Essex Finance Centre
Working Paper Series

Working Paper No 79: 10-2022

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

“H. Peter Boswijk, Giuseppe Cavaliere, Luca De Angelis,A. M. Robert Taylor”

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

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

H. Peter Boswijk[a], Giuseppe Cavaliere[b,c,*], 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

June 2, 2022

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 (2022) 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.

*Correspondence to: Giuseppe Cavaliere, University of Bologna, email giuseppe.cavaliere@unibo.it

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); see also McAleer (2005, 2009), Asai et al., (2006) and McAleer and Medeiros (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 (2022) 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 (2022) 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 (2022) 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 (2022) in the second step based on the first step estimate of the lag length.

The remainder of the paper is organised 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 are contained in the Appendix A.

2 The Heteroskedastic Co-integrated VAR Model

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

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

where $X_t := (X_{1t}, \ldots, X_{pt})'$ and the initial values, $X_{1-k}, \ldots, X_0$, 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 $r_0 \in \{0, \ldots, p\}$ 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 $\varepsilon_t := (\varepsilon_{1t}, \ldots, \varepsilon_{pt})'$ 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* $t = 1, \ldots, T$, *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 (2022). 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 (2022), 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). ♢

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 *et al.* (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 *et al.* (2016). ♢

3 Adaptive Information Criteria

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

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

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 $\mathcal{F}_{t-1} := \{X_{t-1}, \ldots, X_1, X_0, \ldots, X_{1-k}\}$, and the log-likelihood function is given by (see Boswijk and Zu, 2022):

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

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 (2022) 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 (2022). 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,

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

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, $i = 1, \ldots, K$, 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 (2022). In their Lemma 2, Boswijk and Zu (2022) 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 $t = 1, \ldots, T$. 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 (2022) 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

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

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)

[1]The 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), $\pi(k, r) = r(2p-r)+p^2(k-1)$; (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), $\pi(k, r) = r(2p-r+1)+p+p^2(k-1)$.

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

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

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

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

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

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

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 $\hat{k}_{\text{IC}}$. 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 $\hat{k}_{\text{IC}}$ 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 (2022) 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 (2022). Boswijk and Zu (2022) introduce the adaptive PLR statistic for testing the null hypothesis that the true co-integration rank is (no more than) r, $0 \leq r \leq p-1$,

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

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

$$\Delta X_{r,t}^* = \hat{\alpha}^{(r)}\hat{\beta}^{(r)\prime}X_{r,t-1}^* + \sum_{i=1}^{k-1}\hat{\Gamma}_i^{(r)}\Delta X_{r,t-i}^* + \varepsilon_{r,t}^*, \quad t = 1,\ldots,T, \tag{3.7}$$

initialised at $X_{r,j}^* = X_j$, for $j = 1-k,\ldots,0$, 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

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

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 (2022), 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 (2022) 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 (2022) 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

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

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

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

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 (2022) 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

[2]The 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 $r_0 = p$.

over the compact interval $[0, 1]$. This result is a basic building block needed to demonstrate weak consistency[3] 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 $\mathbb{R}$ 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 $T \to \infty$, $h \to 0$ and $Th^2 \to \infty$, 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 $T \to \infty$:*

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

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

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

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

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

[3] An 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, $T_n \xrightarrow{p} \theta$.

Remark 3. The results in Lemma 1 imply that, provided the standard condition that $\frac{c_T}{T} + \frac{1}{c_T} \to 0$, as $T \to \infty$, holds on the penalty term, c_T, then for any lag length $k = 1, \ldots, K$, 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. $\diamondsuit$

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* $\frac{c_T}{T} + \frac{1}{c_T} \to 0$ *as* $T \to \infty$.

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. $\diamondsuit$

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, 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* $T \to \infty$*:*

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

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

The results in Lemma 3 imply that $\hat{k}_{\text{ALS-IC}} \xrightarrow{p} k_0$, again provided $\frac{c_T}{T} + \frac{1}{c_T} \to 0$, as $T \to \infty$. 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 $\frac{c_T}{T} + \frac{1}{c_T} \to 0$ as $T \to \infty$, the variance and the wild bootstrap PLR-tests satisfy:*

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

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

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 $\frac{c_T}{T} + \frac{1}{c_T} \to 0$, as $T \to \infty$, the bootstrap adaptive PLR test-based procedure is an asymptotically accurately capped estimator of the true co-integration rank, r_0. ♢

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. ♢

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 $\frac{c_T}{T} + \frac{1}{c_T} \to 0$ as $T \to \infty$.*

Remark 7. It is easy to show that the condition placed 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 in Cavaliere *et al.* (2018), *inter alia*, 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. ♢

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 1-6.

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

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

with $t = 1 - K, \ldots, T$, $X_{-K} = \Delta X_{-K} = 0$, 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,

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

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

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

$K = 4$ 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 1, while results relating to the sequential procedures from Section 3.4 are reported in Tables 2 and 3 for the IC-based approaches in the case of SV innovations and single volatility break, respectively, and in Tables 4 and 5 for the sequential bootstrap-based procedures, again for the SV and single volatility break cases, respectively. Finally, for comparison purposes, Table 6 reports the results for the joint information-based approaches in the homoskedastic case. Additional Monte Carlo simulations, not reported here in the interests of space but available on request, consider the case of the presence of a "seasonal effect" in the volatility. In particular, in the spirit of Hounyo (2021, cf. Section 4), we consider the case of a periodic variation in volatility where the innovation vector in (5.1) is defined as $\varepsilon_t \sim N(0, \sigma_t^2 I_2)$, with $\sigma_t := s_t \sigma$ where $\{s_t\}$ is the repetition of the sequence $\{1, 1, 2, 4\}$ and $\sigma = 1$. Although this model does not satisfy our Assumption 1, the results we obtained nonetheless suggest that the adaptive versions of the information criteria and bootstrap-based likelihood ratio tests perform at least as well as their standard counterparts.

INSERT TABLES 1-6 HERE

Consider first Table 1 which reports results for determining the co-integration rank r (left two panels of Table 1) and the lag order k (right two panels of Table 1) using the joint ALS-IC-based procedures detailed in Section 3.3 together with their corresponding standard information criteria-based counterparts. In particular, Table 1 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. We 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.

A number of observations can be made from the results reported in Table 1. Consider first the estimators of the co-integration rank.

(i) In the case of autoregressive SV innovations reported in the upper portion of Table 1, 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) their standard counterparts, HQC(k, p) and BIC(k, p). The only exception seen is for 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 seen from the results in Table 1 for BIC and ALS-BIC when $r_0 > 0$.

(ii) In the case of a single volatility break (lower portion of Table 1), for a given penalty 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 1. 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.

(iii) 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.

(iv) For the case where $r_0 = 1$ there is overall little to choose between the estimators based on the BIC and HQC penalties; in particular, those based on the HQC penalty 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 $\gamma = 0.9$.

(v) 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.

(vi) 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 appears to be less pronounced, other things equal, for the adaptive variants of the estimates and for the BIC-based procedures relative to the corresponding HQC-based procedures.

The following observations can also be made concerning the behaviour of the estimators of the autoregressive lag length seen in Table 1.

(i) 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 1. 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.

(ii) The behaviour of each of the lag length estimators 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 would appear to have relatively little impact on the finite sample properties of the lag length estimators.

(iii) 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.

Let us now turn our attention to a discussion of the results in Tables 2-5 which relate to the sequential estimates from Section 3.4.

We first focus attention on the results reported for the two-step IC-based procedures in Tables 2 and 3 for the cases of SV innovations and a 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 where the co-integration rank is 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 approaches discussed above. As an example, the joint ALS-BIC estimate of r in Table 1 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, while the corresponding sequential procedure based on ALS-BIC estimate at both steps, i.e. ALS-BIC$(\hat{k}_{\text{ALS-BIC}}, r)$, selects the true rank 77.1% (90.9%) of the time. Moreover, all of the approaches considered appear to be fairly robust to the choice of whether to use an adaptive or standard information criterion in 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).

We now turn to a discussion of the results for the wild bootstrap PLR procedure [denoted PLR-WB], together with the adaptive PLR procedures of Boswijk and Zu (2022) implemented with either a variance bootstrap [denoted ALR-VB] or a wild bootstrap [denoted ALR-WB] in Tables 4 and 5 for the cases of SV innovations and a single break in volatility, 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. we set $k = k_0$. The other two select the lag length in the first step of the two-step sequential procedure using either standard BIC, $k = \hat{k}_{\text{BIC}}$, or its adaptive counterpart, $k = \hat{k}_{\text{ALS-BIC}}$.[5]

A number of observations can be made from the results reported in Tables 4 and 5.

(i) In the no co-integration case, $r_0 = 0$, the PLR-WB procedure is seen to have better "size" properties than either of the ALR-VB and 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.

(ii) In the co-integrated case, $r_0 = 1$, the ALR-VB and 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.

(iii) In the full rank case, $r_0 = 2$, and with SV innovations (Table 4), the adaptive procedures overall provide slightly better performance than the PLR-WB procedure. Conversely, in the single volatility break case (Table 5), the PLR-WB procedure is, as 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.

(iv) All of the PLR-WB, ALR-VB and ALR-WB procedures appear to be fairly 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 the corresponding infeasible procedures based on a known lag length, $k = k_0$, reveals that the loss in efficacy

[5] In Tables 4 and 5 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 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.

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. This is a comforting result as it suggests there are only small losses in finite sample efficacy from estimating the autoregressive lag length, relative to an infeasible benchmark based on knowledge of the true lag length.

(v) Focusing on the results for the lag length determination reported in the right panel of Table 2, we can observe 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 investigate the potential losses of efficacy seen when using the adaptive methods in the benchmark case of homoskedastic innovations by comparing the results reported in Table 6 for the adaptive IC-based methods with those of their standard counterparts. These results suggest that the performance of the joint adaptive IC-based procedures do not deteriorate to any significant degree when the shocks are homoskedastic, such that the use of adaptive methods is unnecessary. Indeed, when either $r_0 = 1$ or 2, the performance of the ALS-IC-based approaches is similar and sometimes even better than the results for their 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 finite 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 1, 2 and 3 with those in Tables 4 and 5, 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 1, 2 and 3 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 are very similar.

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 $X_t = (X_{1t}, \ldots, X_{5t})'$ 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 (2022). 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) 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, 2022) 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 7. The results in Table 7 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.

INSERT TABLE 7 HERE

We now consider in Table 8 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 8 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 8, respectively. Note that the results reported in the lower panel of Table 8 for the case of $\hat{k} = 2$ reproduce those in Boswijk *et al.*

(2016) and Boswijk and Zu (2022) 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 but the standard BIC information criteria agree on a choice of $\hat{k} = 2$; standard BIC chooses $\hat{k}_{\text{BIC}} = 1$. Therefore, on balance, we would recommend a VAR model of order 2.

Let us next focus on the second step of the sequential procedure and, in particular, on the determination of the co-integration rank obtained by the PLR tests (see the lower panel of Table 8). 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 (2022), 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 of $\hat{r} = 4$ also made at the 5% level made by the adaptive PLR test-based procedures. The co-integration rank of $\hat{r} = 2$ selected using both the adaptive and non-adaptive BIC-based approaches matches that chosen by the (non-adaptive) PLR-WB test procedure. It is worth noting that, when setting $\hat{k} = 1$ as suggested by the (standard) 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, because 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 exacerbated 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.

INSERT TABLE 8 HERE

In summary, overall our results seem strongly in favour of a selection of an autoregressive lag length of 2. 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 (see, for example, 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 (2022), 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.

Acknowledgments

This paper is dedicated to the memory of our dear friend and colleague Mike McAleer. The authors wish to thank Anders Rahbek and Yang Zu for their helpful comments and suggestions. We are grateful to the Editor, Esfandiar Maasoumi, and the anonymous referee for their helpful comments on an earlier version of this paper. The authors also thank participants at the Computation and Financial Econometrics in London (December 2015 and 2017), the Bootstrap Workshop at the Amsterdam School of Economics (November 2015), and the RCEA Time Series Econometrics Workshop in Rimini (June 2016). This research was supported by the Danish Council for Independent Research (DSF Grant 015-00028B) and the Italian Ministry of University and Research (PRIN 2020 Grant 2020B2AKFW).

References

Akaike, H. (1974): A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19, 716–723.

Asai, M., M. McAleer and J. Yu (2006): Multivariate Stochastic Volatility: A Review, *Econometric Reviews* 25, 145–175.

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 cointegration 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 (2022): Adaptive testing for cointegration with nonstationary volatility, *Journal of Business & Economic Statistics*, forthcoming; `https://doi.org/10.1080/07350015.2020.1867558`.

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 nonstationary volatility, *Econometric Theory* 24, 43–71.

Cavaliere, G., L. De Angelis, A. Rahbek and A. M. R. Taylor (2018): Determining the cointegration 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 cointegration rank in VAR models, *Econometrica* 80, 1721–1740.

Cavaliere, G., A. Rahbek and A. M. R. Taylor (2014) Bootstrap determination of the cointegration 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.

Hounyo, U. (2021): A wild bootstrap for dependent data, *Econometric Theory* forthcoming.

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.

McAleer, M. (2005): Automated inference and learning in modeling financial volatility. *Econometric Theory* 21(1), 232–261.

McAleer, M., S. Hoti and F. Chan (2009): Structure and Asymptotic Theory for Multivariate Asymmetric Conditional Volatility, *Econometric Reviews* 28, 422–440.

McAleer, M. and M. C. Medeiros (2008): Realized Volatility: A Review, *Econometric Reviews* 27, 10–45.

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

A Appendix

Notation and preliminary results. Write the unrestricted model (with $r = p$ and $k = K$) without deterministic terms as

$$\Delta X_t = \Pi X_{t-1} + \Psi Z_t + \varepsilon_t = [\Pi : \Psi] W_t + \varepsilon_t = (W_t' \otimes I_p)\, \theta + \varepsilon_t,$$

where $W_t = (X_{t-1}', Z_t')'$ with $Z_t = Z_t^{(K)} = (\Delta X_{t-1}', \dots, \Delta X_{t-K+1}')'$, and where $\theta = \text{vec}[\Pi : \Psi]$.

The lag order restriction $k < K$ implies particular zeros on $\text{vec}\, \Psi$, say,

$$\text{vec}\, \Psi = \begin{pmatrix} \psi_1^{(k)} \\ 0 \end{pmatrix}.$$

The cointegration restriction $\text{rank}\, \Pi \leq r$ implies $\Pi = \alpha\beta'$, and hence

$$\text{vec}\, \Pi = \text{vec}(\alpha\beta') = (I_p \otimes \alpha)\, \text{vec}\, (\beta')\,,$$

where α and β are $p \times r$ matrices. Depending on r, we normalise β as $c'\beta = I_r$, for some known $p \times r$ matrix c of full column rank. Defining $c_\perp$ as the orthogonal complement of c, and $\bar{c} = c(c'c)^{-1}$, this leads to $\beta = \bar{c} + c_\perp \Phi'$ for some $r \times (p - r)$ matrix Φ of unknown parameters; hence

$$\text{vec}\, \Pi = (I_p \otimes \alpha)(\text{vec}(\bar{c}') + (c_\perp \otimes I_r)\phi) =: g^{(r)}(\phi, \alpha), \tag{A.1}$$

where $\phi = \text{vec}\, \Phi$ and the function $g^{(r)}$ is implicitly defined.

With known Σ_t, minus two times the log-likelihood of the unrestricted model, up to an additive constant, is given by

$$\begin{aligned} -2\ell_T(\theta) &= \sum_{t=1}^{T} (\Delta X_t - (W_t' \otimes I_p)\, \theta)'\, \Sigma_t^{-1}\, (\Delta X_t - (W_t' \otimes I_p)\, \theta) \\ &= \sum_{t=1}^{T} \left(\hat{\varepsilon}_t - (W_t' \otimes I_p)\, (\theta - \hat{\theta})\right)'\, \Sigma_t^{-1} \left(\hat{\varepsilon}_t - (W_t' \otimes I_p)\, (\theta - \hat{\theta})\right) \\ &= \sum_{t=1}^{T} \hat{\varepsilon}_t' \Sigma_t^{-1} \hat{\varepsilon}_t + (\theta - \hat{\theta})' \sum_{t=1}^{T} \left(W_t W_t' \otimes \Sigma_t^{-1}\right) (\theta - \hat{\theta}), \end{aligned}$$

where $\hat{\theta}$ is the unrestricted MLE

$$\hat{\theta} = \left[\sum_{t=1}^{T} \left(W_t W_t' \otimes \Sigma_t^{-1}\right)\right]^{-1} \sum_{t=1}^{T} \left(W_t \otimes \Sigma_t^{-1}\right) \Delta X_t,$$

and $\hat{\varepsilon}_t = \Delta X_t - (W_t' \otimes I_p)\, \hat{\theta}$. Estimating different submodels (r, k) involves minimizing $-2\ell_T(\theta)$ over θ under the restriction

$$\theta^{(k,r)} = \begin{pmatrix} g^{(r)}(\phi, \alpha) \\ \psi_1^{(k)} \\ 0 \end{pmatrix},$$

which yields the restricted estimator $\tilde{\theta}^{(k,r)}$.

Using the true value β_0 and hence r_0, define

$$D_T = \begin{bmatrix} T^{-1}c_\perp & T^{-1/2}\beta_0 & 0 \\ 0 & 0 & T^{-1/2}I_{p(K-1)} \end{bmatrix} \otimes I_p,$$

such that

$$D_T'\left(W_t \otimes I_p\right) = T^{-1/2}\begin{pmatrix} T^{-1/2}c_\perp' X_{t-1} \\ \beta_0' X_{t-1} \\ Z_t \end{pmatrix} \otimes I_p.$$

This is used to normalise the factors of the log-likelihood ratio function:

$$\begin{aligned}
\Lambda_T(\theta) &:= -2\left[\ell_T(\theta) - \ell_T(\hat{\theta})\right] \\
&= (\theta - \hat{\theta})' \sum_{t=1}^{T}\left(W_t W_t' \otimes \Sigma_t^{-1}\right)(\theta - \hat{\theta}) \\
&= (\theta - \hat{\theta})' D_T'^{-1}\left[\sum_{t=1}^{T} D_T'\left(W_t W_t' \otimes \Sigma_t^{-1}\right) D_T\right] D_T^{-1}(\theta - \hat{\theta}).
\end{aligned}$$

Note that D_T has been defined such that $D_T^{-1}(\theta - \hat{\theta})$ and the normalised observed information matrix in square brackets are bounded in probability (and the latter has a non-singilar limit). Indeed, as shown by Boswijk and Zu (2022),

$$D_T^{-1}(\hat{\theta} - \theta_0) \xrightarrow{w} \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}^{-1}\begin{pmatrix} S_1 \\ S_2 \end{pmatrix},$$

and

$$\sum_{t=1}^{T} D_T'\left(W_t W_t' \otimes \Sigma_t^{-1}\right) D_T \xrightarrow{w} \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix},$$

where S_1 and J_1 are the limits of the normalised score vector and information matrix of the cointegration parameters ϕ, and S_2 and J_2 are the corresponding limits for the remaining parameters (α and Ψ). Furthermore, Boswijk and Zu (2022) show that the same limit results apply if the true sequence $\{\Sigma_t\}_{t=1}^{T}$ is replaced by the non-parametric estimate $\{\hat{\Sigma}_t\}_{t=1}^{T}$ in the expression for ℓ_T and hence Λ_T.

Extending the above results to the case with a (possibly restricted) constant or linear trend term requires X_{t-1} and possibly Z_t to be extended by such deterministic terms, and a corresponding extension of the matrix D_T. This will not be considered explicitly here.

Finally, it will be convenient to define $\mathrm{LR}(\mathcal{H}_{k_1,r_1}|\mathcal{H}_{k_2,r_2}) := \Lambda_T(\tilde{\theta}^{(k_1,r_1)}) - \Lambda_T(\tilde{\theta}^{(k_2,r_2)})$, the likelihood ratio statistic for $\mathcal{H}_{k_1,r_1}$ against $\mathcal{H}_{k_2,r_2}$, where (k_1, r_1) and (k_2, r_2) are particular values of (k, r) with $k_1 \leq k_2$ and $r_1 \leq r_2$. □

Proof of Lemma 1. To obtain the results of Lemma 1, we proceed to analyse

$$\text{ALS-IC}(k,r) - \text{ALS-IC}(k,r_0) = \Lambda_T(\tilde{\theta}^{(k,r)}) - \Lambda_T(\tilde{\theta}^{(k,r_0)}) + c_T\left[\pi(k,r) - \pi(k,r_0)\right],$$

where $\pi(k,r) = r(2p-r) + p^2(k-1)$. We first consider the case where $k \geq k_0$, such that the chosen lag length is well- (or over-) specified. After that, we consider the case of under-specified dynamics ($k < k_0$).

When $k \geq k_0$, then $\mathcal{H}_{k,r_0}$ is a well-specified model, and hence $\Lambda_T(\tilde{\theta}^{(k,r_0)})$ is the LR statistic for the null hypothesis that the lag length is (less than or) equal to k and the cointegrating rank is r_0 in the unrestricted model. As this null hypothesis is true, $\Lambda_T(\tilde{\theta}^{(k,r_0)})$ will have a limiting null distribution, being the distribution of the sum of the LR statistic in Boswijk and Zu (2022) and a χ^2 random variable. Most importantly, $\Lambda_T(\tilde{\theta}^{(k,r_0)}) = O_p(1)$.

For $r > r_0$, we have $\Lambda_T(\tilde{\theta}^{(k,r)}) - \Lambda_T(\tilde{\theta}^{(k,r_0)}) = -\text{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{k,r})$, which is minus the LR statistic for a true null hypothesis in an overspecified model, and hence it is $O_p(1)$. Because $\pi(k,r) - \pi(k,r_0) > 0$, it follows that

$$\Pr\left(\text{ALS-IC}(k,r) - \text{ALS-IC}(k,r_0) > 0\right) \to 1,$$

provided $c_T \to \infty$.

For $r < r_0$, we have $\Lambda_T(\tilde{\theta}^{(k,r)}) - \Lambda_T(\tilde{\theta}^{(k,r_0)}) = \text{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{k,r_0})$. In this case, the null hypothesis is violated, which will cause the statistic to diverge (to $+\infty$) at the rate $O_p(T)$. To obtain this rate, consider first the simplest case where $r = 0$ and $k_0 = k = K = 1$, so that the estimator of Ψ is zero under both constraints, and $\tilde{\theta}^{(k,r)} = \text{vec}\,\tilde{\Pi}^{(k,r)} = 0$. Therefore,

$$\begin{aligned}
\text{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{k,r_0}) &= \Lambda_T(\tilde{\theta}^{(k,r)}) - \Lambda_T(\tilde{\theta}^{(k,r_0)}) \\
&= \hat{\theta}' D_T'^{-1}\left[\sum_{t=1}^{T} D_T'\left(W_t W_t' \otimes \Sigma_t^{-1}\right) D_T\right] D_T^{-1}\hat{\theta} - \Lambda_T(\tilde{\theta}^{(k,r_0)}).
\end{aligned}$$

Since $D_T^{-1}\hat{\theta} = D_T^{-1}\theta_0 + O_p(1)$, with

$$D_T^{-1}\theta_0 = \left(\begin{bmatrix} T(\beta_{0\perp}' c_\perp)^{-1}\beta_{0\perp}' \\ T^{1/2}(c'\beta_0)^{-1}c' \end{bmatrix} \otimes I_p\right)[\beta_0 \otimes I_p]\,\text{vec}\,\alpha_0 = \begin{pmatrix} 0 \\ T^{1/2}\,\text{vec}\,\alpha_0 \end{pmatrix},$$

and $\Lambda_T(\tilde{\theta}^{(k,r_0)})$ is $O_p(1)$ as before, this leads to $\Lambda_T(\tilde{\theta}^{(k,r)}) - \Lambda_T(\tilde{\theta}^{(k,r_0)}) = O_p(T)$. More generally, we will find that the divergence rate of

$$\Lambda_T(\tilde{\theta}^{(k,r)}) = (\tilde{\theta}^{(k,r)} - \hat{\theta})' D_T'^{-1}\left[\sum_{t=1}^{T} D_T'\left(W_t W_t' \otimes \Sigma_t^{-1}\right) D_T\right] D_T^{-1}(\tilde{\theta}^{(k,r)} - \hat{\theta})$$

will be determined by

$$D_T^{-1}(\tilde{\theta}^{(k,r)} - \hat{\theta}) = D_T^{-1}(\tilde{\theta}^{(k,r)} - \theta_0) + O_p(1),$$

and since θ_0 does not lie in the constrained parameter space such that the difference $\tilde{\theta}^{(k,r)} - \theta_0$ will be $O_p(1)$ but not $o_p(1)$, it follows that $D_T^{-1}(\tilde{\theta}^{(k,r)} - \hat{\theta}) = O_p(T^{1/2})$ as before, and hence $\Lambda_T(\tilde{\theta}^{(k,r)}) = O_p(T)$. The term $c_T\,[\pi(k,r) - \pi(k,r_0)]$ is negative and diverges at the rate c_T; therefore

$$\Pr\left(\text{ALS-IC}(k,r) - \text{ALS-IC}(k,r_0) > 0\right) \to 1$$

provided $c_T/T \to 0$.

Next, consider the case $k < k_0$, so that we are comparing two (dynamically) misspecified models. When $r > r_0$, such that the larger model encompasses the correct cointegration rank, we may use the following decomposition:

$$\text{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{k,r}) = \text{LR}(\mathcal{H}_{K,r_0}|\mathcal{H}_{K,r}) + \text{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{K,r_0}) - \text{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{K,r}), \tag{A.2}$$

which follows from $\mathcal{H}_{k,r_0} \subset \mathcal{H}_{K,r_0} \subset \mathcal{H}_{K,r}$ and $\mathcal{H}_{k,r_0} \subset \mathcal{H}_{k,r} \subset \mathcal{H}_{K,r}$, and equating the sum of the LR statistics for both nested sequences of hypotheses. The first right-hand side expression in (A.2) is the LR statistic for the correct cointegration rank in a well-specified model, and hence $O_p(1)$. The second and third terms in (A.2) are LR statistics for an incorrect lag length against an unrestricted lag length. Both test statistics will diverge, but their difference is $O_p(1)$, as we will now show.

Recall the definition of $g^{(r)}(\phi,\alpha)$ in (A.1), and define the corresponding Jacobian matrix

$$G^{(r)}(\phi,\alpha) = \left[\frac{\partial g^{(r)}(\phi,\alpha)}{\partial \phi'} : \frac{\partial g^{(r)}(\phi,\alpha)}{\partial\,\text{vec}(\alpha)'}\right] = \left[(c_\perp \otimes \alpha) : (\beta \otimes I_p)\right],$$

where β is determined from ϕ as $\text{vec}\,\beta' = \text{vec}((c'c)^{-1}c') + (c_\perp \otimes I_r)\phi$. Next, define the Jacobian matrices evaluated at the true values

$$G_0 = G^{(r_0)}(\phi_0,\alpha_0), \qquad G = G^{(r)}(\phi_0^{(r)},\alpha_0^{(r)}).$$

Here (ϕ_0,α_0) is the true parameter value in the model $\mathcal{H}_{K,r_0}$, and similarly $(\phi_0^{(r)},\alpha_0^{(r)})$ is the true value in the overspecified model $\mathcal{H}_{K,r}$ with $r > r_0$. Note that ϕ and α are not identified in the over-specified model, but one can choose a true value such that $\text{vec}\,\Pi_0 = g^{(r)}(\phi_0^{(r)},\alpha_0^{(r)})$. Using a linearisation of the rank-restricted model, and hence a quadratic approximation of the log-likelihood, we have

$$\tilde{\theta}^{(K,r_0)} - \theta_0 = \begin{pmatrix} \tilde{\pi}^{(K,r_0)} - \pi_0 \\ \tilde{\psi}^{(K,r_0)} - \psi_0 \end{pmatrix} = \begin{bmatrix} G_0 & 0 \\ 0 & I \end{bmatrix} \left(\sum_{t=1}^T \mathbb{W}_t^0 \mathbb{W}_t^{0\prime}\right)^{-1} \sum_{t=1}^T \mathbb{W}_t^0 z_t + o_p(T^{-1/2}),$$

where $z_t = \sigma_t^{-1}\varepsilon_t$ (with σ_t the symmetric square root of Σ_t) and

$$\mathbb{W}_t^0 = \begin{pmatrix} \mathbb{W}_{0t}^0 \\ \mathbb{W}_{1t} \\ \mathbb{W}_{2t} \end{pmatrix} = \begin{pmatrix} G_0'(X_{t-1} \otimes \sigma_t^{-1}) \\ Z_{1t} \otimes \sigma_t^{-1} \\ Z_{2t} \otimes \sigma_t^{-1} \end{pmatrix}.$$

Here the vector of lagged differences Z_t has been partitioned into the retained lags $Z_{1t} = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1})'$ and the excluded lags $Z_{2t} = (\Delta X'_{t-k}, \ldots, \Delta X'_{t-K+1})'$ in the model $\mathcal{H}_{k,r_0}$; with coefficients ψ_1 and ψ_2, respectively. By the same quadratic approximation of the log-likelihood,

$$\mathrm{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{K,r_0}) = \tilde{\psi}_2^{(K,r_0)\prime} \sum_{t=1}^{T} \mathbb{W}^0_{2\cdot 10,t} \mathbb{W}^{0\prime}_{2\cdot 10,t} \tilde{\psi}_2^{(K,r_0)} + o_p(1), \tag{A.3}$$

with

$$\tilde{\psi}_2^{(K,r_0)} = \psi_{2,0} + \left(\sum_{t=1}^{T} \mathbb{W}^0_{2\cdot 10,t} \mathbb{W}^{0\prime}_{2\cdot 10,t} \right)^{-1} \sum_{t=1}^{T} \mathbb{W}^0_{2\cdot 10,t} z_t + o_p(T^{-1/2}),$$

and where

$$\mathbb{W}^0_{2\cdot 10,t} = \mathbb{W}_{2t} - \sum_{t=1}^{T} \mathbb{W}_{2t} (\mathbb{W}^{0\prime}_{0t}, \mathbb{W}'_{1t}) \left(\sum_{t=1}^{T} \begin{bmatrix} \mathbb{W}^0_{0t}\mathbb{W}^{0\prime}_{0t} & \mathbb{W}^0_{0t}\mathbb{W}'_{1t} \\ \mathbb{W}_{1t}\mathbb{W}^{0\prime}_{0t} & \mathbb{W}_{1t}\mathbb{W}'_{1t} \end{bmatrix} \right)^{-1} \begin{pmatrix} \mathbb{W}^0_{0t} \\ \mathbb{W}_{1t} \end{pmatrix}, \tag{A.4}$$

the least-squares residual of a regression of $\mathbb{W}_{2t}$ on $\mathbb{W}^0_{0t}$ and $\mathbb{W}_{1t}$. By the same derivations, an approximation analogous to (A.3) applies to $\mathrm{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{K,r})$ for $r > r_0$, but with $\tilde{\psi}_2^{(K,r_0)}$ replaced by $\tilde{\psi}_2^{(K,r)}$, and $\mathbb{W}^0_{2\cdot 10,t}$ replaced by $\mathbb{W}_{2\cdot 10,t}$, which in turn is defined by (A.4) with $\mathbb{W}^0_{0t}$ replaced by $\mathbb{W}_{0t} = G'(X_{t-1} \otimes \sigma_t^{-1})$. This leads to the following result:

$$\begin{aligned}
\mathrm{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{K,r_0}) - \mathrm{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{K,r}) &= \tilde{\psi}_2^{(K,r_0)\prime} \sum_{t=1}^{T} \mathbb{W}^0_{2\cdot 10,t} \mathbb{W}^{0\prime}_{2\cdot 10,t} \tilde{\psi}_2^{(K,r_0)} \\
&\quad - \tilde{\psi}_2^{(K,r)\prime} \sum_{t=1}^{T} \mathbb{W}_{2\cdot 10,t} \mathbb{W}'_{2\cdot 10,t} \tilde{\psi}_2^{(K,r)} + o_p(1) \\
&= \psi'_{2,0} \sum_{t=1}^{T} \left(\mathbb{W}^0_{2\cdot 10,t} \mathbb{W}^{0\prime}_{2\cdot 10,t} - \mathbb{W}_{2\cdot 10,t} \mathbb{W}'_{2\cdot 10,t} \right) \psi_{2,0} \\
&\quad + 2\psi'_{2,0} \sum_{t=1}^{T} \left(\mathbb{W}^0_{2\cdot 10,t} - \mathbb{W}_{2\cdot 10,t} \right) z_t \\
&\quad + \sum_{t=1}^{T} z'_t \mathbb{W}^{0\prime}_{2\cdot 10,t} \left(\sum_{t=1}^{T} \mathbb{W}^0_{2\cdot 10,t} \mathbb{W}^{0\prime}_{2\cdot 10,t} \right)^{-1} \sum_{t=1}^{T} \mathbb{W}^0_{2\cdot 10,t} z_t \\
&\quad - \sum_{t=1}^{T} z'_t \mathbb{W}'_{2\cdot 10,t} \left(\sum_{t=1}^{T} \mathbb{W}_{2\cdot 10,t} \mathbb{W}'_{2\cdot 10,t} \right)^{-1} \sum_{t=1}^{T} \mathbb{W}_{2\cdot 10,t} z_t \\
&\quad + o_p(1).
\end{aligned} \tag{A.5}$$

The third and fourth terms in the final right-hand side expression are $O_p(1)$, since they represent essentially the two likelihood ratio statistics under the null hypothesis $\psi_2 = 0$. We will now analyse the first two terms.

Because $\mathcal{H}_{K,r_0}$ is nested in $\mathcal{H}_{K,r}$, it follows that the column space of G_0 is a subset of the column space of G. Without loss of generality (after suitable rotation), we may write

$G = [G_0 : G^*]$ for some matrix G^*, orthogonal to G_0. Using standard derivations involving projection matrices, this leads to

$$\mathbb{W}^0_{2\cdot 10,t} - \mathbb{W}_{2\cdot 10,t} = \sum_{t=1}^T \mathbb{W}_{2t}\mathbb{W}^{*\prime}_{0t}\left(\sum_{t=1}^T \mathbb{W}^*_{0t}\mathbb{W}^{*\prime}_{0t}\right)^{-1}\mathbb{W}^*_{0t},$$

with $\mathbb{W}^*_{0t}$ the least-squares residual of a regression of $G^{*\prime}(X_{t-1}\otimes\sigma_t^{-1})$ on $\mathbb{W}^0_{0t}$ and $\mathbb{W}_{1t}$; and

$$\sum_{t=1}^T \left(\mathbb{W}^0_{2\cdot 10,t}\mathbb{W}^{0\prime}_{2\cdot 10,t} - \mathbb{W}_{2\cdot 10,t}\mathbb{W}_{2\cdot 10,t}\right) = \sum_{t=1}^T \mathbb{W}_{2t}\mathbb{W}^{*\prime}_{0t}\left(\sum_{t=1}^T \mathbb{W}^*_{0t}\mathbb{W}^{*\prime}_{0t}\right)^{-1}\mathbb{W}^*_{0t}\mathbb{W}'_{2t}.$$

It can be shown that $G^{*\prime}(X_{t-1}\otimes\sigma_t^{-1})$ selects I(1) linear combinations from X_{t-1}, which implies

$$\sum_{t=1}^T \mathbb{W}^*_{0t}\mathbb{W}^{*\prime}_{0t} = O_p(T^2), \qquad \sum_{t=1}^T \mathbb{W}^*_{0t}\mathbb{W}'_{2t} = O_p(T), \qquad \sum_{t=1}^T \mathbb{W}^*_{0t}z_t = O_p(T),$$

and substituting this in (A.5) leads to $\text{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{K,r_0}) - \text{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{K,r}) = O_p(1)$. Hence, because $\pi(k,r) - \pi(k,r_0) > 0$, it follows that

$$\Pr\left(\text{ALS-IC}(k,r) - \text{ALS-IC}(k,r_0) > 0\right) \to 1$$

if $c_T \to \infty$.

For $k < k_0, r < r_0$, the proof follows from a combination of ingredients from the previous two cases: now

$$\text{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{k,r_0}) = \text{LR}(\mathcal{H}_{K,r}|\mathcal{H}_{K,r_0}) + \text{LR}(\mathcal{H}_{k,r}|\mathcal{H}_{K,r}) - \text{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{K,r_0}).$$

The first right-hand side term will diverge at the rate $O_p(T)$, analogous to the result for $k \geq k_0$, $r < r_0$; and the final two terms together will be $O_p(1)$ as in the case $k < k_0$, $r \geq r_0$. This again leads to the required result. □

Proof of Lemma 2. As in the proof of Lemma 1, we start with expressing the ALS-IC difference in terms of likelihood ratio statistics and $\pi(k,r)$. For $k_0 < k \leq K$,

$$\text{ALS-IC}(k,r_0) - \text{ALS-IC}(k_0,r_0) = -\text{LR}(\mathcal{H}_{k_0,r_0}|\mathcal{H}_{k,r_0}) + c_T\left[\pi(k,r_0) - \pi(k_0,r_0)\right].$$

The first right-hand side term is an LR test statistic for a true null hypothesis in a well-specified model, and hence $O_p(1)$. Because $\pi(k,r_0) - \pi(k_0,r_0) > 0$, the ALS-IC diverges provided $c_T \to \infty$, which proves part (i).

For $0 < k < k_0$,

$$\text{ALS-IC}(k,r_0) - \text{ALS-IC}(k_0,r_0) = \text{LR}(\mathcal{H}_{k,r_0}|\mathcal{H}_{k_0,r_0}) + c_T\left[\pi(k,r_0) - \pi(k_0,r_0)\right].$$

The first right-hand side term is an LR statistic for a false null hypothesis in a well-specified model, and hence will diverge at the rate $O_p(T)$; see the proof of Lemma 1, case $k < k_0$, $r > r_0$. Since $\pi(k, r_0) - \pi(k_0, r_0) < 0$ in thise case, the ALS-IC diverges provided $c_T = o(T)$, which proves part (ii). □

Proof of Theorem 1. The theorem is a direct extension of Theorem 1 of Cavaliere *et al.* (2018) to the case of ALS-based information criteria. Making use of Lemmas 1 and 2, the line of the proof is exactly the same as in their proof. □

Proof of Lemma 3. The proof is analogous to the proof of Lemma 2; the difference is that the true cointegrating rank r_0 in Lemma 2 has been replaced here by $p \geq r_0$. Therefore, the LR test statistics are now for a true or false null hypothesis in an over-specified model; but this does not affect the divergence rates, hence the same results obtain. □

Proof of Theorem 2. It follows from Boswijk and Zu (2022), Theorem 3, that when using the true lag length k_0, the bootstrap PLR-tests have correct size and are consistent, i.e., for the chosen significance level η, and as $T \to \infty$,

$$\begin{aligned} \Pr\left(\hat{r}^*(k_0) < r\right) &\to 0, \\ \Pr\left(\hat{r}^*(k_0) = r_0\right) &\to 1 - \eta. \end{aligned}$$

Together with Lemma 3, this implies the result of Theorem 2, analogously to the proof of Theorem 3 of Cavaliere *et al.* (2018). □

Proof of Theorem 3. The theorem is a direct extension of Theorem 2 of Cavaliere *et al.* (2018) to the case of ALS-based information criteria. Making use of Lemmas 1–3, the line of the proof is exactly the same as in their proof. □

TABLE 1: Joint procedures for determining the co-integration rank and the lag length. VAR(2) model with rank $r_0 = 0, 1, 2$

		Co-integration rank determination												Lag length determination											
		ALS-HQC(k,r)			ALS-BIC(k,r)			HQC(k,r)			BIC(k,r)			ALS-HQC(k,r)			ALS-BIC(k,r)			HQC(k,r)			BIC(k,r)		
$r_0=0$		Autoregressive Stochastic Volatility																							
γ	T	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4
0.0	50	**65.6**	28.1	6.3	**81.4**	14.8	3.8	**72.0**	25.8	2.2	**89.7**	10.0	0.3	**75.9**	10.9	13.2	**88.4**	6.0	5.6	**78.8**	12.3	8.9	**92.2**	5.7	2.1
	100	**80.2**	14.5	5.3	**90.2**	6.4	3.4	**76.2**	21.7	2.1	**91.2**	8.1	0.7	**84.3**	8.1	7.6	**93.4**	2.2	4.4	**72.9**	14.6	12.5	**89.6**	7.4	3.0
0.1	50	**66.1**	27.2	6.7	**79.7**	15.9	4.4	**73.5**	23.4	3.1	**89.9**	9.4	0.7	68.9	**15.5**	15.6	82.4	**10.5**	7.1	72.9	**15.0**	12.1	88.6	**9.4**	2.0
	100	**80.7**	13.7	5.6	**90.5**	5.9	3.6	**79.1**	19.3	1.6	**93.5**	6.3	0.2	78.3	**13.9**	7.8	90.1	**5.5**	4.4	63.3	**21.8**	14.9	84.9	**11.7**	3.4
0.5	50	**65.1**	28.7	6.2	**78.7**	18.1	3.2	**68.3**	27.6	4.1	**84.3**	14.4	1.3	3.3	**76.1**	20.6	7.1	**82.0**	10.9	5.0	**78.3**	16.7	14.0	**79.8**	6.2
	100	**81.8**	12.6	5.6	**90.1**	5.9	4.0	**77.6**	20.7	1.7	**92.5**	7.1	0.4	0.2	**88.5**	11.3	1.0	**93.0**	6.0	0.7	**77.3**	22.0	2.3	**89.4**	8.3
0.9	50	**47.8**	40.8	11.4	**70.4**	23.8	5.8	**50.6**	42.1	7.3	**79.4**	18.7	1.9	0.0	**77.0**	23.0	0.0	**87.7**	12.3	0.0	**80.1**	19.9	0.0	**92.3**	7.7
	100	**71.4**	22.1	6.4	**86.7**	9.6	3.7	**69.5**	27.6	2.9	**89.5**	10.1	0.4	0.0	**88.7**	11.3	0.0	**94.5**	5.5	0.0	**78.1**	21.9	0.0	**91.8**	8.2
$r_0=1$																									
0.0	50	8.2	**73.8**	18.0	19.6	**70.1**	10.3	14.6	**71.0**	14.4	32.4	**60.4**	7.2	**79.1**	10.6	10.2	**87.8**	6.2	6.0	**79.4**	11.5	9.1	**91.8**	5.6	2.6
	100	0.4	**83.6**	16.0	4.3	**85.4**	10.3	2.5	**84.8**	12.7	7.7	**85.3**	7.0	**83.9**	7.8	8.3	**92.6**	2.6	4.8	**73.5**	14.6	11.9	**89.5**	7.3	3.2
0.1	50	9.6	**71.1**	19.3	26.3	**62.5**	11.2	15.6	**70.1**	14.3	39.8	**53.3**	6.9	68.4	**15.8**	15.8	82.1	**9.2**	8.6	71.1	**18.5**	10.4	86.9	**10.4**	2.7
	100	0.5	**84.2**	15.3	3.9	**86.8**	9.2	1.2	**84.6**	14.2	7.0	**86.0**	7.0	77.6	**14.9**	7.5	90.0	**5.3**	4.7	62.9	**23.4**	13.7	82.5	**14.1**	3.4
0.5	50	2.5	**79.9**	17.6	9.8	**77.6**	12.6	5.6	**80.0**	14.4	19.6	**73.1**	7.3	2.5	**77.3**	20.2	8.3	**79.6**	12.1	4.5	**78.5**	17.0	17.2	**76.4**	6.4
	100	0.0	**85.8**	14.2	0.4	**91.3**	8.3	0.7	**86.6**	12.7	1.9	**91.3**	6.8	0.0	**88.6**	11.4	0.6	**93.8**	5.6	0.7	**78.2**	21.1	1.9	**91.4**	6.7
0.9	50	0.3	**79.4**	20.3	0.6	**86.4**	13.0	0.3	**84.5**	15.2	1.2	**88.7**	10.1	0.0	**80.2**	19.8	0.0	**89.9**	10.1	0.0	**81.6**	18.4	0.0	**93.7**	6.3
	100	0.0	**85.9**	14.1	0.0	**91.7**	8.3	0.2	**89.3**	10.5	0.2	**94.7**	5.1	0.0	**85.8**	14.2	0.0	**92.9**	7.1	0.0	**74.7**	25.3	0.0	**89.8**	10.2
$r_0=2$																									
0.0	50	0.9	3.1	**96.0**	2.8	4.3	**92.9**	2.0	4.8	**93.2**	5.8	5.3	**88.9**	**78.0**	11.4	10.6	**87.3**	6.5	6.2	**78.7**	12.5	8.8	**91.8**	5.4	2.8
	100	0.0	0.2	**99.8**	0.1	0.4	**99.5**	0.1	1.0	**98.9**	0.2	1.3	**98.5**	**85.5**	7.2	7.3	**93.1**	2.7	4.2	**76.5**	13.1	10.4	**91.8**	6.2	2.0
0.1	50	0.9	3.1	**96.0**	4.0	3.8	**92.2**	1.8	3.5	**94.7**	7.0	4.6	**88.4**	68.2	**16.9**	14.9	82.4	**9.4**	8.2	72.6	**17.8**	9.6	87.6	**9.3**	3.1
	100	0.0	0.1	**99.9**	0.0	0.2	**99.8**	0.0	0.7	**99.3**	0.1	0.9	**99.0**	76.9	**14.7**	8.4	90.5	**4.6**	4.9	63.4	**24.9**	11.7	84.5	**12.4**	3.1
0.5	50	0.1	1.8	**98.1**	0.9	2.4	**96.7**	0.1	3.4	**96.5**	3.7	3.1	**93.2**	1.7	**79.2**	19.1	6.0	**84.1**	9.9	3.4	**81.0**	15.6	9.9	**85.5**	4.6
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.6	**99.4**	0.0	0.4	**99.6**	0.1	**88.4**	11.5	0.6	**94.5**	4.9	0.5	**79.4**	20.1	1.5	**91.3**	7.2
0.9	50	0.0	0.2	**99.8**	0.0	0.4	**99.6**	0.0	1.4	**98.6**	0.1	1.2	**98.7**	0.0	**78.4**	21.6	0.0	**88.8**	11.2	0.0	**79.5**	20.5	0.0	**91.8**	8.2
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.1	**99.9**	0.0	0.0	**100.0**	0.0	**87.6**	12.4	0.0	**93.5**	6.5	0.0	**76.8**	23.2	0.0	**91.6**	8.4
$r_0=0$		Single Volatility Break																							
0.0	50	**64.3**	27.1	8.6	**83.2**	14.2	2.6	**58.8**	31.8	9.4	**83.5**	14.8	1.7	**67.2**	15.0	17.8	**85.1**	9.2	5.7	**63.3**	16.0	20.7	**88.1**	8.7	3.2
	100	**83.1**	13.8	3.1	**95.9**	3.3	0.8	**69.7**	24.4	5.9	**91.0**	7.5	1.5	**90.3**	7.3	2.4	**97.4**	2.3	0.3	**71.4**	14.9	13.7	**93.6**	5.5	0.9
0.1	50	**63.2**	28.1	8.7	**79.3**	17.3	3.4	**58.7**	32.5	8.8	**80.1**	17.4	2.5	61.4	**21.4**	17.2	80.4	**13.5**	6.1	58.2	**21.6**	20.2	86.4	**10.3**	3.3
	100	**83.6**	13.5	2.9	**95.6**	3.4	1.0	**67.8**	26.7	5.5	**88.9**	9.8	1.3	82.4	**13.8**	3.8	95.5	**3.9**	0.6	63.2	**20.7**	16.1	88.2	**9.8**	2.0
0.5	50	**62.0**	28.5	9.5	**76.0**	19.5	4.5	**53.1**	35.7	11.2	**72.8**	22.9	4.3	2.8	**72.1**	25.1	7.0	**82.2**	10.8	6.7	**62.5**	30.8	15.4	**73.6**	11.0
	100	**81.4**	16.1	2.5	**95.1**	4.6	0.3	**64.5**	29.6	5.9	**87.5**	11.2	1.3	0.0	**92.3**	7.7	0.3	**98.6**	1.1	0.1	**76.8**	23.1	1.2	**92.5**	6.3
0.9	50	**49.4**	33.4	17.1	**71.1**	20.8	8.1	**40.7**	41.1	18.2	**70.3**	22.3	7.4	0.0	**72.9**	27.1	0.0	**87.7**	12.3	0.0	**64.4**	35.6	0.0	**88.8**	11.2
	100	**77.2**	16.6	6.2	**92.1**	6.5	1.4	**57.4**	31.2	11.4	**81.5**	15.1	3.4	0.0	**92.3**	7.7	0.0	**98.0**	2.0	0.0	**72.2**	27.8	0.0	**93.1**	6.9
$r_0=1$																									
0.0	50	8.3	**74.2**	17.5	16.2	**73.2**	10.6	8.8	**71.7**	19.5	24.9	**65.3**	9.8	**66.9**	15.1	18.0	**84.5**	8.8	6.7	**62.1**	15.9	22.0	**86.4**	9.5	4.1
	100	0.1	**89.3**	10.6	0.9	**94.3**	4.8	1.3	**79.9**	18.8	2.9	**87.2**	9.9	**91.5**	6.7	1.8	**98.4**	1.2	0.4	**71.7**	15.7	12.6	**93.3**	5.3	1.4
0.1	50	8.0	**75.1**	16.9	21.5	**68.9**	9.6	9.3	**69.8**	20.9	29.7	**60.0**	10.3	58.8	**22.5**	18.7	79.7	**12.5**	7.8	53.2	**22.7**	24.1	82.5	**12.1**	5.4
	100	0.0	**89.4**	10.6	2.3	**93.2**	4.5	0.8	**80.6**	18.6	3.0	**87.1**	9.9	81.9	**15.3**	2.8	94.9	**4.5**	0.6	59.8	**23.9**	16.3	87.5	**10.6**	1.9
0.5	50	3.0	**79.3**	17.7	9.3	**79.3**	11.4	6.3	**72.5**	21.2	15.3	**72.0**	12.7	2.7	**71.1**	26.2	8.3	**79.5**	12.2	4.5	**60.8**	34.7	14.3	**71.5**	14.2
	100	0.0	**90.1**	9.9	0.6	**94.4**	5.0	0.5	**80.3**	19.2	1.4	**89.4**	9.2	0.0	**93.6**	6.4	0.2	**98.3**	1.5	0.1	**76.0**	23.9	0.8	**93.5**	5.7
0.9	50	0.2	**78.7**	21.1	0.3	**86.7**	13.0	0.6	**72.1**	27.3	0.9	**82.6**	16.5	0.0	**71.7**	28.3	0.0	**85.9**	14.1	0.0	**62.6**	37.4	0.0	**84.8**	15.2
	100	0.0	**86.9**	13.1	0.0	**92.9**	7.1	0.0	**77.8**	22.2	0.0	**87.2**	12.8	0.0	**90.6**	9.4	0.0	**97.5**	2.5	0.0	**72.4**	27.6	0.0	**92.3**	7.7
$r_0=2$																									
0.0	50	1.6	6.3	**92.1**	3.3	5.3	**91.4**	2.8	13.3	**83.9**	5.0	9.9	**85.1**	**70.0**	12.9	17.1	**86.3**	7.5	6.2	**60.5**	16.6	22.9	**88.2**	7.6	4.2
	100	0.0	0.0	**100.0**	0.0	0.1	**99.9**	0.1	0.9	**99.0**	0.2	0.5	**99.3**	**91.4**	6.3	2.3	**98.8**	1.0	0.2	**71.9**	16.2	11.9	**92.9**	5.9	1.2
0.1	50	1.1	5.4	**93.5**	4.1	5.6	**90.3**	1.4	10.2	**88.4**	7.1	10.0	**82.9**	61.6	**18.9**	19.5	80.7	**12.6**	6.7	53.1	**20.9**	26.0	81.5	**13.0**	5.5
	100	0.0	0.1	**99.9**	0.0	0.0	**100.0**	0.0	0.7	**99.3**	0.2	0.7	**99.1**	81.8	**14.8**	3.4	94.6	**4.9**	0.5	62.1	**22.8**	15.1	85.6	**12.2**	2.2
0.5	50	0.3	3.6	**96.1**	1.5	3.9	**94.6**	0.2	7.7	**92.1**	5.1	6.3	**88.6**	2.0	**70.6**	27.4	7.0	**80.3**	12.7	4.0	**61.8**	34.2	14.0	**74.1**	11.9
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.5	**99.5**	0.2	0.5	**99.3**	0.0	**94.0**	6.0	0.2	**98.8**	1.0	0.1	**72.9**	27.0	0.6	**92.4**	7.0
0.9	50	0.1	1.4	**98.5**	0.1	1.2	**98.7**	0.0	2.3	**97.7**	0.1	2.4	**97.5**	0.0	**73.2**	26.8	0.0	**87.3**	12.7	0.0	**63.6**	36.4	0.0	**86.1**	13.9
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	**92.1**	7.9	0.0	**98.8**	1.2	0.0	**72.1**	27.9	0.0	**93.4**	6.6

TABLE 2: Sequential information criteria-based procedures for determining the lag length (Step I) and the co-integration rank (Step II). VAR(2) model with rank $r_0 = 0, 1, 2$ (AR stochastic volatility)

			ALS-HQC			ALS-BIC			HQC			BIC			ALS-HQC			ALS-BIC			HQC			BIC		
			$\gamma = 0$												$\gamma = 0.1$											
Step I		T	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4
		50	**75.6**	12.7	11.7	**86.9**	5.8	7.3	**79.2**	11.4	9.4	**92.5**	4.8	2.7	68.6	**16.8**	14.6	84.1	**8.7**	7.2	70.7	**18.8**	10.5	87.9	**9.3**	2.8
		100	**85.5**	6.7	7.8	**93.5**	2.1	4.4	**74.1**	14.4	11.5	**89.8**	7.3	2.9	78.4	**13.5**	8.1	88.3	**6.0**	5.7	64.4	**23.6**	12.0	84.1	**13.2**	2.7
Step II		T	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2
$\hat{r}_{\text{HQC}}$	$r_0=0$	50	**71.1**	26.5	2.4	**72.8**	24.8	2.3	**70.2**	27.1	2.7	**73.5**	24.4	2.1	**73.7**	23.5	2.8	**74.9**	22.0	3.1	**74.2**	22.8	3.0	**76.2**	20.8	3.0
		100	**80.2**	17.6	2.2	**80.7**	17.0	2.3	**80.2**	17.2	2.6	**80.8**	16.9	2.3	**77.5**	20.9	1.6	**78.3**	20.0	1.7	**75.9**	22.0	2.1	**77.4**	20.3	2.3
	$r_0=1$	50	9.4	**75.9**	14.7	8.4	**76.7**	14.9	9.7	**75.0**	15.3	7.8	**76.6**	15.6	11.5	**75.1**	13.4	12.3	**74.3**	13.4	11.4	**75.0**	13.6	11.8	**75.0**	13.2
		100	1.0	**83.4**	15.6	1.0	**83.3**	15.7	1.5	**82.9**	15.6	1.0	**82.7**	16.3	0.4	**87.4**	12.2	0.6	**87.2**	12.2	0.7	**86.9**	12.4	0.3	**87.1**	12.6
	$r_0=2$	50	1.7	3.4	**94.9**	1.2	2.5	**96.3**	1.1	2.5	**96.4**	0.4	2.0	**97.6**	1.9	4.4	**93.7**	1.1	3.2	**95.7**	0.9	4.4	**94.7**	0.8	3.1	**96.1**
		100	0.0	0.2	**99.8**	0.0	0.2	**99.8**	0.1	0.5	**99.4**	0.0	0.2	**99.8**	0.0	1.2	**98.8**	0.0	1.0	**99.0**	0.2	1.4	**98.4**	0.2	1.0	**98.8**
$\hat{r}_{\text{BIC}}$	$r_0=0$	50	**87.5**	11.7	0.8	**88.4**	10.6	1.0	**86.8**	12.1	1.1	**88.7**	10.4	0.9	**87.8**	11.7	0.5	**89.6**	9.9	0.5	**88.4**	10.8	0.8	**89.6**	9.6	0.8
		100	**93.9**	5.9	0.2	**93.9**	5.9	0.2	**93.4**	6.2	0.4	**94.2**	5.6	0.2	**93.2**	6.6	0.2	**93.4**	6.4	0.2	**92.4**	7.2	0.4	**92.4**	7.4	0.2
	$r_0=1$	50	30.5	**61.9**	7.6	30.4	**62.1**	7.5	30.6	**61.8**	7.6	29.0	**63.0**	8.0	35.5	**57.7**	6.8	37.2	**57.0**	5.8	35.1	**58.1**	6.8	37.3	**56.8**	5.9
		100	5.7	**86.3**	8.0	5.8	**86.4**	7.8	6.9	**85.3**	7.8	5.1	**86.6**	8.3	7.2	**85.8**	7.0	7.4	**85.9**	6.7	6.8	**86.5**	6.7	6.2	**86.9**	6.9
	$r_0=2$	50	6.1	5.4	**88.5**	4.8	3.8	**91.4**	4.3	5.0	**90.7**	3.8	2.9	**93.3**	8.9	5.7	**85.4**	7.9	4.1	**88.0**	7.5	4.9	**87.6**	6.8	3.9	**89.3**
		100	0.4	0.6	**99.0**	0.2	0.6	**99.2**	0.6	1.0	**98.4**	0.1	0.7	**99.2**	0.5	1.3	**98.2**	0.5	1.0	**98.5**	0.7	1.6	**97.7**	0.4	1.0	**98.6**
$\hat{r}_{\text{ALS-HQC}}$	$r_0=0$	50	**66.9**	24.7	8.4	**68.8**	22.4	8.8	**69.6**	23.8	6.6	**71.7**	21.7	6.5	**68.4**	22.0	9.6	**70.3**	20.0	9.7	**71.5**	21.0	7.5	**71.8**	20.3	7.8
		100	**80.8**	13.1	6.1	**81.1**	12.4	6.5	**83.6**	11.6	4.8	**83.0**	11.9	5.1	**75.1**	15.7	9.2	**74.7**	15.7	9.6	**79.7**	14.4	5.9	**76.8**	15.7	7.5
	$r_0=1$	50	10.8	**72.0**	17.2	9.6	**72.3**	18.1	11.6	**71.7**	16.8	9.3	**74.2**	16.5	12.2	**70.2**	17.7	11.6	**70.9**	17.5	12.8	**70.7**	16.4	12.0	**71.4**	16.7
		100	0.8	**82.4**	16.8	0.5	**81.8**	17.7	1.6	**83.2**	15.1	0.6	**83.2**	16.2	0.2	**82.0**	17.8	0.1	**81.5**	18.4	1.4	**83.2**	15.4	0.6	**83.2**	16.2
	$r_0=2$	50	1.6	2.8	**95.6**	0.9	1.8	**97.3**	1.6	3.0	**95.4**	0.6	1.3	**98.1**	2.0	3.5	**94.5**	0.9	1.9	**97.2**	1.9	3.4	**94.7**	1.1	1.7	**97.2**
		100	0.1	0.6	**99.3**	0.0	0.3	**99.7**	0.1	0.5	**99.4**	0.0	0.0	**100.0**	0.1	0.3	**99.6**	0.1	0.2	**99.7**	0.1	0.3	**99.6**	0.1	0.0	**99.9**
$\hat{r}_{\text{ALS-BIC}}$	$r_0=0$	50	**81.0**	12.5	6.4	**82.1**	11.1	6.7	**83.4**	11.7	4.9	**84.8**	10.3	4.9	**82.3**	10.9	6.8	**83.9**	9.0	7.1	**85.4**	10.2	4.4	**85.9**	9.6	4.5
		100	**89.6**	5.3	5.0	**89.5**	5.0	5.4	**91.8**	4.6	3.6	**91.1**	5.1	3.8	**87.2**	5.8	7.0	**86.5**	6.2	7.3	**90.1**	5.8	4.1	**87.7**	7.0	5.3
	$r_0=1$	50	31.4	**56.6**	11.9	30.2	**57.1**	12.7	33.0	**56.2**	10.9	31.1	**58.2**	10.7	31.8	**57.1**	11.2	33.0	**56.3**	10.7	33.7	**56.6**	9.7	33.9	**56.5**	9.6
		100	4.2	**84.0**	11.8	4.3	**82.8**	12.9	7.7	**83.1**	9.1	4.7	**84.6**	10.6	5.5	**83.2**	11.3	4.8	**83.0**	12.2	7.8	**83.0**	9.2	5.5	**84.7**	9.8
	$r_0=2$	50	5.7	5.1	**89.2**	4.2	3.5	**92.3**	6.0	5.4	**88.6**	4.0	2.9	**93.1**	8.5	5.1	**86.4**	7.0	3.3	**89.7**	8.3	5.0	**86.7**	7.4	3.1	**89.5**
		100	0.3	0.8	**98.9**	0.2	0.4	**99.4**	0.7	0.8	**98.5**	0.1	0.3	**99.6**	0.2	0.6	**99.2**	0.1	0.3	**99.6**	0.6	1.0	**98.4**	0.2	0.3	**99.5**

			ALS-HQC			ALS-BIC			HQC			BIC			ALS-HQC			ALS-BIC			HQC			BIC		
			$\gamma = 0.5$												$\gamma = 0.9$											
Step I		T	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4
		50	3.4	**76.4**	20.2	7.3	**83.1**	9.6	4.7	**77.0**	18.3	11.9	**82.1**	6.0	0.0	**79.0**	21.0	0.0	**89.2**	10.8	0.0	**79.6**	20.4	0.0	**91.6**	8.4
		100	0.0	**87.8**	12.2	0.3	**93.3**	6.4	0.6	**78.5**	20.9	1.5	**91.8**	6.7	0.0	**86.8**	13.2	0.0	**94.0**	6.0	0.0	**76.7**	23.3	0.0	**92.5**	7.5
Step II		T	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2
$\hat{r}_{\text{HQC}}$	$r_0=0$	50	**65.1**	30.6	4.3	**65.0**	30.8	4.2	**62.8**	32.4	4.8	**63.3**	32.2	4.5	**52.5**	39.1	8.4	**54.6**	37.7	7.7	**51.3**	40.4	8.3	**55.2**	37.0	7.8
		100	**75.6**	22.5	1.9	**75.8**	22.3	1.9	**75.9**	22.5	1.6	**75.5**	22.7	1.8	**68.2**	27.8	4.0	**68.7**	27.5	3.8	**68.2**	27.3	4.5	**68.5**	27.5	4.0
	$r_0=1$	50	3.7	**81.7**	14.6	3.6	**81.8**	14.6	4.2	**80.2**	15.6	4.4	**80.0**	15.6	0.4	**81.6**	18.0	0.4	**82.6**	17.0	0.4	**82.0**	17.6	0.3	**82.6**	17.1
		100	0.2	**87.6**	12.2	0.2	**87.6**	12.2	0.4	**87.4**	12.2	0.3	**87.5**	12.2	0.0	**87.0**	13.0	0.0	**87.6**	12.4	0.1	**87.1**	12.8	0.0	**87.0**	13.0
	$r_0=2$	50	0.2	2.8	**97.0**	0.3	2.3	**97.4**	0.6	2.8	**96.6**	0.5	2.0	**97.5**	0.0	1.5	**98.5**	0.0	1.2	**98.8**	0.0	1.5	**98.5**	0.0	1.0	**99.0**
		100	0.0	0.7	**99.3**	0.0	0.7	**99.3**	0.0	0.7	**99.3**	0.0	0.8	**99.2**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**
$\hat{r}_{\text{BIC}}$	$r_0=0$	50	**84.2**	14.5	1.3	**83.3**	15.5	1.2	**82.5**	15.6	1.9	**81.7**	16.8	1.5	**77.1**	19.5	3.4	**78.7**	18.5	2.8	**76.3**	20.1	3.6	**79.4**	17.8	2.8
		100	**91.8**	7.9	0.3	**91.8**	7.8	0.4	**90.9**	8.8	0.3	**91.1**	8.6	0.3	**88.5**	11.0	0.5	**89.0**	10.5	0.5	**88.6**	10.9	0.5	**89.0**	10.5	0.5
	$r_0=1$	50	13.3	**79.7**	7.0	13.3	**79.8**	6.9	12.9	**79.2**	7.9	13.9	**78.5**	7.6	1.4	**88.2**	10.4	1.2	**89.3**	9.5	0.8	**88.8**	10.4	0.6	**89.7**	9.7
		100	1.6	**92.7**	5.7	1.2	**93.1**	5.7	2.1	**91.9**	6.0	1.7	**92.3**	6.0	0.1	**93.7**	6.2	0.1	**94.1**	5.8	0.2	**93.0**	6.8	0.0	**93.6**	6.4
	$r_0=2$	50	1.5	4.1	**94.4**	2.1	3.3	**94.6**	1.9	3.8	**94.3**	2.7	3.0	**94.3**	0.1	2.2	**97.7**	0.0	1.6	**98.4**	0.1	2.3	**97.6**	0.0	1.3	**98.7**
		100	0.0	0.8	**99.2**	0.0	0.8	**99.2**	0.0	0.8	**99.2**	0.1	0.7	**99.2**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**
$\hat{r}_{\text{ALS-HQC}}$	$r_0=0$	50	**67.0**	24.0	9.0	**65.3**	25.1	9.6	**66.3**	25.2	8.5	**64.3**	26.8	8.9	**52.3**	34.4	13.4	**53.3**	33.7	13.0	**52.5**	34.9	12.6	**55.1**	33.2	11.7
		100	**81.6**	11.6	6.7	**81.3**	11.5	7.1	**83.4**	11.1	5.5	**82.1**	12.4	5.5	**68.1**	22.7	9.2	**68.2**	22.6	9.3	**69.7**	22.0	8.3	**69.5**	21.3	9.1
	$r_0=1$	50	4.1	**79.9**	16.0	3.8	**80.2**	16.0	3.3	**79.7**	17.0	3.2	**80.2**	16.6	0.2	**79.5**	20.3	0.2	**80.1**	19.7	0.3	**81.1**	18.6	0.3	**80.9**	18.8
		100	0.5	**86.6**	12.9	0.2	**86.5**	13.3	0.1	**85.9**	14.0	0.1	**85.7**	14.2	0.0	**82.0**	18.0	0.0	**81.7**	18.3	0.0	**83.6**	16.4	0.0	**83.7**	16.3
	$r_0=2$	50	0.5	2.1	**97.4**	0.6	1.2	**98.2**	0.8	2.3	**96.9**	0.7	1.4	**97.9**	0.0	1.1	**98.9**	0.0	1.0	**99.0**	0.0	0.9	**99.1**	0.0	0.7	**99.3**
		100	0.0	0.2	**99.8**	0.0	0.1	**99.9**	0.0	0.2	**99.8**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**
$\hat{r}_{\text{ALS-BIC}}$	$r_0=0$	50	**80.2**	14.0	5.8	**78.0**	15.2	6.8	**80.3**	14.6	5.1	**77.8**	16.2	6.0	**70.9**	20.0	9.1	**72.6**	18.8	8.6	**72.6**	19.4	8.0	**74.4**	18.2	7.3
		100	**91.0**	3.7	5.3	**90.7**	3.9	5.4	**92.7**	3.3	4.0	**92.3**	3.8	3.9	**85.1**	9.4	5.5	**85.4**	9.0	5.6	**86.0**	8.5	5.5	**86.3**	7.8	5.8
	$r_0=1$	50	13.2	**77.0**	9.8	12.8	**77.1**	10.1	11.9	**77.1**	11.0	11.2	**78.1**	10.7	0.6	**86.0**	13.4	0.5	**86.7**	12.8	1.0	**87.8**	11.2	0.7	**87.6**	11.7
		100	2.1	**90.1**	7.8	1.1	**90.9**	8.0	1.0	**90.3**	8.7	0.5	**90.3**	9.2	0.0	**87.9**	12.1	0.0	**87.8**	12.2	0.3	**89.2**	10.5	0.0	**89.8**	10.2
	$r_0=2$	50	1.5	4.2	**94.3**	1.9	2.6	**95.5**	2.6	4.0	**93.4**	2.7	2.6	**94.7**	0.2	1.7	**98.1**	0.0	1.2	**98.8**	0.3	1.6	**98.1**	0.2	1.1	**98.7**
		100	0.0	0.4	**99.6**	0.0	0.2	**99.8**	0.2	0.4	**99.4**	0.0	0.2	**99.8**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**

TABLE 3: Sequential information criteria-based procedures for determining the lag length (Step I) and the co-integration rank (Step II). VAR(2) model with rank $r_0 = 0, 1, 2$ (Single volatility break)

			ALS-HQC			ALS-BIC			HQC			BIC			ALS-HQC			ALS-BIC			HQC			BIC		
			$\gamma = 0$												$\gamma = 0.1$											
Step I		T	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4
		50	**68.5**	16.5	15.0	**86.0**	8.6	5.4	**61.6**	18.1	20.3	**87.3**	9.2	3.5	62.7	**18.1**	19.2	81.4	**12.2**	6.4	54.5	**20.9**	24.6	82.2	**13.1**	4.7
		100	**91.0**	7.0	2.0	**97.4**	1.9	0.7	**69.9**	15.4	14.7	**92.7**	6.4	0.9	81.4	**14.7**	3.9	94.7	**5.0**	0.3	62.0	**21.1**	16.9	87.1	**10.5**	2.4
Step II		T	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2
	$r_0=0$	50	**57.6**	33.8	8.6	**61.1**	30.8	8.1	**55.4**	35.3	9.3	**61.5**	30.7	7.8	**56.1**	33.9	10.0	**58.9**	32.3	8.8	**52.5**	37.6	9.9	**59.1**	32.2	8.7
		100	**65.7**	27.7	6.6	**66.8**	26.9	6.3	**63.6**	29.5	6.9	**65.6**	27.9	6.5	**66.7**	27.0	6.3	**68.1**	26.0	5.9	**64.0**	28.7	7.3	**67.3**	26.4	6.3
$\hat{r}_{\text{HQC}}$	$r_0=1$	50	6.1	**71.8**	22.1	5.4	**73.1**	21.5	7.0	**70.3**	22.7	5.0	**72.7**	22.3	8.9	**69.6**	21.5	7.8	**71.3**	20.9	8.6	**68.3**	23.1	7.6	**71.3**	21.1
		100	0.4	**79.1**	20.5	0.2	**79.0**	20.8	0.8	**78.7**	20.5	0.4	**79.1**	20.5	0.3	**80.8**	18.9	0.2	**81.6**	18.2	0.5	**79.5**	20.0	0.4	**81.0**	18.6
	$r_0=2$	50	1.2	8.0	**90.8**	0.5	5.1	**94.4**	1.3	10.1	**88.6**	0.6	4.4	**95.0**	1.1	9.3	**89.5**	0.5	6.4	**93.0**	1.4	11.7	**86.9**	0.3	6.5	**93.2**
		100	0.0	0.1	**99.9**	0.0	0.0	**100.0**	0.0	0.8	**99.2**	0.0	0.0	**100.0**	0.0	0.1	**99.9**	0.0	0.1	**99.9**	0.0	0.5	**99.5**	0.0	0.1	**99.9**
	$r_0=0$	50	**78.6**	18.7	2.7	**80.6**	16.9	2.5	**76.8**	20.2	3.0	**80.9**	16.8	2.3	**78.1**	17.0	4.9	**80.0**	15.8	4.2	**74.4**	19.9	5.7	**80.0**	15.7	4.3
		100	**88.0**	10.7	1.3	**88.3**	10.6	1.1	**85.5**	12.8	1.7	**88.0**	10.7	1.3	**88.1**	10.1	1.8	**88.9**	9.4	1.7	**87.0**	10.9	2.1	**88.1**	10.0	1.9
$\hat{r}_{\text{BIC}}$	$r_0=1$	50	21.6	**66.1**	12.3	21.3	**67.0**	11.7	22.3	**65.0**	12.7	20.8	**67.2**	12.0	26.2	**63.9**	9.9	26.8	**63.6**	9.6	25.3	**63.7**	11.0	25.4	**64.7**	9.9
		100	2.5	**87.5**	10.0	1.7	**88.5**	9.8	4.9	**85.4**	9.7	2.0	**88.0**	10.0	3.5	**86.9**	9.6	3.2	**87.5**	9.3	4.2	**85.5**	10.3	3.4	**86.5**	10.1
	$r_0=2$	50	5.5	12.1	**82.4**	3.7	8.4	**87.9**	5.7	14.3	**80.0**	3.2	7.6	**89.2**	5.9	14.5	**79.5**	5.2	10.7	**84.0**	7.1	16.1	**76.8**	5.0	10.3	**84.7**
		100	0.0	0.4	**99.6**	0.0	0.1	**99.9**	0.2	1.8	**98.0**	0.1	0.1	**99.8**	0.0	0.3	**99.7**	0.0	0.2	**99.8**	0.4	1.0	**98.6**	0.0	0.2	**99.8**
	$r_0=0$	50	**70.0**	25.6	4.4	**72.4**	23.3	4.3	**68.4**	27.3	4.3	**74.3**	22.8	2.9	**66.4**	28.1	5.5	**70.8**	24.4	4.8	**63.9**	30.7	5.4	**71.3**	23.8	4.9
		100	**84.2**	14.3	1.5	**84.8**	14.0	1.2	**83.0**	15.8	1.2	**84.9**	13.9	1.2	**83.4**	14.7	1.9	**85.4**	12.9	1.7	**82.6**	15.6	1.8	**84.6**	13.7	1.7
$\hat{r}_{\text{ALS-HQC}}$	$r_0=1$	50	8.2	**77.0**	14.8	7.4	**77.9**	14.7	10.1	**75.2**	14.7	7.1	**79.0**	13.9	9.5	**77.3**	13.2	8.7	**79.0**	12.3	9.7	**75.8**	14.5	8.3	**79.2**	12.5
		100	0.2	**88.5**	11.3	0.0	**88.8**	11.2	0.6	**87.6**	11.8	0.1	**88.7**	11.2	0.2	**89.8**	10.0	0.0	**90.5**	9.5	0.9	**89.0**	10.1	0.3	**90.0**	9.7
	$r_0=2$	50	1.0	6.9	**92.1**	0.4	3.5	**96.1**	1.4	8.0	**90.6**	0.4	2.7	**96.9**	1.6	5.4	**92.9**	0.7	2.4	**96.8**	2.2	7.9	**89.9**	0.5	2.4	**97.1**
		100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.3	**99.7**	0.0	0.0	**100.0**	0.0	0.2	**99.8**	0.0	0.1	**99.9**	0.0	0.2	**99.8**	0.0	0.0	**100.0**
	$r_0=0$	50	**83.3**	14.3	2.4	**86.2**	11.4	2.4	**83.1**	14.8	2.1	**88.1**	10.7	1.2	**83.6**	13.9	2.5	**86.9**	11.3	1.8	**82.8**	14.6	2.6	**87.1**	11.1	1.8
		100	**95.7**	3.6	0.7	**96.1**	3.3	0.6	**95.2**	4.1	0.7	**96.0**	3.4	0.6	**95.7**	3.6	0.7	**96.5**	2.9	0.6	**95.6**	3.8	0.6	**96.3**	3.1	0.6
$\hat{r}_{\text{ALS-BIC}}$	$r_0=1$	50	24.3	**66.4**	9.3	24.5	**66.5**	9.0	27.1	**64.2**	8.7	25.2	**66.8**	8.0	31.2	**62.4**	6.4	31.5	**62.9**	5.6	30.7	**62.0**	7.3	31.0	**63.8**	5.2
		100	2.4	**93.2**	4.4	1.3	**94.2**	4.5	5.4	**90.4**	4.2	2.0	**93.7**	4.3	2.9	**92.5**	4.6	2.4	**93.3**	4.3	4.7	**89.6**	5.7	2.7	**92.6**	4.7
	$r_0=2$	50	6.0	9.3	**84.7**	3.9	5.2	**90.9**	7.3	11.4	**81.3**	3.7	4.7	**91.6**	7.2	9.9	**82.8**	5.7	6.6	**87.6**	8.8	12.0	**79.2**	5.6	6.6	**87.8**
		100	0.0	0.3	**99.7**	0.0	0.1	**99.9**	0.1	1.3	**98.6**	0.0	0.1	**99.9**	0.0	0.2	**99.8**	0.0	0.1	**99.9**	0.5	0.1	**99.4**	0.0	0.0	**100.0**
			$\gamma = 0.5$												$\gamma = 0.9$											
Step I		T	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4	$k=1$	2	3,4
		50	2.4	**71.4**	26.2	7.3	**81.1**	11.6	4.1	**62.6**	33.3	12.5	**74.2**	13.3	0.0	**70.5**	29.5	0.0	**87.8**	12.2	0.0	**60.7**	39.3	0.0	**84.0**	16.0
		100	0.0	**93.1**	6.9	0.4	**98.5**	1.1	0.1	**76.5**	23.4	1.0	**93.0**	6.0	0.0	**92.2**	7.8	0.0	**98.6**	1.4	0.0	**75.9**	24.1	0.0	**94.4**	5.6
Step II		T	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2	$r=0$	1	2
	$r_0=0$	50	**50.7**	36.6	12.7	**50.1**	36.9	13.0	**47.9**	38.3	13.8	**48.1**	38.2	13.7	**40.9**	42.3	16.8	**42.9**	42.0	15.1	**39.0**	42.5	18.5	**42.4**	42.2	15.4
		100	**64.8**	28.3	6.9	**65.7**	27.7	6.6	**62.6**	29.5	7.9	**64.1**	29.0	6.9	**56.1**	32.9	11.0	**56.5**	32.7	10.8	**54.3**	34.0	11.7	**55.5**	33.4	11.1
$\hat{r}_{\text{HQC}}$	$r_0=1$	50	4.1	**74.7**	21.2	4.0	**74.7**	21.3	4.6	**74.1**	21.3	4.2	**74.4**	21.4	0.2	**71.9**	27.9	0.1	**72.9**	27.0	0.2	**70.7**	29.1	0.1	**72.6**	27.3
		100	0.1	**79.9**	20.0	0.1	**80.1**	19.8	0.3	**79.6**	20.1	0.1	**79.8**	20.1	0.0	**75.8**	24.2	0.0	**76.3**	23.7	0.0	**74.6**	25.4	0.0	**76.1**	23.9
	$r_0=2$	50	0.5	5.4	**94.1**	0.4	3.4	**96.2**	0.6	6.6	**92.8**	1.0	3.7	**95.3**	0.1	1.6	**98.3**	0.0	0.4	**99.6**	0.1	1.8	**98.1**	0.0	0.5	**99.5**
		100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.3	**99.7**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**
	$r_0=0$	50	**71.7**	22.9	5.4	**70.6**	23.9	5.5	**68.1**	25.7	6.2	**67.4**	26.5	6.1	**65.3**	27.3	7.4	**67.3**	25.9	6.8	**63.1**	28.3	8.6	**66.7**	26.2	7.1
		100	**87.1**	11.3	1.6	**87.2**	11.2	1.6	**86.0**	12.0	2.0	**86.4**	11.7	1.9	**81.0**	15.5	3.5	**81.6**	15.3	3.1	**79.8**	16.7	3.5	**80.4**	16.3	3.3
$\hat{r}_{\text{BIC}}$	$r_0=1$	50	13.8	**73.2**	13.0	12.3	**75.5**	12.2	13.7	**73.1**	13.2	13.3	**74.1**	12.6	1.0	**82.1**	16.9	0.5	**82.9**	16.6	1.3	**80.9**	17.8	0.6	**83.0**	16.4
		100	0.7	**88.6**	10.7	0.5	**88.8**	10.7	2.3	**86.5**	11.2	0.9	**88.4**	10.7	0.0	**86.1**	13.9	0.0	**86.7**	13.3	0.0	**85.9**	14.1	0.0	**86.5**	13.5
	$r_0=2$	50	2.4	8.5	**89.1**	2.5	6.0	**91.5**	3.2	10.5	**86.3**	4.5	5.7	**89.8**	0.3	2.6	**97.1**	0.0	0.9	**99.1**	0.2	3.1	**96.7**	0.1	1.2	**98.7**
		100	0.0	0.2	**99.8**	0.0	0.1	**99.9**	0.1	0.5	**99.4**	0.0	0.2	**99.8**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.1	**99.9**	0.0	0.0	**100.0**
	$r_0=0$	50	**62.5**	30.6	6.9	**61.6**	31.5	6.9	**58.4**	33.9	7.7	**59.2**	33.7	7.1	**49.5**	38.7	11.8	**53.0**	36.4	10.6	**48.4**	38.4	13.2	**53.3**	36.0	10.7
		100	**83.4**	14.9	1.7	**84.2**	14.1	1.7	**82.7**	15.9	1.4	**83.1**	15.2	1.7	**75.1**	21.4	3.5	**75.5**	21.3	3.2	**73.5**	23.0	3.5	**75.2**	21.6	3.2
$\hat{r}_{\text{ALS-HQC}}$	$r_0=1$	50	4.1	**81.1**	14.8	3.7	**82.3**	14.0	4.6	**79.4**	16.0	4.5	**81.5**	14.0	0.1	**79.5**	20.4	0.1	**81.6**	18.3	0.1	**79.0**	20.9	0.1	**81.3**	18.6
		100	0.0	**87.9**	12.1	0.0	**88.3**	11.7	0.1	**88.1**	11.8	0.0	**88.2**	11.8	0.0	**87.5**	12.5	0.0	**88.2**	11.8	0.0	**86.9**	13.1	0.0	**87.6**	12.4
	$r_0=2$	50	0.5	4.0	**95.5**	0.6	2.0	**97.4**	0.6	4.7	**94.7**	1.3	2.1	**96.6**	0.0	1.3	**98.7**	0.0	0.1	**99.9**	0.0	1.5	**98.5**	0.0	0.2	**99.8**
		100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**
	$r_0=0$	50	**77.1**	19.6	3.3	**77.2**	19.3	3.5	**75.9**	20.7	3.4	**75.3**	21.5	3.2	**71.0**	24.1	4.9	**74.7**	21.1	4.2	**70.7**	24.6	4.7	**73.9**	21.8	4.3
		100	**96.6**	3.3	0.1	**96.9**	3.0	0.1	**95.6**	4.1	0.3	**96.3**	3.6	0.1	**91.6**	7.7	0.7	**92.4**	7.1	0.5	**90.6**	8.9	0.5	**91.5**	7.9	0.6
$\hat{r}_{\text{ALS-BIC}}$	$r_0=1$	50	11.9	**79.9**	8.2	10.9	**81.0**	8.1	13.4	**78.4**	8.2	12.6	**79.7**	7.7	0.4	**87.7**	11.9	0.2	**89.3**	10.5	0.8	**86.8**	12.4	0.2	**89.5**	10.3
		100	0.3	**94.5**	5.2	0.2	**95.0**	4.8	1.2	**93.5**	5.3	0.4	**94.3**	5.3	0.0	**93.4**	6.6	0.0	**93.7**	6.3	0.0	**93.1**	6.9	0.0	**93.5**	6.5
	$r_0=2$	50	3.0	6.0	**91.0**	3.1	3.2	**93.7**	3.9	7.6	**88.5**	5.3	3.3	**91.4**	0.3	2.6	**97.1**	0.1	0.4	**99.5**	0.2	2.7	**97.1**	0.1	0.8	**99.1**
		100	0.0	0.1	**99.9**	0.0	0.0	**100.0**	0.1	0.2	**99.7**	0.0	0.1	**99.9**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**

TABLE 4: Sequential bootstrap-based procedures for determining the lag length (Step I) and the co-integration rank (Step II). VAR(2) model with rank $r_0 = 0, 1, 2$ (AR Stochastic Volatility)

		Co-integration rank determination ($r_0 = 0$)									Co-integration rank determination ($r_0 = 1$)									Lag length determination		
$k = k_0$		PLR-WB			ALR-VB			ALR-WB			PLR-WB			ALR-VB			ALR-WB			BIC		
γ	T	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$k = 1$	2	3,4
0.0	50	**93.0**	6.5	0.5	**79.5**	15.9	4.6	**82.4**	15.0	2.6	36.8	**58.6**	4.5	18.9	**71.6**	9.5	19.4	**70.5**	10.1	-	-	-
	100	**94.6**	4.8	0.6	**82.8**	13.5	3.7	**82.0**	15.1	2.9	6.3	**88.3**	5.4	0.8	**87.7**	11.5	3.0	**85.1**	11.9	-	-	-
0.1	50	**93.4**	6.0	0.6	**81.9**	14.2	3.9	**80.2**	16.5	3.3	44.7	**50.5**	4.8	25.8	**65.6**	8.5	24.9	**66.0**	9.1	-	-	-
	100	**93.5**	5.8	0.7	**87.1**	10.2	2.7	**80.2**	16.9	2.9	8.5	**86.2**	5.4	3.0	**87.9**	9.1	1.5	**86.0**	12.5	-	-	-
0.5	50	**94.0**	5.3	0.7	**82.2**	14.9	2.9	**81.0**	15.6	3.4	17.2	**77.9**	4.9	6.7	**81.6**	11.7	6.6	**81.7**	11.6	-	-	-
	100	**93.6**	5.7	0.7	**86.0**	10.8	3.2	**81.0**	15.6	3.4	1.3	**92.8**	5.9	0.0	**90.6**	9.4	0.7	**87.4**	11.9	-	-	-
0.9	50	**93.7**	5.7	0.6	**79.2**	17.4	3.4	**82.7**	14.2	3.1	1.4	**92.9**	5.7	0.3	**87.4**	12.3	0.4	**86.0**	13.6	-	-	-
	100	**93.3**	6.3	0.4	**83.0**	12.4	4.6	**79.8**	16.1	4.1	0.1	**95.1**	4.8	0.0	**87.9**	12.1	0.2	**85.5**	14.4	-	-	-
$k = \hat{k}_{\text{BIC}}$																						
0.0	50	**91.8**	7.7	0.5	**79.5**	15.9	4.6	**82.3**	15.1	2.6	35.3	**59.8**	4.9	19.0	**71.5**	9.5	19.3	**70.5**	10.2	**91.5**	6.4	2.2
	100	**94.0**	5.4	0.6	**82.7**	13.7	3.6	**81.9**	15.3	2.8	6.6	**88.0**	5.4	0.7	**87.8**	11.5	3.0	**85.3**	11.7	**90.8**	5.5	3.7
0.1	50	**93.2**	6.4	0.4	**82.0**	14.1	3.9	**81.1**	15.8	3.1	41.5	**53.7**	4.7	25.6	**65.9**	8.5	24.7	**66.4**	8.9	88.3	**9.2**	2.5
	100	**93.6**	5.7	0.7	**87.0**	10.4	2.6	**80.5**	16.1	3.4	8.1	**87.2**	4.7	2.8	**88.4**	8.8	1.7	**85.9**	12.4	83.1	**13.5**	3.5
0.5	50	**89.2**	9.0	1.8	**82.4**	14.7	2.9	**80.2**	16.4	3.4	19.0	**76.2**	4.8	6.6	**81.5**	11.9	6.8	**81.4**	11.8	10.6	**83.6**	5.7
	100	**92.8**	6.5	0.7	**85.7**	11.1	3.2	**81.1**	15.5	3.4	1.8	**92.1**	6.1	0.0	**90.6**	9.4	0.7	**87.3**	12.0	1.6	**91.8**	6.6
0.9	50	**92.1**	7.3	0.6	**79.0**	17.5	3.5	**82.7**	14.2	3.1	2.1	**91.9**	6.0	0.3	**87.5**	12.2	0.4	**85.9**	13.7	0.0	**91.9**	8.1
	100	**93.4**	6.2	0.4	**83.0**	12.5	4.5	**79.9**	16.0	4.1	0.3	**95.1**	4.6	0.0	**87.8**	12.2	0.3	**85.4**	14.4	0.0	**91.7**	8.3
$k = \hat{k}_{\text{ALS-BIC}}$																						
0.0	50	**91.8**	7.8	0.4	**79.7**	15.6	4.7	**82.5**	14.8	2.7	37.8	**57.4**	4.8	18.9	**71.5**	9.6	19.2	**70.5**	10.3	**85.9**	7.4	6.7
	100	**94.7**	4.7	0.6	**82.8**	13.2	4.0	**81.9**	15.2	2.9	7.2	**87.6**	5.2	0.8	**87.7**	11.5	2.9	**85.2**	11.9	**94.4**	1.6	4.0
0.1	50	**93.2**	6.5	0.3	**81.9**	14.4	3.7	**81.0**	15.9	3.1	42.4	**52.8**	4.8	25.5	**66.1**	8.5	24.5	**66.6**	8.9	84.4	**9.2**	6.5
	100	**94.1**	5.1	0.8	**86.8**	10.4	2.8	**80.4**	16.5	3.1	9.0	**86.9**	4.1	2.8	**88.5**	8.7	1.7	**85.8**	12.5	89.2	**5.7**	5.1
0.5	50	**90.6**	8.0	1.4	**82.1**	14.8	3.1	**80.3**	16.4	3.3	19.2	**76.2**	4.6	6.8	**81.5**	11.7	6.6	**81.7**	11.6	6.5	**83.8**	9.7
	100	**93.3**	6.0	0.7	**85.9**	11.0	3.1	**81.0**	15.5	3.5	1.8	**92.4**	5.8	0.0	**90.6**	9.4	0.7	**87.1**	12.3	0.6	**93.4**	6.0
0.9	50	**92.1**	7.2	0.7	**79.3**	17.3	3.4	**82.7**	14.2	3.1	2.7	**91.2**	6.1	0.3	**87.5**	12.2	0.3	**86.1**	13.6	0.0	**88.9**	11.1
	100	**93.4**	6.1	0.5	**82.9**	12.5	4.6	**79.9**	16.2	3.9	0.3	**95.0**	4.7	0.0	**88.1**	11.9	0.2	**85.5**	14.4	0.0	**93.2**	6.8

$k = k_0$		Co-integration rank determination ($r_0 = 2$)								
0.0	50	2.8	6.9	**90.3**	0.4	2.8	**96.8**	0.8	2.0	**97.2**
	100	0.1	1.4	**98.5**	0.0	0.4	**99.6**	0.5	0.5	**99.0**
0.1	50	6.5	9.3	**84.2**	1.4	6.5	**92.1**	1.8	4.2	**94.0**
	100	0.4	1.4	**98.2**	0.0	0.5	**99.5**	0.1	0.6	**99.3**
0.5	50	0.9	3.1	**96.0**	0.2	0.9	**98.9**	0.1	0.9	**99.0**
	100	0.1	0.3	**99.6**	0.0	0.3	**99.7**	0.0	0.4	**99.6**
0.9	50	0.1	1.3	**98.6**	0.0	0.3	**99.7**	0.0	0.5	**99.5**
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.1	**99.9**
$k = \hat{k}_{\text{BIC}}$										
0.0	50	2.7	7.4	**89.9**	0.5	2.9	**96.6**	0.7	2.1	**97.2**
	100	0.2	2.2	**97.6**	0.0	0.4	**99.6**	0.5	0.5	**99.0**
0.1	50	5.6	6.7	**87.7**	1.5	5.7	**92.8**	1.6	3.7	**94.7**
	100	0.3	1.7	**98.0**	0.0	0.7	**99.3**	0.1	0.6	**99.3**
0.5	50	2.7	3.9	**93.4**	0.1	0.8	**99.1**	0.1	1.0	**98.9**
	100	0.0	0.8	**99.2**	0.0	0.3	**99.7**	0.0	0.4	**99.6**
0.9	50	0.1	1.9	**98.0**	0.0	0.3	**99.7**	0.0	0.5	**99.5**
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.1	**99.9**
$k = \hat{k}_{\text{ALS-BIC}}$										
0.0	50	3.6	7.8	**88.6**	0.5	2.8	**96.7**	0.7	2.0	**97.3**
	100	0.2	1.8	**98.0**	0.0	0.4	**99.6**	0.5	0.5	**99.0**
0.1	50	6.1	7.4	**86.5**	1.5	5.7	**92.8**	1.6	3.7	**94.7**
	100	0.3	2.1	**97.6**	0.0	0.6	**99.4**	0.1	0.6	**99.3**
0.5	50	2.5	3.7	**93.8**	0.1	0.8	**99.1**	0.1	0.9	**99.0**
	100	0.0	0.5	**99.5**	0.0	0.3	**99.7**	0.0	0.4	**99.6**
0.9	50	0.2	2.3	**97.5**	0.0	0.3	**99.7**	0.0	0.4	**99.6**
	100	0.0	0.2	**99.8**	0.0	0.0	**100.0**	0.0	0.1	**99.9**

TABLE 5: Sequential bootstrap-based procedures for determining the lag length (Step I) and the co-integration rank (Step II). VAR(2) model with rank $r_0 = 0, 1, 2$ (Single Volatility Break)

		Co-integration rank determination ($r_0 = 0$)									Co-integration rank determination ($r_0 = 1$)									Lag length determination		
$k = k_0$		PLR-WB			ALR-VB			ALR-WB			PLR-WB			ALR-VB			ALR-WB			BIC		
γ	T	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$r = 0$	1	2	$k = 1$	2	3,4
0.0	50	**93.5**	5.8	0.7	**86.2**	11.6	2.2	**88.9**	8.9	2.2	45.9	**48.6**	5.5	15.4	**77.3**	7.3	19.0	**71.8**	9.2	-	-	-
	100	**94.5**	4.9	0.6	**92.2**	6.9	1.0	**91.9**	7.0	1.2	6.0	**88.0**	6.0	0.1	**94.0**	5.9	0.1	**93.0**	6.9	-	-	-
0.1	50	**92.6**	6.8	0.6	**84.5**	14.2	1.3	**88.3**	10.1	1.6	54.9	**39.9**	5.2	24.3	**68.5**	7.2	29.9	**61.9**	8.2	-	-	-
	100	**94.3**	5.1	0.6	**91.0**	7.6	1.4	**90.7**	7.9	1.4	12.6	**81.7**	5.7	1.3	**91.8**	6.9	1.1	**91.0**	7.9	-	-	-
0.5	50	**92.5**	7.0	0.5	**83.7**	14.8	1.5	**85.5**	12.9	1.6	28.6	**64.6**	6.8	6.3	**85.1**	8.6	9.6	**81.1**	9.3	-	-	-
	100	**94.6**	4.5	0.9	**90.6**	8.0	1.4	**89.1**	9.3	1.6	1.3	**93.3**	5.4	0.0	**93.9**	6.1	0.0	**92.9**	7.1	-	-	-
0.9	50	**93.8**	5.6	0.6	**83.9**	13.1	3.0	**88.0**	9.5	2.5	2.2	**92.4**	5.4	0.2	**91.3**	8.5	0.2	**90.2**	9.6	-	-	-
	100	**95.0**	4.3	0.7	**90.1**	8.8	1.1	**89.7**	8.5	1.8	0.0	**94.7**	5.3	0.0	**92.7**	7.3	0.0	**91.2**	8.8	-	-	-
$k = \hat{k}_{\text{BIC}}$																						
0.0	50	**90.3**	8.6	1.1	**86.1**	11.7	2.2	**88.9**	9.0	2.1	43.6	**50.3**	6.1	15.4	**77.2**	7.4	19.1	**71.9**	9.0	**86.7**	9.5	3.8
	100	**93.6**	5.7	0.7	**92.3**	6.8	1.0	**92.0**	6.9	1.2	6.6	**87.4**	6	0.1	**94.2**	5.7	0.1	**93.0**	6.9	**92.4**	5.7	1.9
0.1	50	**88.6**	10.3	1.1	**84.2**	14.1	1.7	**88.6**	9.8	1.6	51.3	**42.6**	6.1	23.2	**69.9**	6.9	30.0	**61.8**	8.2	80.7	**13.5**	5.8
	100	**91.7**	7.8	0.5	**90.4**	8.2	1.4	**89.6**	8.9	1.5	10.9	**83.7**	5.4	1.3	**92.2**	6.5	1.1	**90.8**	8.1	87.0	**10.9**	2.1
0.5	50	**83.6**	14.7	1.7	**83.7**	14.8	1.5	**85.6**	12.7	1.7	28.9	**63.7**	7.4	6.2	**85.2**	8.6	9.7	**81.3**	9.0	12.8	**74.8**	12.4
	100	**93.4**	5.8	0.8	**90.3**	8.3	1.4	**89.0**	9.4	1.6	2.0	**92.5**	5.5	0.0	**93.8**	6.2	0.0	**93.0**	7.0	1.3	**93.3**	5.4
0.9	50	**90.7**	8.3	1.0	**84.0**	13.1	2.9	**87.8**	9.6	2.6	3.0	**90.9**	6.1	0.1	**91.4**	8.5	0.3	**90.4**	9.3	0.0	**86.6**	13.4
	100	**93.9**	5.2	0.9	**90.0**	8.9	1.1	**89.8**	8.6	1.6	0.0	**94.7**	5.3	0.0	**92.6**	7.4	0.0	**91.1**	8.9	0.0	**92.0**	8.0
$k = \hat{k}_{\text{ALS-BIC}}$																						
0.0	50	**90.6**	8.4	1.0	**86.0**	11.8	2.2	**88.9**	9.0	2.1	43.9	**50.2**	5.9	15.7	**76.9**	7.4	18.9	**72.1**	9.0	**85.1**	9.6	5.3
	100	**94.5**	4.9	0.6	**92.2**	6.9	1.0	**91.9**	7.0	1.2	6.2	**87.7**	6.1	0.1	**94.0**	5.9	0.1	**93.0**	6.9	**98.6**	1.2	0.2
0.1	50	**88.8**	10.0	1.2	**84.4**	13.8	1.8	**88.5**	9.9	1.6	52.4	**41.5**	6.1	23.3	**69.4**	7.3	30.3	**61.5**	8.2	78.4	**14.2**	7.4
	100	**91.8**	7.8	0.4	**90.5**	8.1	1.4	**89.6**	8.9	1.5	11.4	**83.5**	5.1	1.3	**92.1**	6.6	0.9	**91.2**	7.9	95.5	**4.1**	0.4
0.5	50	**86.9**	11.6	1.5	**83.7**	14.8	1.5	**85.8**	12.5	1.7	28.8	**64.2**	7.0	6.3	**85.1**	8.6	9.5	**81.5**	9.0	7.2	**82.0**	10.8
	100	**94.7**	4.4	0.9	**90.6**	8.0	1.4	**89.1**	9.3	1.6	1.6	**93.1**	5.3	0.0	**93.9**	6.1	0.0	**92.9**	7.1	0.5	**98.4**	1.1
0.9	50	**91.7**	7.4	0.9	**83.9**	13.1	3.0	**87.7**	9.9	2.4	2.8	**91.3**	5.9	0.2	**91.3**	8.5	0.2	**90.4**	9.4	0.0	**88.3**	11.7
	100	**94.9**	4.3	0.8	**90.1**	8.8	1.1	**89.7**	8.5	1.8	0.0	**94.7**	5.3	0.0	**92.7**	7.3	0.0	**91.1**	8.9	0.0	**97.8**	2.2

$k = k_0$		Co-integration rank determination ($r_0 = 2$)								
0.0	50	5.1	15.1	**79.8**	0.2	25.8	**74.0**	0.5	32.0	**67.5**
	100	0.0	0.5	**99.5**	0.0	11.9	**88.1**	0.0	17.7	**82.3**
0.1	50	12.9	18.4	**68.7**	1.5	29.7	**68.8**	2.7	31.7	**65.6**
	100	0.1	1.0	**98.9**	0.0	14.0	**86.0**	0.0	18.1	**81.9**
0.5	50	1.4	9.8	**88.8**	0.1	19.0	**80.9**	0.2	23.4	**76.4**
	100	0.0	0.1	**99.9**	0.0	10.6	**89.4**	0.0	16.6	**83.4**
0.9	50	0.0	1.1	**98.9**	0.0	14.8	**85.2**	0.0	18.8	**81.2**
	100	0.0	0.0	**100.0**	0.0	14.7	**85.3**	0.0	18.5	**81.5**
$k = \hat{k}_{\text{BIC}}$										
0.0	50	5.5	18.4	**76.1**	0.2	25.5	**74.3**	0.4	31.9	**67.7**
	100	0.1	1.1	**98.8**	0.0	12.1	**87.9**	0.0	17.9	**82.1**
0.1	50	9.0	19.2	**71.8**	1.4	29.8	**68.8**	3.2	29.9	**66.9**
	100	0.0	1.1	**98.9**	0.0	14.2	**85.8**	0.0	18.6	**81.4**
0.5	50	4.4	12.4	**83.2**	0.1	18.8	**81.1**	0.2	23.8	**76.0**
	100	0.0	0.3	**99.7**	0.0	10.7	**89.3**	0.0	16.7	**83.3**
0.9	50	0.4	1.9	**97.7**	0.0	14.4	**85.6**	0.0	18.9	**81.1**
	100	0.0	0.0	**100.0**	0.0	14.5	**85.5**	0.0	18.7	**81.3**
$k = \hat{k}_{\text{ALS-BIC}}$										
0.0	50	5.9	18.7	**75.4**	0.2	25.8	**74.0**	0.4	31.9	**67.7**
	100	0.1	0.7	**99.2**	0.0	11.9	**88.1**	0.0	17.7	**82.3**
0.1	50	9.6	19.8	**70.6**	1.4	29.6	**69.0**	3.2	29.6	**67.2**
	100	0.0	0.4	**99.6**	0.0	13.9	**86.1**	0.0	18.7	**81.3**
0.5	50	2.9	12.8	**84.3**	0.1	18.7	**81.2**	0.1	23.7	**76.2**
	100	0.0	0.1	**99.9**	0.0	10.7	**89.3**	0.0	16.5	**83.5**
0.9	50	0.2	1.9	**97.9**	0.0	14.5	**85.5**	0.0	18.9	**81.1**
	100	0.0	0.0	**100.0**	0.0	14.6	**85.4**	0.0	18.6	**81.4**

TABLE 6: Sequential joint procedures for determining the co-integration rank and the lag length. VAR(2) model with rank $r_0 = 0, 1, 2$ (Homoskedastic DGP)

		Co-integration rank determination												Lag length determination											
		ALS-HQC(k,r)			ALS-BIC(k,r)			HQC(k,r)			BIC(k,r)			ALS-HQC(k,r)			ALS-BIC(k,r)			HQC(k,r)			BIC(k,r)		
$r_0 = 0$																									
γ	T	r = 0	1	2	r = 0	1	2	r = 0	1	2	r = 0	1	2	k = 1	2	3,4	k = 1	2	3,4	k = 1	2	3,4	k = 1	2	3,4
0.0	50	**69.3**	27.3	3.4	**86.9**	11.9	1.2	**82.9**	16.3	0.8	**95.9**	3.9	0.2	**89.3**	6.3	4.4	**96.6**	2.3	1.1	**96.5**	2.6	0.9	**99.8**	0.2	0.0
	100	**84.3**	14.6	1.1	**96.2**	3.5	0.3	**89.1**	10.3	0.6	**98.5**	1.4	0.1	**97.8**	2.0	0.2	**99.6**	0.4	0.0	**99.1**	0.9	0.0	**99.7**	0.3	0.0
0.1	50	**69.6**	25.6	4.8	**86.6**	11.5	1.9	**82.6**	14.8	2.6	**96.0**	3.6	0.4	81.6	**13.1**	5.3	95.5	**3.8**	0.7	93.8	**5.3**	0.9	98.6	**1.4**	0.0
	100	**83.0**	14.9	2.1	**95.7**	3.9	0.4	**88.6**	9.4	2.0	**97.8**	2.1	0.1	87.2	**11.8**	1.0	97.1	**2.9**	0.0	92.0	**7.6**	0.4	98.4	**1.6**	0.0
0.5	50	**64.9**	29.7	5.4	**83.3**	15.2	1.5	**77.0**	20.6	2.4	**92.4**	6.5	1.1	2.0	**88.3**	9.7	4.8	**93.1**	2.1	3.0	**93.0**	4.0	8.6	**91.2**	0.2
	100	**79.9**	18.0	2.1	**94.5**	5.3	0.2	**84.3**	14.2	1.5	**97.4**	2.4	0.2	0.0	**96.6**	3.4	0.1	**99.6**	0.3	0.0	**97.9**	2.1	0.1	**99.8**	0.1
0.9	50	**44.1**	44.5	11.4	**67.7**	29.1	3.2	**57.3**	36.3	6.4	**82.3**	15.8	1.9	0.0	**87.3**	12.7	0.0	**96.8**	3.2	0.0	**94.3**	5.7	0.0	**99.0**	1.0
	100	**71.1**	25.1	3.8	**92.0**	7.3	0.7	**77.7**	19.4	2.9	**95.3**	4.4	0.3	0.0	**96.4**	3.6	0.0	**99.7**	0.3	0.0	**98.0**	2.0	0.0	**99.9**	0.1
$r_0 = 1$																									
0.0	50	2.4	**79.2**	18.4	11.9	**78.9**	9.2	5.9	**81.0**	13.1	30.3	**65.4**	4.3	**88.1**	8.4	3.5	**96.6**	2.8	0.6	**94.8**	4.1	1.1	**99.2**	0.8	0.0
	100	0.0	**90.8**	9.2	0.4	**95.4**	4.2	0.1	**92.6**	7.3	0.9	**95.6**	3.5	**96.8**	2.7	0.5	**99.6**	0.4	0.0	**97.7**	2.1	0.2	**99.8**	0.2	0
0.1	50	2.7	**82.8**	14.5	16.4	**75.1**	8.5	8.8	**80.4**	10.8	40.8	**55.0**	4.2	82.3	**14.6**	3.1	94.4	**4.8**	0.8	90.4	**8.7**	0.9	98.6	**1.4**	0
	100	0.0	**89.1**	10.9	0.2	**95.3**	4.5	0.0	**91.2**	8.8	1.3	**95.3**	3.4	86.4	**12.4**	1.2	97.0	**3.0**	0.0	90.4	**9.1**	0.5	98.4	**1.6**	0
0.5	50	0.8	**85.0**	14.2	3.7	**89.2**	7.1	1.7	**87.0**	11.3	11.7	**83.1**	5.2	1.0	**88.6**	10.4	4.1	**92.7**	3.2	1.6	**93.1**	5.3	11.2	**88**	0.8
	100	0.0	**86.8**	13.2	0.0	**94.5**	5.5	0.0	**88.4**	11.6	0.0	**95.8**	4.2	0.0	**96.1**	3.9	0.0	**99.7**	0.3	0.0	**97.6**	2.4	0.0	**100**	0
0.9	50	0.0	**80.0**	20.0	0.1	**90.2**	9.7	0.0	**83.7**	16.3	0.1	**92.9**	7.0	0.0	**88.0**	12.0	0.0	**97.1**	2.9	0.0	**94.0**	6.0	0.0	**99.2**	0.8
	100	0.0	**90.1**	9.9	0.0	**96.3**	3.7	0.0	**91.3**	8.7	0.0	**97.5**	2.5	0.0	**94.9**	5.1	0.0	**99.2**	0.8	0.0	**97.0**	3.0	0.0	**99.6**	0.4
$r_0 = 2$																									
0.0	50	0.2	0.8	**99.0**	0.7	0.3	**99.0**	0.2	0.4	**99.4**	1.5	0.6	**97.9**	**89.3**	7.5	3.2	**96.9**	2.6	0.5	**96.0**	3.1	0.9	**98.9**	1.0	0.1
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	**96.8**	2.9	0.3	**99.7**	0.3	0.0	**97.4**	2.5	0.1	**100.0**	0.0	0.0
0.1	50	0.0	0.6	**99.4**	0.4	0.6	**99.0**	0.0	0.1	**99.9**	3.5	1.0	**95.5**	82.7	**12.8**	4.5	94.0	**5.2**	0.8	91.0	**7.8**	1.2	98.3	**1.6**	0.1
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	87.6	**11.9**	0.5	97.4	**2.6**	0.0	91.3	**8.6**	0.1	98.5	**1.5**	0.0
0.5	50	0.0	0.4	**99.6**	0.2	0.4	**99.4**	0.0	0.3	**99.7**	2.4	0.2	**97.4**	0.5	**88.9**	10.6	3.3	**93.7**	3.0	1.1	**93.9**	5.0	7.7	**91.1**	1.2
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	**96.2**	3.8	0.0	**99.6**	0.4	0.0	**97.4**	2.6	0.1	**99.9**	0.0
0.9	50	0.0	0.1	**99.9**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	**88.8**	11.2	0.0	**97.1**	2.9	0.0	**94.9**	5.1	0.0	**99.2**	0.8
	100	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	0.0	**100.0**	0.0	**96.2**	3.8	0.0	**99.9**	0.1	0.0	**97.5**	2.5	0.0	**100.0**	0.0

TABLE 7: 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 8: 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)				
$\hat{r}_{\text{IC}}(\hat{k}=1)$	4	2	5	3
$\hat{r}_{\text{IC}}(\hat{k}=2)$	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.