Table A, the performance of the adaptive version of the information criteria, i.e. ALS-HQC(k, p) and ALS-BIC(k, p), is overall superior than (or at least as good as) the standard counterparts, HQC(k, p) and BIC(k, p). The only exception is the case of no co-integration, $r_0 = 0$, where the standard BIC outperforms its adaptive counterpart. However, this is likely to be an artefact of the tendency of the standard BIC to under-fit the ‘true’ value of the co-integration rank which can be noted from the results in Table A for BIC and ALS-BIC when $r_0 > 0$.
- In the case of a single volatility break (lower portion of Table A), for a given penalty

⁵We do not consider the ALS-AIC estimator nor its standard counterpart in the Monte Carlo experiments because the poor performance of AIC-based approaches in finite samples is documented in many contributions in the literature (see e.g., Kapetanios, 2004; Wang and Bessler, 2005; Cavaliere *et al.*, 2015; Cavaliere *et al.*, 2016). Additional simulations show that, also in the case of adaptive estimation, this criterion tends to overestimate both the true co-integration rank and the lag length. Nevertheless, the adaptation with respect to the variance matrix profile considerably improves the finite sample performance of the AIC-based approach. These results are available on request.

choice, i.e. HQC or BIC, the adaptive estimator is more efficacious, and often considerably so, than the standard estimator in all but two of the cases reported in Table A. As an example, while ALS-HQC selects the correct value of r 81.4% of the time when $r_0 = 0$, $\gamma = 0.5$ and $T = 100$, the standard HQC picks the correct rank only 64.5% of the time.

- In the no co-integration case, $r_0 = 0$, the ALS-BIC penalty delivers superior performance to the ALS-HQC, with the same ordering holding for the approaches based on the standard BIC and HQC criteria. In particular we see that for both the adaptive and standard cases the HQC penalty over-fits the co-integration rank considerably more often than the BIC penalty. The degree of over-fitting is, however, smaller for the ALS-HQC *vis-à-vis* the standard HQC criterion.
- For the case where $r_0 = 1$ there is little to choose overall between the estimators based on the BIC and HQC penalties; in particular, those based on HQC again tend to over-fit the rank to a greater degree than those based on the BIC penalty, with this effect again lessened for the adaptive version of the estimator. In contrast, the BIC-based estimators can tend to under-fit for $T = 50$, excepting $\gamma = 0.9$.
- For the full rank case, $r_0 = 2$, the HQC-based estimators generally select $r_0 = 2$ more often than the corresponding BIC-based estimators, although this is likely to some degree to be an artefact of the tendency of the former to over-fit discussed above.
- For all of the estimators considered, the lag length and the magnitude of the lag parameter, γ , can have a considerable impact on the finite sample behaviour of the co-integration rank estimators. This impact seems less pronounced, other things equal for the adaptive variants of the estimates and for the BIC-based procedures relative to the HQC-based procedures.

Next consider the estimators of the autoregressive lag length.

- As observed above for co-integration rank estimation, the adaptive information criteria outperform their standard counterparts in selecting the true autoregressive lag length, k_0 , in almost all of the cases reported in Table A. These differences can again be large and generally tend to be larger, other things equal, for the HQC penalty than for the BIC penalty. As an example, while the ALS-HQC estimate of k selects the correct lag length 91.5% of the time when $r_0 = 1$, $\gamma = 0.0$ and $T = 100$, the standard HQC estimate selects the correct lag length 71.7% of the time.
- The behaviour of each lag length estimator considered is very similar, other things equal, across the three values of the co-integration rank considered. Consequently, the value of

the true co-integration rank seems to have relatively little impact on the finite sample properties of the lag length estimators.

- The HQC-based procedures are superior to the BIC-based procedures for $\gamma = 0.1$, presumably because of the greater tendency of the HQC-based procedures to over-fit, a tendency which is clearly seen for the larger values of γ considered, most notably with the non-adaptive versions of the estimators.

Consider next the results in Tables B-E which report results for the sequential estimates from section 3.4.

First focus on the results for the two-step IC-based procedures reported in Tables B and C for the cases of SV innovations and single volatility break, respectively. In particular, we report the empirical frequencies with which both standard and adaptive IC-based procedures select the lag length, k , at the first step ('Step I' in the tables) and those with which they select a co-integration rank, r , of zero, one or two at the second step ('Step II' in the tables), using the lag length estimated at the first step by each standard and adaptive information criterion, $IC(k, p)$ and $ALS-IC(k, p)$.

The results for the co-integration rank determined using the same information criterion at both steps of the sequential procedure are overall similar to the results for the corresponding joint IC-based approach. As an example, while the joint ALS-BIC estimate of r in Table A selects the correct co-integration rank 77.6% (91.3%) of the time when $r_0 = 1$, $\gamma = 0.5$ and $T = 50$ ($T = 100$) in the SV case, the corresponding sequential procedure based on ALS-BIC estimate at both steps, i.e. $ALS-BIC(\hat{k}_{ALS-BIC}, r)$, selects the true rank 77.1% (90.9%) of the time. Moreover, all of the information criteria appear to be quite robust to the choice of the information criterion at the first step of the sequential procedure as the results for the co-integration rank determination appear very similar using either $ALS-IC(k, p)$ or $IC(k, p)$.

Next consider the results for the wild bootstrap PLR procedure [denoted PLR-WB], together with the adaptive PLR procedures of Boswijk and Zu (2016) implemented with either a variance bootstrap [denoted ALR-VB] or a wild bootstrap [denoted ALR-WB] in Tables D and E for the cases of SV innovations and single volatility break, respectively. For each of these we report the empirical frequencies with which they select a co-integration rank, r , of zero, one or two. We report results for three case for the lag length used in these procedures. The first is an infeasible version based on knowledge of the true lag length, i.e. they set $k = k_0$. The other two selected the lag length in the first step of the two-step sequential procedure using either standard BIC, $k = \hat{k}_{BIC}$, or its adaptive counterpart, $k = \hat{k}_{ALS-BIC}$.⁶

⁶In Tables D and E we focus on BIC-based approaches for the selection of k because these provide the best overall performance, see e.g. Cavaliere *et al.* (2018). Moreover, we only report the results for the lag length

A number of observations can be made from the results reported in Tables D and E.

- In the no co-integration case, $r_0 = 0$, the PLR-WB procedure is seen to have better “size” properties than either the ALR-VB or ALR-WB procedures which both tend to over-estimate the co-integration rank to a greater degree than does the PLR-WB procedure. The behaviour of the ALR-VB or ALR-WB procedures for $r_0 = 0$ are fairly similar.
- In the co-integrated case, $r_0 = 1$, the ALR-VB or ALR-WB procedures both show a significantly higher empirical probability of selecting the correct rank, $r_0 = 1$, than does the PLR-WB procedure which displays a tendency to under-fit the co-integration rank most notably for $T = 50$. In the $r_0 = 1$ case the ALR-VB procedure appears to be slightly more efficacious than the ALR-WB procedure.
- In the full rank case, $r_0 = 2$, and SV innovations (Table D), the adaptive procedures provide an overall slightly better performance than the PLR-WB procedure. Conversely, in the single volatility break case (Table E), the PLR-WB procedure is, like in the zero rank case, more efficacious than either the ALR-VB or ALR-WB procedures both of which display a consistent tendency to under-fit the rank. As with the the $r_0 = 1$ case, the ALR-VB procedure appears to be slightly superior to the ALR-WB procedure for both the heteroskedastic cases considered.
- All of the PLR-WB, ALR-VB and ALR-WB procedures appear to be quite robust to the choice of the lag length made at the first step of the sequential procedure. In particular, a comparison of the results for $k = \hat{k}_{\text{BIC}}$ and $k = \hat{k}_{\text{ALS-BIC}}$ with those for a known lag length, $k = k_0$, reveals that the loss in efficacy shown by the procedures for determining the true co-integration rank at the second step due to the estimation of the unknown lag length at the first step appears very small and in some cases even negligible.
- Focusing on the results for the lag length determination reported in the right panel of Table B, we note that, overall, ALS-BIC appears to be more reliable than the corresponding standard BIC. For example, in the case of a single volatility break and $\gamma = 0.5$ the selection frequency of the true lag order, $\hat{k} = k_0 = 2$, for ALS-BIC is 82.0% (98.4%) against 74.8% (93.3%) for BIC when $T = 50$ ($T = 100$).

Finally, we consider the potential loss of efficacy of the adaptive methods in the benchmark case of homoskedastic innovations by comparing the results reported in Table F for the adaptive IC-based methods with those of their standard counterparts. These results show that the

determination for the case of $r_0 = 1$. The results for $r_0 = 0$ and 2 are very similar and thus, in the interest of space, are not reported.

performance of the joint adaptive IC-based procedures does not deteriorate much in presence of homoskedasticity, that is when they are not needed. Indeed, when either $r_0 = 1$ or 2 , the performance of the ALS-IC-based approaches is similar and sometimes even better than the corresponding standard counterparts. Conversely, in the case of no co-integration, $r_0 = 0$, standard BIC and HQC-based approaches outperform their adaptive counterparts. However, as pointed out above, this is mainly an artefact of the overall tendency of the standard criteria, especially BIC, to under-fit the true co-integration rank.

To conclude this section, we compare the small sample behaviour of the adaptive information criteria-based methods with that of the adaptive PLR test-based approaches. By comparing the results reported in Tables A, B and C with those in Tables D and E, we observe that, for the co-integrated case ($r_0 = 1$), the finite sample performance of either joint or sequential ALS-IC is similar to that of adaptive PLR test-based procedures. Conversely, when $r_0 = 0$ the PLR test-based procedures outperform the adaptive information criteria-based approaches, while this behaviour is reversed when $r_0 = 2$ and $T = 50$. In the case of full rank and $T = 100$, the performance of the methods considered are similar. Finally, by comparing the results in Tables A, B and C for the joint and the sequential information criteria-based approaches for selecting the lag length, we note that the ability of these methods to determine k is very similar to one another.

6 An Empirical Application: US Term Structure of Interest Rates

In this section we provide an empirical application of the adaptive information criteria-based approaches to the term structure of interest rates in the US. In particular, we analyse the time series $X_t = (X_{1t}, \dots, X_{5t})'$ of monthly zero yields from January 1970 to December 2012, for maturities equal to 3 months (X_{1t}), 1 year (X_{2t}), 3 years (X_{3t}), 5 years (X_{4t}), and 10 years (X_{5t}).

The co-integration analysis of X_t has already been considered by Boswijk *et al.* (2016) and Boswijk and Zu (2016). In particular, in order to account for the unconditional heteroskedasticity present in the data, sequential procedures where the lag length is selected at the first step according to (standard) $\text{HQC}(k, p)$, and then the co-integration rank of the system is determined using either PLR-WB (Boswijk *et al.*, 2016) or adaptive PLR tests (Boswijk and Zu, 2016) were adopted. Here we apply the adaptive information-based methods to estimate the co-integration rank and autoregressive lag order of the system and compare these results with those obtained in the two previous analyses cited above. In what follows, the VAR models are fitted with a restricted trend and, for all methods, the maximum number of lags considered is

$K = 4$. The number of bootstrap samples used in the bootstrap algorithms is $B = 999$.

We first focus on the joint determination of the co-integration rank and lag length using adaptive joint information criterion-based procedures as outlined in Section (3.3) and the standard counterparts. These results are reported in Table 1. The results in Table 1 show that all of the joint information criteria, both adaptive and non-adaptive, agree on selecting a lag length of $\tilde{k} = 2$. Moreover, both standard and adaptive versions of the joint BIC-based approach delivers the same estimate of the co-integration rank, namely $\tilde{r}_{\text{BIC}} = \tilde{r}_{\text{ALS-BIC}} = 2$. Conversely, the joint HQC-based approaches select a higher co-integration rank. Specifically, the co-integration rank selected using the (standard) joint HQC-based approach is 3, i.e., $\tilde{r}_{\text{HQC}} = 3$, whereas $\tilde{r}_{\text{ALS-HQC}} = 4$ is obtained using the adaptive version.

We now consider in Table 2 the results obtained using the sequential procedures for determining the lag length and then the co-integration rank. In particular, the upper panel of Table 2 shows the results for the selection of k in the first step of the sequential procedure, whereas the results for the determination of r at the second step using information criteria and PLR tests are reported in the middle and lower panels of Table 2, respectively. Note that the results reported in the lower panel of Table 2 for the case of $\hat{k} = 2$ reproduced those in Boswijk *et al.* (2016) and Boswijk and Zu (2016) who use standard HQC to select the lag length and therefore they set $\hat{k} = 2$. The results for the first step of the sequential procedure show that all of the information criteria agree with the choice of $\hat{k} = 2$, except the standard BIC which chooses $\hat{k}_{\text{BIC}} = 1$. Therefore, on balance we would recommend a VAR model of order 2.

Then we focus on the second step of the sequential procedure and, in particular, on the determination of the co-integration rank obtained by PLR tests (cf. lower panel of Table 2). For $\hat{k} = 2$, the results for the adaptive and non-adaptive bootstrap-based PLR test procedures vary according to the nominal significance level considered. In particular, at a standard 5% level we select $\hat{r} = 2$ using the (non-adaptive) PLR-WB procedure, whereas the two adaptive PLR methods yield $\hat{r} = 4$, again replicating the results in Boswijk *et al.* (2016) and Boswijk and Zu (2016), respectively. Using a 1% significance level, we still select a co-integration rank of 2 using the (standard) PLR-WB but we would now select $\hat{r} = 3$ using the two adaptive PLR test-based procedures. The results for the information criteria used in the second step of the sequential procedure show that, using $\hat{k} = 2$, HQC-based approaches in both adaptive and non-adaptive form agree with the selection at the 5% level of the adaptive PLR test-based procedures, that is $\hat{r} = 4$, whereas the co-integration rank estimated using both the adaptive and non-adaptive BIC-based approaches match the result for the (non-adaptive) PLR-WB test procedure and select $\hat{r} = 2$. It is worth noting that, when setting $\hat{k} = 1$ as suggested by the (standard) $\text{BIC}(k, p)$, the results for both information criteria and PLR tests in step 2 of the sequential procedure are much more variable across the methods with the rank selected

anywhere between 2 and 5. Therefore, we would not recommend the conclusions based on $\hat{k} = 1$. In particular, since BIC uses a stricter penalty term than HQC, we would expect, other things equal, that BIC-based approaches will often select a lower lag length and/or co-integration rank than HQC-based approaches. Moreover, this tendency of standard BIC might be sharpened by the presence of heteroskedasticity in the data, thus allowing the adaptation with the respect to the volatility process to deliver more reliable results in small samples.

In summary, overall our results seem strongly in favour of a selection of two lags. However, the selected co-integration rank varies according to the method used. In particular, the joint and sequential (for $\hat{k} = 2$) HQC-based approaches select a co-integration rank of 4, while the joint and sequential (for $\hat{k} = 2$) BIC-based approaches select rank 2. The sequential procedures based on PLR tests and $\hat{k} = 2$ select $\hat{r} = 2$ in non-adaptive form, $\hat{r} = 3$ when using a 1% significance level and $\hat{r} = 4$ when using a 5% significance level. This is in some ways consistent with the findings for BIC and HQC-based methods since decreasing the significance level is qualitatively the same as using a stricter penalty in the information criterion. Finally, it is worth noting that the choice of a rank equal to 4 implies the presence of a single stochastic trend driving the five yields and is in line with the (weak-form) expectation hypothesis of interest rates (Campbell and Shiller, 1987), which implies that the (long-term) level factor - but not the slope nor the curvature - of the interest rate yield curve is a random walk process, so that $\beta' X_t$ consists of spreads $X_{it} - X_{1t}$ for $i = 2, 3, 4, 5$.

7 Conclusions

In this paper we have proposed new methods for determining the co-integration rank and the lag order in heteroskedastic VAR models which exploit the time variation in the unconditional error variance matrix. In particular, we have proposed adaptive information criteria-based approaches to jointly determine the co-integration rank and the autoregressive lag length. Provided standard conditions hold on the penalty term hold, these methods are proved to be weakly consistent for co-integration rank and lag order determination. We have also demonstrated that the adaptive PLR rank determination procedure of Boswijk and Zu (2016), originally developed under the assumption of a known autoregressive lag length, remains asymptotically valid when a consistent lag length estimate, such as that provided by an adaptive information criterion, is used. Monte Carlo experiments reported indicate that the adaptive information criteria-based approaches generally outperform standard methods in finite samples when non-stationary volatility is present in the data.

References

- Akaike, H. (1974): A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19, 716–723.
- Boswijk, H.P. (1995): Identifiability of cointegrated systems, Tinbergen Institute Discussion Paper # 95-78, <http://dare.uva.nl/document/2/163397>.
- Boswijk, H.P., G. Cavaliere, 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.
- Boswijk, H.P. and Y. Zu (2016): Adaptive testing for cointegration with nonstationary volatility, Department of Economics Discussion Paper No. 13/08, City University of London.
- Campbell, J.Y. and R.J. Shiller (1987): Cointegration and tests of present value model, *Journal of Political Economy* 95, 1062–1088.
- Cavaliere, G. and A.M.R. Taylor (2008): Bootstrap unit root tests for time series with non-stationary volatility, *Econometric Theory* 24, 43–71.
- Cavaliere, G., L. De Angelis, A. Rahbek and A.M.R. Taylor (2018): Determining the co-integration rank in heteroskedastic VAR models of unknown order, *Econometric Theory*, 34, 349–382.
- 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 (2010): 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.
- 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.

- 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.
- Hansen, P.R. (2002): Generalized reduced rank regression, Brown University Working Paper No. 2002-02, <http://ssrn.com/abstract=302859>.
- Hansen, P.R. (2003): Structural changes in the cointegrated vector autoregressive model, *Journal of Econometrics* 114, 261–295.
- Haug, A.A. (1996): Test for cointegration: A Monte Carlo comparison, *Journal of Econometrics* 71, 89–115.
- Johansen, S. (1996): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.
- Kapetanios, G. (2004): The asymptotic distribution of the cointegration rank estimator under the Akaike information criterion, *Econometric Theory* 20, 735–742.
- 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
- McConnell, M.M. and G. Perez Quiros (2000): Output fluctuations in the United States: what has changed since the early 1980s?, *American Economic Review* 90, 1464–1476.
- Nielsen, B. (2006): Order determination in general vector autoregressions, *IMS Lecture Notes - Monograph Series* 52, 93–112.
- Patilea, V., H. Raïssi (2012): Adaptive estimation of vector autoregressive models with time-varying variance: Application to testing linear causality in mean, *Journal of Statistical Planning and Inference* 142, 2891–2912.
- Schwarz, G. (1978): Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.
- Sensier, M., and D. van Dijk (2004): Testing for volatility changes in U.S. macroeconomic time series, *Review of Economics and Statistics* 86, 833–839.
- Wang, Z. and D.A. Bessler (2005): A Monte Carlo study on the selection of cointegration rank using information criteria, *Econometric Theory* 21, 593–620.

Xu, K.-L., P.C.B. Phillips (2008): Adaptive estimation of autoregressive models with time-varying variances, *Journal of Econometrics* 142, 265–280.

Table 1: Co-integration rank and lag length determination for the term structure of interest rates in the US using standard and adaptive joint information-based procedures, $IC(k, r)$ and $ALS-IC(k, r)$.

	HQC	BIC	ALS-HQC	ALS-BIC
\tilde{k}	2	2	2	2
\tilde{r}	3	2	4	2

Note: VAR models are fitted with a restricted constant. The maximum number of lags is $K = 4$.

Table 2: Co-integration rank and lag length determination for the term structure of interest rates in the US using standard and adaptive sequential procedures.

	HQC	BIC	ALS-HQC	ALS-BIC		
Step I \hat{k}	2	1	2	2		
Step II (IC)						
$\hat{r}_{IC}(\hat{k} = 1)$	4	2	5	3		
$\hat{r}_{IC}(\hat{k} = 2)$	4	2	4	2		
Step II (PLR)						
	PLR-WB		ALR-VBS		ALR-WBS	
	$\hat{k} = 1$	$\hat{k} = 2$	$\hat{k} = 1$	$\hat{k} = 2$	$\hat{k} = 1$	$\hat{k} = 2$
$r = 0$	0.000	0.000	0.000	0.000	0.000	0.000
$r = 1$	0.000	0.000	0.000	0.000	0.000	0.000
$r = 2$	0.015	0.091	0.000	0.000	0.000	0.001
$r = 3$	0.270	0.309	0.010	0.038	0.000	0.011
$r = 4$	0.914	0.806	0.082	0.172	0.035	0.124
$\hat{r}_{(\eta=0.10)}^*$	3	3	5	4	5	4
$\hat{r}_{(\eta=0.05)}^*$	3	2	4	4	5	4
$\hat{r}_{(\eta=0.01)}^*$	2	2	3	3	4	3

Notes: VAR models are fitted with a restricted constant. The maximum number of lags is $K = 4$.

‘PLR-WB’ denotes the (non-adaptive) wild bootstrap PLR test-based approach; ‘ALR-VB’ denotes the adaptive PLR test based on the volatility bootstrap; ‘ALR-WB’ denotes the adaptive PLR test based on the wild bootstrap. The number B of bootstrap samples used in the wild bootstrap algorithm is 999.