Harris, David and Leybourne, Stephen J. and Taylor, A.M. Robert (2016) Tests of the co-integration rank in VAR models in the presence of a possible break in trend at an unknown point. Journal of Econometrics, 192 (24). pp. 451-467. ISSN 0304-4076

Tests of the Co-integration Rank in VAR Models in the Presence of a Possible Break in Trend at an Unknown Point*

David Harris[a], Stephen J. Leybourne[b] and A.M. Robert Taylor[c]

[a] Department of Econometrics and Business Statistics, Monash University

[b] School of Economics, University of Nottingham

[c] Essex Business School, University of Essex

November 2015

Abstract

In this paper we consider the problem of testing for the co-integration rank of a vector autoregressive process in the case where a trend break may potentially be present in the data. It is known that un-modelled trend breaks can result in tests which are incorrectly sized under the null hypothesis and inconsistent under the alternative hypothesis. Extant procedures in this literature have attempted to solve this inference problem but require the practitioner to either assume that the trend break date is known or to assume that any trend break cannot occur under the co-integration rank null hypothesis being tested. These procedures also assume the autoregressive lag length is known to the practitioner. All of these assumptions would seem unreasonable in practice. Moreover in each of these strands of the literature there is also a presumption in calculating the tests that a trend break is known to have happened. This can lead to a substantial loss in finite sample power in the case where a trend break does not in fact occur. Using information criteria based methods to select both the autoregressive lag order and to choose between the trend break and no trend break models, using a consistent estimate of the break fraction in the context of the former, we develop a number of procedures which deliver asymptotically correctly sized and consistent tests of the co-integration rank regardless of whether a trend break is present in the data or not. By selecting the no break model when no trend break is present, these procedures also avoid the potentially large power losses associated with the extant procedures in such cases.

Keywords: Co-integration rank; vector autoregression; error-correction model; trend break; break point estimation; information criteria.

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

*We are grateful to the Guest Editor, Jörg Breitung, and three anonymous referees for their helpful and constructive comments on earlier versions of this paper. Taylor gratefully acknowledges financial support provided by the Economic and Social Research Council of the United Kingdom under research grant ES/M01147X/1. Correspondence to: Robert Taylor, Essex Business School, University of Essex, Colchester, CO4 3SQ, U.K. *E-mail*: robert.taylor@essex.ac.uk

1 Introduction

Macroeconomic series are typically characterized by piecewise linear (or broken) trend functions; see, inter alia, Stock and Watson (1996, 1999, 2005) and Perron and Zhu (2005). Such breaks in the trend function might occur following a period of major economic upheaval or a political regime change.

In the univariate setting this has spurred a large literature on testing for an autoregressive unit root when a trend break may be present in the data. The first proper theoretical treatment of this problem was given by Perron (1989) who showed that unit root tests which fail to account for a trend break present in the data have non-pivotal limiting null distributions and are inconsistent under stable root alternatives. Assuming the putative break date to be known, Perron (1989) proposed new unit root tests which avoid these problems by modelling the trend break. However, if a break does not occur this approach loses considerable finite sample power through the inclusion of an unnecessary trend break regressor. Subsequent approaches have focussed on the case where the break date is unknown. Zivot and Andrews (1992) base a test on the most negative of a sequence, taken across all possible break dates, of the Perron (1989) statistics, while Perron (1997) first estimates the trend break location and then uses the Perron (1989) test for the estimated break date. The limiting distributions of the Zivot and Andrews (1992) tests depend on the magnitude of the trend break parameter which renders them infeasible in practice. The Perron (1997) approach is also problematic in that the break point estimator has a non-degenerate limit distribution when no break is present, with the result that the associated unit root test has a different large sample null distribution *vis-à-vis* the case where a trend break is present. Size-controlled inference can then only be achieved by using so-called conservative critical values corresponding to the case where no break is present, with an associated loss of efficiency where a break is present. As a result, Carrion-i-Silvestre *et al.* (2009), Harris *et al.* (2009) and Kim and Perron (2009) advocate approaches based on the use of pre-tests for the presence of a trend break.

In the vector time series setting, un-modelled trend breaks cause similar problems for the co-integration rank tests of Johansen (1995). For example, Inoue (1999) documents large losses in finite sample power with the standard trace and maximum eigenvalue tests of Johansen (1995) when an un-modelled trend break is present in the data. As we will show in the simulation results we report in this paper, an un-modelled trend break also causes substantial over-sizing in the standard rank tests, consistent with the findings for standard unit root tests in Perron (1989). Surprisingly then, the literature on testing for co-integration rank in the presence of breaks in the deterministic trend function is relatively sparse compared to the univariate case.

In the context of co-integration rank tests of the type considered in Johansen (1995), Johansen *et al.* (2000) develop likelihood ratio tests, analogous to those considered in the univariate case in Perron (1989), for the case where the break in the trend function occurs at a *known* point. Like Perron (1989) they consider both level break and trend break models, and extend to allow for multiple breaks in the trend function. Saikkonen and Lütkepohl (2000) for a level break (but no trend break) at a known date, Lütkepohl *et al.* (2003) for a level break (no trend break) at an *unknown* point, and Trenkler *et al.* (2007) for a trend break at a known date, propose further co-integration rank tests, in each case using the pseudo-GLS de-trending method outlined in Saikkonen and Lütkepohl (2000). All of these procedures assume that the autoregressive lag length is known to the practitioner. The approaches taken in the last three of these papers also differ from the approach taken in Johansen *et al.* (2000) according to how the data generating process [DGP] under consideration is constructed. While they adopt a components DGP, forming the observed process as the sum of the deterministic variables and an indeterministic vector autoregressive [VAR] process, Johansen *et al.* (2000), follow Johansen (1995) and place the deterministic variables directly into the VAR equation. Finally, Inoue (1999), who also assumes a known autoregressive lag order, develops Zivot and Andrews (1992) type co-integration rank tests by calculating with-break implementations of the Johansen (1995) tests over all possible break dates and basing a test on the most positive of these.

Relative to the developments seen in the univariate case, significant drawbacks therefore still exist with the currently available co-integration rank tests which allow for a break in the deterministic trend. Firstly, the approach in Inoue (1999) is infeasible in practice because, like Zivot and Andrews (1992), it cannot allow a trend break to occur under the null. In practice the co-integration rank of a system of variables is established by the sequential procedure outlined in Johansen (1995). Here one first tests the null hypothesis that the co-integration rank, r say, is zero against the hypothesis that $r = n$, n being the dimension of the system. If this null is accepted the procedure stops. Otherwise one sequentially tests the null hypotheses that $r = 1, 2, ...$, against the alternative that $r = n$, until the null cannot be rejected. If the true rank is r^*, the test for $r = r^*$ in this procedure will not be size controlled even asymptotically when a break is present. Second, the tests considered in Johansen *et al.* (2000) and Trenkler *et al.* (2007) both assume that the trend break date is known to the practitioner. Moreover, these tests essentially assume that a trend break does indeed occur and, hence, would be expected to unnecessarily sacrifice a considerable degree of finite sample power when no break occurs. Indeed it should be noted that Lütkepohl *et al.* (2003) need to impose that a break does occur otherwise they run into the same problems outlined above for the Perron (1997) procedure where no break is present. Finally, all of these procedures take the autoregressive lag length as fixed and known.

The aim of this paper is to address these drawbacks with the existing tests in the literature. In order to focus attention on what we believe to be the empirically most relevant case, we follow Trenkler *et al.* (2007) and consider only the leading example of the trend break case, but allowing for the possibility of a simultaneous level break. We propose new testing procedures which in spirit generalise the approach taken by Carrion-i-Silvestre *et al.* (2009), Harris *et al.* (2009) and Kim and Perron (2009) to the setting of testing for co-integration rank. We consider two possible approaches depending on whether the deterministic component is included additively as in Trenkler *et al.* (2007) or directly into the co-integrated vector autoregressive [VAR] equation as in Johansen *et al.* (2000). In either case the first step in the procedure is based on the use of a consistent estimator of the break date. In the context of the component DGP of Trenkler *et al.* (2007), a multivariate generalisation of the first difference trend break estimator used in Harris *et al.* (2009) is proposed, along with a corresponding estimator obtained from the levels of the data, while for the Johansen *et al.* (2000) set-up a maximum likelihood estimator of the break date is used. Based on these break date estimators, for each of the two approaches an information-based method using a Schwarz-type criterion is then employed to select between the version of the model which includes a trend break (included at the relevant estimated break date) and that which does not. Each of the proposed procedures also employs a Schwarz-type criterion to select the autoregressive lag length. Conventional trace-type co-integration rank tests are then computed appropriate to the model selected by these Schwarz-type criteria.

For each of the proposed procedures we establish that: (i) the estimator of the break fraction is consistent for the true break fraction; (ii) the information-based methods based on this estimator consistently select between the with-break and without-break variants of the model, and (iii) the resulting trace tests can be validly compared to known break date critical values in trend break case and to the without break critical values in the no break case. A consequence of our results is that, at least in large samples, the information-based methods we propose allow us to correctly identify whether we need to allow for a trend break in the model or not. This then implies that where a break is not present we will not see the loss in efficiency that is incurred by including a redundant trend break regressor in the model, and at the same time where a trend break is present we will not see the potentially large impact on the size and power properties of the rank tests that result from omitting the trend break. We present Monte Carlo simulation evidence which suggests that the procedure based on the Johansen *et al.* (2000) set-up is preferred and generally works very well even for a relatively small sample size such that the finite sample performance of this procedure is quite close to that seen for the benchmark rank tests which would obtain with knowledge of whether a trend break was present or not. The key findings of our Monte Carlo simulation exercise are presented here, while a more detailed set of results can be found in the accompanying supplement, Harris *et al.* (2015).

The paper is organised as follows. Section 2 details our reference co-integrated VAR model. Section 3 outlines our new procedures for co-integration rank testing which allow for the possibility that series under test display a trend break at an unknown point in the sample. Section 4 analyses the large sample properties of these methods. Results from our Monte Carlo simulation study are reported in section 5. Section 6 concludes. All proofs are contained in the Appendix. In the following $\xrightarrow{d}$ and $\xrightarrow{p}$ are used to denote weak convergence and convergence in probability, respectively; $1_{(\cdot)}$ denotes the usual indicator function; $\lfloor\cdot\rfloor$ denotes the integer part of its argument; $x := y$ and $y =: x$ each indicate that x is defined by y; $x \vee y$ and $x \wedge y$ indicate the maximum and minimum, respectively, of x and y; I_k denotes the $k \times k$ identity matrix. The notation 0 is used generically in context to denote a $j \times k$ matrix of zeroes. If a is of full column rank $n < m$, then $a_\perp$ is an $m \times (m-n)$ full column rank matrix satisfying $a'_\perp a = 0$; for any square matrix, A, $|A|$ denotes its determinant, and $\mathrm{tr}(A)$ its trace.

2 The Trend Break Co-integrated VAR Model

Following Trenkler *et al.* (2007), we consider the n-dimensional time series process $y_t := (y_{1t}, ..., y_{nt})'$, $t = 1, ..., T$, generated according to the following DGP

$$y_t = \mu_{0,0} d_{0,t}(0) + \mu_{1,0} d_{1,t}(0) + \mu_{0,1} d_{0,t}(b) + \mu_{1,1} d_{1,t}(b) + u_t, \tag{1}$$

for which we have defined the step (or level break) dummy $d_{0,t}(b) := 1_{(t>b)}$, and then for any $k = 0, \pm 1, \pm 2, ...$, also defined $d_{k,t}(b) := \Delta^{-k} d_{0,t}(b)$, where $\Delta := (1 - L)$ denotes the usual first difference filter in the lag operator, L, such that $L^k y_t = y_{t-k}$. Then, as special cases of this generic definition, we have that $d_{1,t}(0) = 1$ and $d_{1,t}(0) = t$ in (1) are the usual constant and linear trend terms, while $d_{1,t}(b) = 0 \vee (t - b)$ is a trend break dummy. The parameter vectors $\mu_{i,j}$, $i, j = 0, 1$, in (1) are all $n \times 1$. The model in (1) is therefore generated as the sum of a constant, linear trend, level shift and change in the trend slope at time b, together with a stochastic component u_t which we specify below. As is standard, for the purposes of the large sample results which follow, we assume that the break date depends on the sample size such that the break occurs at a fixed fraction of the sample size; that is, we parameterise the breakpoint in terms of the break fraction λ where $0 < \lambda_L \le \lambda \le \lambda_U < 1$, by $b = \lfloor \lambda T \rfloor$. Notice therefore that b is constrained to lie in the set $B := [\lfloor T\lambda_L \rfloor, \lfloor T\lambda_U \rfloor]$. It can be seen that a trend break exists in y_t only if $\mu_{1,1} \neq 0$ in (1) (ie. where at least one element of the vector $\mu_{1,1}$ is non-zero); unlike the previous contributions to this literature outlined in section 1, we will not assume that $\mu_{1,1} \neq 0$. In this paper our focus is on trend breaks and so while we allow for a simultaneous level break through $d_{0,t}(b)$ we will not explicitly consider the case where a level break but no trend break occurs since tests appropriate for this setting have already been developed in Saikkonen and Lütkepohl (2000), Lütkepohl *et al.* (2003) and Johansen *et al.* (2000), *inter alia*; that is, we will assume that $\mu_{0,1} = 0$ when $\mu_{1,1} = 0$ in what follows. In contrast to Trenkler *et al.* (2007), we do not assume in what follows that the break fraction λ is known. Where the distinction is important we will distinguish between a generic possible break fraction (break point) and the true break fraction (break point) by using λ^* (b^*) for the latter.

The model in (1) is completed by specifying the usual pth order reduced rank VAR (or co-integrated VAR) indeterministic version of the model of Johansen (1995) for u_t; that is,

$$\Delta u_t = \alpha \beta' u_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta u_{t-j} + e_t, \quad t = 1, ..., T \tag{2}$$

where $u_t := (u_{1t}, ..., u_{nt})'$, $e_t := (e_{1t}, ..., e_{nt})'$, and where the initial values, $u_{1-p}, ..., u_0$, are taken to be fixed in the statistical analysis. The co-integration parameters α and β are $(n \times r)$-dimensional, while the parameters $\{\Gamma_i\}_{i=1,..,p-1}$ on the stationary lagged dependent variables are each $(n \times n)$-dimensional. The innovation process $\{e_t\}$ in (2) is taken to satisfy the following relatively weak globally stationary martingale difference assumption taken from Cavaliere *et al.* (2010); see also Davidson (1994,pp.454-455) for further discussion:

Assumption 1: *The innovations $\{e_t\}$ form a martingale difference sequence with respect to the filtration $\mathcal{F}_t$, where $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ for $t = ..., -1, 0, 1, 2, ...$, satisfying: (i) the global homoskedasticity condition: $\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}(e_t e_t' | \mathcal{F}_{t-1}) \xrightarrow{p} \Sigma$, where Σ is full-rank, and (ii) $\mathrm{E}\|e_t\|^4 \le K < \infty$.*

As is routine, we also impose the standard so-called '$I(1, r)$ conditions' of Johansen (1995) on the parameters of (2) in order to rule out, for example, explosive processes.

Assumption 2: *The following conditions hold on the parameters of* (2)*: (i) The autoregressive lag* *order p satisfies $1 \leq p < \infty$; (ii) $|(I_n - \sum_{j=1}^{p-1} \Gamma_j z^j)(1-z) - \alpha\beta' z| = 0$ implies $|z| > 1$ or $z = 1$, and (iii) $|\alpha'_\perp \Gamma \beta_\perp| \neq 0$, where $\Gamma := (I_n - \sum_{j=1}^{p-1} \Gamma_j)$.*

Under Assumption 2, u_t is integrated of order one ($I(1)$) with co-integration rank r, and the co-integrating relations $\beta' u_t - E\left(\beta' u_t\right)$ are stationary. Part (i) of Assumption 2 assumes that the lag length parameter p is finite, but crucially does not assume that it is known to the practitioner.

An alternative formulation of (1)-(2) is considered in Johansen *et al.* (2000). Specifically, and as demonstrated in Trenkler *et al.* (2007), multiplying (1) through by the lag polynomial $(I_n - \sum_{j=1}^{p-1} \Gamma_j L^j)\Delta - \alpha\beta' L$ and re-arranging yields the vector error correction mechanism [VECM] form

$$\begin{aligned} \Delta y_t &= \delta_{0,0} d_{0,t}(0) + \delta_{0,1} d_{0,t}(b) + \sum_{j=0}^{p-1} \delta_{-1,j} d_{-1,t}(b+j) \\ &\quad + \alpha\left(\beta' y_{t-1} + \delta'_{1,0} d_{1,t-1}(0) + \delta'_{1,1} d_{1,t-1}(b)\right) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + e_t \qquad (3) \end{aligned}$$

where $d_{-1,t}(b) := 1_{(t=b+1)}$ is an impulse dummy; cf. Equation (5) of Trenkler *et al.* (2007), where explicit formulae for the $\delta_{i,j}$, $i,j = 0,1$ and $\delta_{-1,j}$, $j = 0, ..., p-1$, coefficient vectors are provided. As Trenkler *et al.* (2007) note, the VECM representation in (3) is a re-parameterised form of equation (6) of Johansen *et al.* (2000) for the special case thereof of a single break in trend. Notice that the VECM form in (3) includes a (broken) linear trend but does so in such a way that its coefficients are restricted to exclude the possibility of a quadratic (broken) trend in y_t.

Regardless of whether we work with the components form in (1)-(2), as in Trenkler *et al.* (2007), or the VECM form in (3) as in Johansen *et al.* (2000), our interest in this paper is focussed on the problem of testing the usual null hypothesis that the co-integration rank is (less than or equal to) r, denoted $H(r)$, against $H(n)$, but crucially without assuming any prior knowledge of whether $\mu_{1,1} = 0$ or $\mu_{1,1} \neq 0$ in (1), and in the case where $\mu_{1,1} \neq 0$ without prior knowledge of the trend break location, λ. In the next section we outline our proposed procedures for achieving this. These procedures all employ a trend break fraction estimator in a first step and then use standard information-based methods to select between the model with trend break and the corresponding model without.

3 Co-integration Rank Test Procedures

As discussed in the previous section, we now explore general approaches suggested by the structures of each of (1)–(2) and (3).

The procedures which we will develop in relation to the VECM form in (3) are based on a Gaussian (quasi) likelihood approach, in which the break date estimation, selection between the with-break and without-break models, de-trending and co-integration rank testing are all done within the usual reduced rank regression framework of Johansen (1995) and Johansen *et al.* (2000). In contrast to Johansen *et al.* (2000) the (maximum) number of potential trend breaks is restricted to be one, but the presence of a trend break is not assumed and, where a trend break is present, the break date is treated as unknown. The break date estimation is obtained by using (quasi) maximum likelihood estimation [MLE] on (3) under $H(r)$. Given this breakpoint, an adaptation of the usual Schwarz information criterion[1] [SC] is then used to select between the model with a break and the model with break excluded (i.e. (3) with $\delta_{0,1} = 0$, $\delta_{1,1} = 0$), both estimated under $H(r)$. The usual trace test for $H(r)$ is then performed on the selected model, using critical values appropriate to the selected model. In what follows we will refer to this procedure as SC-VECM.

For the components form in (1)–(2) a possible breakpoint estimator to use is the least squares (minimum residual sum of squares) estimator for the location of a level break in the first differences of (1). From a likelihood perspective, this estimator treats u_t as a simple vector random walk (i.e. $p = 1$, $r = 0$) regardless of the actual values of p and r. This estimator can be viewed as a multivariate generalisation of the corresponding trend break estimator used in Harris *et al.* (2009) which is based on applying the univariate level break estimator proposed in Bai (1994) to the first differences of the data. In that sense, the multivariate trend break estimator we consider here is also a special case of the multivariate (level break) estimator considered in Qu and Perron (2007), but applied here to the first differences. A natural SC step following this approach is to choose between the with-break and without-break models in the simple random walk model. As in the SC-VECM procedure, the usual trace test follows this selection. This procedure will be referred to as SC-DIFF in what follows.

The final procedure we present, referred to as SC-VAR in what follows, carries out the breakpoint estimation and SC selection between the with-break and without-break models in an unrestricted VAR, i.e. with $r = n$. This then permits a full comparison of the reasonable models in which to carry out the break specification; that is, SC-VECM is constructed under $H(r)$, SC-DIFF under $H(0)$, and SC-VAR under $H(n)$.

Before we lay out these procedures in detail, a short discussion comparing these possible approaches would seem useful. There is no obvious reason to predict, *a priori*, why one should prefer one of these procedures over the other in practice and it is our intention to take an ambivalent stance and outline each. We will then compare their finite sample performance using Monte Carlo methods in section 5. And indeed these results suggest that no one of these procedures dominates all of the others in all situations, although our Monte Carlo results do suggest that the SC-VAR approach is the least efficacious of the three. The motivation for the SC-VECM and SC-VAR approaches is perhaps clearer and more natural in that they are based on the likelihood function throughout. The SC-DIFF approach is more *ad hoc* in nature, with the break date estimation and model selection procedure imposing $r = 0, p = 1$.

We now detail the SC-VECM approach in section 3.1, followed by the SC-DIFF and SC-VAR approaches in sections 3.2 and 3.3 respectively. In each of these three procedures we outline below

[1] Although our focus in this paper is on the use of SC-type information criterion, analogous procedures based on any consistent information criterion, such as the Hannan-Quinn [HQ] information criterion, would have the same asymptotic properties as we report for the procedures in this paper. Unreported simulations suggest that the SC-type procedures considered here display superior finite sample performance to corresponding procedures based on HQ. As in the discussion in Remark 2 below, the HQ-type procedures showed a tendency to retain a trend break too often when it was not present.

the autoregressive lag length, p, will be chosen from the set of candidate values $p \in \{1,...,\bar{p}\}$, where $\bar{p}$ denotes the maximum lag length considered by the practitioner. As is standard we assume in what follows that $\bar{p}$ is at least as large as the true autoregressive lag order, denoted p^*.

3.1 The SC-VECM Procedure

First define a generic reduced rank regression of the form

$$\begin{aligned} Z_0 &= Z_1\gamma\alpha' + X_p\Psi' + \mathcal{E} \\ &= (Y_1 : X_1)\begin{pmatrix} \beta \\ \delta_1 \end{pmatrix}\alpha' + (X_0 : Z_{\Delta,p})\begin{pmatrix} \delta_0' \\ \Gamma' \end{pmatrix} + \mathcal{E} \end{aligned}$$

where $\delta_1 := \left(\delta_{1,0}' : \delta_{1,1}'\right)'$, $\delta_0 := (\delta_{0,0} : \delta_{0,1})$,

$$\underset{(T-p)\times n}{Z_0} := \begin{pmatrix} \Delta y_{p+1}' \\ \vdots \\ \Delta y_T' \end{pmatrix}, \; Y_1 := \begin{pmatrix} y_p' \\ \vdots \\ y_{T-1}' \end{pmatrix}, \; Z_{\Delta,p} := \begin{pmatrix} \Delta y_p' & \dots & \Delta y_2' \\ \vdots & & \vdots \\ \Delta y_{T-1}' & \dots & \Delta y_{T-(p-1)}' \end{pmatrix}, \; \mathcal{E} := \begin{pmatrix} \varepsilon_{p+1}' \\ \vdots \\ \varepsilon_T' \end{pmatrix} \tag{4}$$

and where X_0 and X_1 are each matrices of deterministic terms. The idea is that X_1 will contain the broken linear trend (if included) and the linear trend, while X_0 will contain the level shift (if included) and the constant term. The maximised quasi log-likelihood associated with (3), based on the additional assumption that e_t is Gaussian, is then given by the usual expression,

$$\begin{aligned} \hat{\ell}_T\left(r; X_0, X_1, p\right) &= -\frac{T}{2}\log\left|\frac{Z_0'\bar{P}_{X,p}Z_0}{T}\right| \\ &\quad -\frac{T}{2}\sum_{i=1}^{r}\log\left(1-\nu_i\left(\left(Z_0'\bar{P}_{X_p}Z_0\right)^{-1}Z_0'\bar{P}_{X_p}Z_1\left(Z_1'\bar{P}_{X_p}Z_1\right)^{-1}Z_1'\bar{P}_{X_p}Z_0\right)\right) \end{aligned} \tag{5}$$

where $\nu_i\left(M\right)$ denotes the i^{th} largest eigenvalues of the matrix M and $\bar{P}_X := I - X\left(X'X\right)^{-1}X'$ is the OLS orthogonal projection matrix on any X; see Chapter 6 of Johansen (1995) for a detailed discussion of the general approach.

For any possible break fraction $\lambda \in [\lambda_L, \lambda_U]$, define the $(T-p)\times 1$ vectors $\iota_\lambda := \left(d_{0,t}\left(\lfloor\lambda T\rfloor\right)\right)_{t=p+1}^{T}$ and $\tau_\lambda := \left(d_{1,t}\left(\lfloor\lambda T\rfloor\right)\right)_{t=p}^{T-1}$. The VECM in (3) with no trend break then has $X_1 := \tau_0$ and $X_0 := \iota_0$, while the VECM with a trend break has $X_1 := D_{1,\lambda} = (\tau_0 : \tau_\lambda)$ and $X_0 := D_{0,\lambda} = (\iota_0 : \iota_\lambda)$. The p impulse dummies included in (3) are asymptotically negligible but can be included in X_0 by defining $\varsigma_{\lambda,p} := \left(d_{-1,t}\left(\lfloor\lambda T\rfloor\right), d_{-1,t}\left(\lfloor\lambda T\rfloor+1\right), \dots,\; d_{-1,t}\left(\lfloor\lambda T\rfloor+p-1\right)\right)_{t=p+1}^{T}$ and redefining $D_{0,\lambda} := (\iota_0 : \iota_\lambda : \varsigma_{\lambda,p})$.

The SC-VECM procedure can then be described as follows.

SC-VECM Procedure:

Step 1. For each of $p = 1,...,\bar{p}$, define the MLE of the breakpoint under $H(r)$, viz.,

$$\hat{b}_{r,p} := \arg\max_{b\in B}\hat{\ell}_T\left(r; D_{0,b/T}, D_{1,b/T}; p\right). \tag{6}$$

The corresponding break fraction estimator is then defined as $\hat{\lambda}_{r,p} := \hat{b}_{r,p}/T$.

Step 2. Define the SC for the model including the trend break to be

$$SC_1(p; r, \lambda) := -2\hat{\ell}_T(r; D_{0,\lambda}, D_{1,\lambda}, p) + \left(n + r + 2 + n^2 p\right) \log T,$$

with selected lag length $\hat{p}_{1,r} := \arg\min_{p\in\{1,\ldots,\bar{p}\}} SC_1\left(p; n, \hat{\lambda}_{r,p}\right)$, where $\hat{\lambda}_{r,p}$ is the estimate of λ^* obtained in Step 1. Notice therefore that $\hat{p}_{1,r}$ is selected under $H(n)$.

Step 3. Define the SC for the model excluding the trend break to be

$$SC_0(p; r) := -2\hat{\ell}_T(r; \iota_0, \tau_0; p) + \left(n^2 p\right) \log T,$$

with selected lag length $\hat{p}_0 := \arg\min_{p\in\{1,\ldots,\bar{p}\}} SC_0(p; n)$. Again notice that $\hat{p}_0$ is selected under $H(n)$.

Step 4. Choose the model with trend break by setting: $\hat{p} = \hat{p}_{1,r}$ and $(X_0, X_1) = (D_{0,\hat{\lambda}_{r,\hat{p}}}, D_{1,\hat{\lambda}_{r,\hat{p}}})$ if

$$\text{SC-VECM} : SC_1\left(\hat{p}_{1,r}; r, \hat{\lambda}_{r,\hat{p}_{1,r}}\right) \leq SC_0(\hat{p}_0; r);$$

and setting $\hat{p} = \hat{p}_0$ and $(X_0, X_1) = (\iota_0, \tau_0)$ otherwise.

Step 5. The trace test statistic of $H(r)$ against $H(n)$ is then given by

$$q_T(X_0, X_1; \hat{p}) := 2\left(\hat{\ell}_T(n; X_0, X_1, \hat{p}) - \hat{\ell}_T(r; X_0, X_1, \hat{p})\right).$$

□

Remark 1. Observe that the SC in Step 4 of the SC-VECM procedure can be expressed in terms of the likelihood ratio decision rule to include the trend break if $2(\hat{\ell}_T(r; D_{0,\hat{\lambda}_{r,\hat{p}_{1,r}}}, D_{1,\hat{\lambda}_{r,\hat{p}_{1,r}}}, \hat{p}_{1,r}) - \hat{\ell}_T(r; \iota_0, \tau_0; \hat{p}_0)) \geq (n + r + 2) \log T$. This is analogous for testing for the presence of a trend break at the random fraction $\hat{\lambda}_{r,\hat{p}_{1,r}}$, and as such it is related to a sup-LR type statistic in the spirit of Andrews (1993), but where the decision rule is based not on a fixed critical value but on a Schwarz-type penalty. As such, Step 4 is then essentially a pre-test for the presence of a break which, by design, has size which shrinks to zero as the sample size diverges. The same requirement is needed on the trend break pre-tests used in the univariate testing analogue of the problem considered here in Harris *et al.* (2009) and Carrion-i-Silvestre *et al.* (2009).

Remark 2. The part of the SC-type penalty which corresponds to the trend break in the VECM is $(n + r + 2) \log T$. There are n parameters in $\delta_{0,1}$, r parameters in $\delta_{1,1}$ and the unknown breakpoint parameter is given a penalty of 2, the latter following from the theoretical results provided in Zhang and Siegmund (2007), Kurozumi and Tuvaandorj (2011) and Kim (2012). Consistent with the theoretical arguments provided by these authors, we found the choice of 2 for the breakpoint parameter in the penalty function gave better finite sample results than a penalty of 1, in that the latter did not appear to penalise the inclusion of the break sufficiently strongly, such that the trend break was retained too often when no break was in fact present, resulting in correspondingly lower power in that case; see the accompanying supplement, Harris *et al.* (2015).

Remark 3. In the SC-VECM procedure the lag length is selected for both the model including a break and the model excluding a break. Although the breakpoint estimation and break selection is done under $H(r)$ in SC-VECM, it is necessary to select p, in both the model including a break and the model excluding a break, under $H(n)$ (i.e. from the VAR in levels). It is well-known that failure to do so leads to power losses for the trace test; see Lütkepohl (2005) and Lütkepohl and Saikkonen (1999), *inter alia.* This will be done for all of the procedures outlined in this paper.

Remark 4. When a trend break is present in the DGP, the lag length estimator $\hat{p}_0$ may be inconsistent for the true lag length because it is based on a misspecified deterministic specification. Nevertheless, as shown in Theorem 1 below, the selection of the trend break in step 4 is consistent, implying that the resulting lag length estimator $\hat{p}$ is consistent whether or not a trend break is present in the DGP. An alternative approach would be to re-define the SC-VECM decision rule in step 4 as $SC_1\left(\hat{p}_{1,r}; r, \hat{\lambda}_{r,\hat{p}_{1,r}}\right) \leq SC_0\left(\hat{p}_{1,r}; r\right)$, so that only the lag length estimator $\hat{p}_{1,r}$ is used. The asymptotic results in Theorem 1 would be unchanged by this, but unreported simulations found that the SC-VECM procedure proposed above results in co-integration tests with superior finite sample properties. The same comments apply to the SC-DIFF and SC-VAR procedures subsequently outlined in sections 3.2 and 3.3 respectively.

Remark 5. If a sequence of tests of $H(r)$ is carried out for $r = 0, 1, \ldots$, the lag length in SC-VECM is re-selected for each test, noting that $\hat{\lambda}_{r,p}$ is recomputed for each value of r. Perron and Qu (2007) find, in a different context, that the re-selection of p for each r can produce improvements in finite sample properties. The incorporation of their modified selection criterion would also be possible in our context but is left for future research.

3.2 The SC-DIFF Procedure

The SC-DIFF procedure is motivated by the components form of the model (1). Taking first differences of (1) yields

$$\Delta y_t = \mu_{1,0} d_{0,t}(0) + \mu_{0,1} d_{-1,t}(b) + \mu_{1,1} d_{0,t}(b) + v_t, \quad t = 2, ..., T \tag{7}$$

where $v_t := \Delta u_t$. Observe that (7) coincides with the VECM form in (3) if $r = 0$ and $p = 1$, in which case $v_t = e_t$. More generally, v_t will be a stationary linear process disturbance. In matrix form (7) can be written as

$$\begin{aligned} Z_0^{(1)} &= \begin{pmatrix} \iota_0^{(1)} & \varsigma_{\lambda,1}^{(1)} & \iota_\lambda^{(1)} \end{pmatrix} \begin{pmatrix} \mu_{1,0}' \\ \mu_{0,1}' \\ \mu_{1,1}' \end{pmatrix} + V \\ &= D_\lambda^{(1)} \mu' + V. \end{aligned} \tag{8}$$

where $Z_0^{(1)} := (\Delta y_t')_{t=2}^T$, $\iota_\lambda^{(1)} := (d_{0,t}(\lfloor \lambda T \rfloor))_{t=2}^T$, $\varsigma_{\lambda,1}^{(1)} := (d_{-1,t}(\lfloor \lambda T \rfloor))_{t=2}^T$, and $V := (v_t')_{t=2}^T$. The (1) superscript denotes that these matrices contain observations for $t = 2, \ldots, T$, consistent with a model with $p = 1$, as opposed to the corresponding SC-VECM matrices that contain observations for $t = p + 1, \ldots, T$. The idea is that the breakpoint and then the presence or absence of the break can then both be decided in the context of (8).

The SC-DIFF procedure can then be described as follows.

SC-DIFF Procedure:

Step 1. Use $\hat{b}_{0,1}$ defined in (6) and the resulting $\hat{\lambda}_{0,1} := \hat{b}_{0,1}/T$; that is the breakpoint and break fraction estimates are obtained setting $r = 0$ and $p = 1$.

Step 2. Choose the model with trend break by setting: $(X_0, X_1) = (D_{0,\hat{\lambda}_{0,1}}, D_{1,\hat{\lambda}_{0,1}})$ if

$$\text{SC-DIFF} : SC_1\left(1; 0, \hat{\lambda}_{0,1}\right) \leq SC_0\left(1; 0\right),$$

where $SC_1(.;.,.)$ and $SC_0(.;.)$ are as defined in Steps 2 and 3, respectively, of SC-VECM; and setting $(X_0, X_1) = (\iota_0, \tau_0)$ otherwise.

Step 3. If the break is selected in Step 2, set

$$\hat{p} = \hat{p}_{1,0} := \arg\min_{p \in \{1,\ldots,\bar{p}\}} SC_1\left(p; n, \hat{\lambda}_{0,1}\right).$$

If the break is not selected in Step 2, set

$$\hat{p} := \arg\min_{p \in \{1,\ldots,\bar{p}\}} SC_0\left(p; n\right).$$

Step 4. The trace test statistic of $H(r)$ against $H(n)$ is then given by

$$q_T\left(X_0, X_1; \hat{p}\right) := 2\left(\hat{\ell}_T\left(n; X_0, X_1, \hat{p}\right) - \hat{\ell}_T\left(r; X_0, X_1, \hat{p}\right)\right).$$

$\square$

Remark 6. Notice that the estimator $\hat{b}_{0,1}$ is the value of b that minimises the generalised variance of the OLS residuals from (8); that is, $\hat{b}_0 = \arg\min_{b \in B}\left|\hat{\Sigma}_1\left(b/T\right)\right|$, where $\hat{\Sigma}_1(\lambda) := T^{-1}(Z_0^{(1)\prime}\bar{P}_{D_\lambda^{(1)}}Z_0^{(1)})$. This estimator can therefore be viewed as the multivariate extension of the trend break estimator discussed in Harris *et al.* (2009) which is based on applying the univariate level break estimator proposed in Bai (1994) to the first differences of the data.

Remark 7. The SC-DIFF approach imposes $r = 0$ and $p = 1$ for the breakpoint estimator and break selection steps. Although based on a misspecified model when either $r > 0$ or $p > 1$, we will demonstrate in section 4 that the SC-DIFF method is still able to consistently discriminate between the trend break and no trend break models in such cases.

Remark 8. Notice that, in contrast to the SC-VECM procedure of section 3.1, the SC-DIFF procedure uses only a single breakpoint estimator, $\hat{b}_{0,1}$, across all r and p. As a result, the lag length selection is the same for every r if a sequence of $H(r)$, $r = 0, 1, \ldots$, hypotheses are being tested in a sequential procedure.

3.3 The SC-VAR Procedure

The SC-VAR procedure is as follows.

SC-VAR Procedure:

Step 1. Use the breakpoint estimator $\hat{b}_{n,p}$, and corresponding break fraction estimator $\hat{\lambda}_{n,p} := \hat{b}_{n,p}/T$, obtained under $H(n)$.

Step 2. Select the lag length in the model with break as

$$\hat{p}_{1,n} := \arg \min_{p \in \{1,\ldots,\bar{p}\}} SC_1 \left(p; n, \hat{\lambda}_{n,p}\right).$$

Step 3. Select the lag length in the model without break as

$$\hat{p}_0 := \arg \min_{p \in \{1,\ldots,\bar{p}\}} SC_0 \left(p; n\right).$$

Step 4. Choose the model with trend break by setting: $\hat{p} = \hat{p}_{1,n}$ and $(X_0, X_1) = (D_{0,\hat{\lambda}_{n,\hat{p}}}, D_{1,\hat{\lambda}_{n,\hat{p}}})$ if

$$\text{SC-VAR} : SC_1 \left(\hat{p}_{1,n}; n, \hat{\lambda}_{n,\hat{p}_{1,n}}\right) \leq SC_0 \left(\hat{p}_0; n\right);$$

and setting $\hat{p} = \hat{p}_0$ and $(X_0, X_1) = (\iota_0, \tau_0)$ otherwise.

Step 5. The trace test statistic of $H(r)$ against $H(n)$ is then given by

$$q_T \left(X_0, X_1; \hat{p}\right) := 2 \left(\hat{\ell}_T \left(n; X_0, X_1, \hat{p}\right) - \hat{\ell}_T \left(r; X_0, X_1, \hat{p}\right)\right).$$

□

Remark 9. The SC-VAR decision criterion used in Step 4 can, like the SC-VECM criterion, also be expressed in terms of the VECM log-likelihoods defined in (5) with $r = n$ as

$$2 \left(\hat{\ell}_T \left(n; D_{0,\hat{\lambda}_{n,\hat{p}}}, D_{1,\hat{\lambda}_{n,\hat{p}}}\right) - \hat{\ell}_T \left(n; \iota_0, \tau_0\right)\right) \geq (n + r + 2) \log T$$

and so again has a likelihood ratio pre-test interpretation; cf. Remark 1. It can also be seen that the SC-VAR decision criterion carries out the trend break versus no trend break selection step under the alternative hypothesis, $H(n)$.

Remark 10. Notice that, in common with the SC-DIFF procedure but unlike the SC-VECM procedure, the selected lag length used in the SC-VAR procedure is the same for every r. This results from the fact that the breakpoint estimator in Step 1 is computed under $H(n)$.

Remark 11. Monte Carlo simulations reported in the accompanying supplement, Harris *et al.* (2015), reveal that the breakpoint estimator $\hat{b}_{n,p}$ used in Step 1 yields a procedure with quite poor finite sample properties. We found that substituting this with the breakpoint estimator $\hat{b}_{0,1}$ from the SC-DIFF procedure led to considerable improvements in the finite sample properties of the SC-VAR procedure relative to using $\hat{b}_{n,p}$. This change has no impact on the large sample properties of the SC-VAR procedure.

4 Asymptotic Analysis

In this section we establish the large sample behaviour of the SC-VECM, SC-DIFF and SC-VAR procedures outlined in section 3. In particular we demonstrate that the break fraction estimators $\hat{\lambda}_{r,p}$ are consistent at rate $O_p(T^{-1})$ in the case where a trend break occurs. We then demonstrate that the associated SC-VECM, SC-DIFF and SC-VAR information-based selection criteria based on these estimators all consistently discriminate between the relevant with trend break and without trend break models. We then establish the limiting null distributions of the resulting trace test statistics from these procedures, highlighting where these coincide with distributions which are known and tabulated in the literature and tabulating selected asymptotic critical values otherwise.

Before we present our main theorem, we need first to define the following functional which will feature in the representations given for the limiting distributions of the trace statistics which obtain from the SC-VECM, SC-DIFF and SC-VAR procedures. To that end, define

$$Q_{n-r}(F,W) := \operatorname{tr}\left(\int_0^1 dW(s)\,F(s)'\,ds\left(\int_0^1 F(s)\,F(s)'\,ds\right)^{-1}\int_0^1 F(s)\,dW(s)'\right),$$

where $W(\cdot)$ is an $(n-r)$-dimensional standard Brownian motion and $F(\cdot)$ is a generic process derived from $W(\cdot)$, the details of which are given on a case-by-case basis in the following Theorem.

Theorem 1 *Let $\{y_t\}$ be generated according to DGP (1)-(2) under Assumptions 1 and 2. Let the true co-integrating rank be denoted r^*. Then:*

(a) If $\mu_{1,1} \neq 0$ in (1), so that a trend break occurs, then:

A1. The trend break fraction estimators $\hat{\lambda}_{r,p}$ are consistent for the true break fraction, λ^*, for any r and $p \leq \bar{p}$, and satisfy

$$\hat{\lambda}_{r,p} - \lambda^* = O_p\left(T^{-1}\right). \tag{9}$$

A2. For any $r \leq r^*$, each of the three SC criteria select the model with a trend break with probability converging to one as $T \to \infty$; viz.,

$$\begin{aligned} \text{SC-VECM} &: \Pr\left(SC_1\left(\hat{p}_{1,r}; r, \hat{\lambda}_{r,\hat{p}_{1,r}}\right) \leq SC_0\left(\hat{p}_0; r\right)\right) \to 1 \\ \text{SC-DIFF} &: \Pr\left(SC_1\left(1; 0, \hat{\lambda}_{0,1}\right) \leq SC_0\left(1; 0\right)\right) \to 1 \\ \text{SC-VAR} &: \Pr\left(SC_1\left(\hat{p}_{1,n}; n, \hat{\lambda}_{n,\hat{p}_{1,n}}\right) \leq SC_0\left(\hat{p}_0; n\right)\right) \to 1 \end{aligned}$$

A3. For any $r \leq r^*$, $\hat{\lambda}$ satisfying (9) and $\hat{p}_1 = \hat{p}_{1,r}, \hat{p}_{1,0}$ or $\hat{p}_{1,n}$, the trace tests constructed using the estimated breakpoint and lag length are asymptotically equivalent to the trace tests based on the true breakpoint and lag length; that is, $q_T\left(D_{0,\hat{\lambda}}, D_{1,\hat{\lambda}}; \hat{p}_1\right) - q_T\left(D_{0,\lambda^*}, D_{1,\lambda^*}; p^*\right) \xrightarrow{p} 0$.

A4. The trace test statistics at the true breakpoint have the asymptotic null distributions

$$q_T\left(D_{0,\lambda^*}, D_{1,\lambda^*}; p^*\right) \xrightarrow{d} Q_{n-r_0}\left(F_{1,\lambda^*}, W\right) \tag{10}$$

where F_{1,λ^*} coincides with the process F_u defined in Equation (3.3) of Johansen et al. (2000,p.223) setting their parameter q equal to 2.

(b) If $\mu_{1,1} = 0$ in (1), so that no break occurs, then:

B1. For any $r \leq r^*$ each of the three SC criteria select the model without trend break with probability converging to one as $T \to \infty$; viz.,

$$\begin{aligned} \text{SC-VECM} &: \Pr\left(SC_1\left(\hat{p}_{1,r}; r, \hat{\lambda}_{r,\hat{p}_{1,r}}\right) > SC_0\left(\hat{p}_0; r\right)\right) \to 1 \\ \text{SC-DIFF} &: \Pr\left(SC_1\left(1; 0, \hat{\lambda}_{0,1}\right) > SC_0\left(1; 0\right)\right) \to 1 \\ \text{SC-VAR} &: \Pr\left(SC_1\left(\hat{p}_{1,n}; n, \hat{\lambda}_{n,\hat{p}_{1,n}}\right) > SC_0\left(\hat{p}_0; n\right)\right) \to 1 \end{aligned}$$

B2. The asymptotic null distribution of the trace test statistics without trend break are given by

$$q_T\left(\iota_0, \tau_0; \hat{p}_0\right) \xrightarrow{d} Q_{n-r_0}\left(F_0, W\right) \tag{11}$$

where F_0 is given in equation (11.11) of Johansen (1995).

Some remarks are in order.

Remark 12. The result in part A1 of Theorem 1 demonstrates that the break fraction estimators $\hat{\lambda}_{r,p}$ are consistent for the true break fraction, λ^* at rate $O_p(T^{-1})$. This rate holds regardless of the true co-integrating rank, r^*. Moreover, it also holds regardless of the true autoregressive lag length, p^*, since correct specification of the lag length is not necessary for the consistent estimation of the break fraction.

Remark 13. The rate of consistency established for the break fraction estimator in part A1 is crucial to the results in A2 and A3 which together show that where a trend break is present the trace statistics based on these estimated trend break points are asymptotically equivalent under the null hypothesis to the corresponding trace tests based on the true (unknown) break point for each of the SC-VECM, SC-DIFF and SC-VAR procedures. The limiting null distribution of the trace statistic, $q_T\left(D_{0,\lambda^*}, D_{1,\lambda^*}; p^*\right)$, given in (10), coincides with the limiting distribution given in Theorem 3.1 of Johansen *et al.* (2000); critical values from this distribution can be obtained either from Table 1 below (calculated by direct simulation methods using 10,000 replications) or can be calculated from the response surface given in Table 4 of Johansen *et al.* (2000,p.229) setting their parameter $q = 2$.

Remark 14. Where no trend break is present, the results in B1 and B2 show that the SC-VECM, SC-DIFF and SC-VAR procedures all correctly select the no break model for sufficiently large samples. The resulting no break trace statistic $q_T\left(\iota_0, \tau_0; p^*\right)$ has the usual restricted linear trend limiting distribution given in equations (11.9) and (11.11) of Theorem 11.1 of Johansen (1995) and tabulated in Table 15.4 of Johansen (1995).

Remark 15. The results given in A4 and B2 hold when the null hypothesis $H\left(r^*\right)$ that the co-integration rank is r^* is true. These results therefore imply that the trace tests from each of the SC-VECM, SC-DIFF and SC-VAR procedures will all be asymptotically correctly sized (when using the asymptotic critical values discussed in each procedure) regardless of whether a trend break occurs.

Remark 16. As in Johansen (1995) and Johansen *et al.* (2000), under $H(r)$, the r largest eigenvalues included in (5), generically denoted $\hat{\nu}_1, \ldots, \hat{\nu}_r$ here, converge in probability to positive numbers[2], while $T\hat{\nu}_{r+1}, \ldots, T\hat{\nu}_p$ are of $O_p(1)$. This holds both for the no trend break case and for the trend break case when evaluated at the true break fraction, λ^*. As a consequence it is straightforward to show that the trace tests which result from the SC-VECM, SC-DIFF and SC-VAR procedures will be consistent at rate $O_p(T)$ when the true co-integration rank is such that $r^* > r$. This result holds regardless of whether a trend break is present in the data or not. This implies, therefore, that the usual sequential approach to determining the co-integration rank[3] outlined in Johansen (1995) can still be employed using the trace tests which obtain from either the SC-VECM, SC-DIFF or SC-VAR procedures. In particular, these sequential approaches will lead to the selection of the correct co-integrating rank with probability $(1-\xi)$ in large samples, again regardless of whether a trend break occurs or not.

5 Finite Sample Simulations

5.1 Simulation design

In this section we report on a Monte Carlo simulation exercise designed to assess the finite sample performance of the trace co-integration tests of the SC-VECM and SC-DIFF procedures. We adopt the following VAR(2) simulation DGP,

$$y_t = \begin{pmatrix} \underset{(n-r)\times 1}{y_t^{(1)}} \\ \underset{r\times 1}{y_t^{(0)}} \end{pmatrix} = \begin{pmatrix} \underset{(n-r)\times 1}{\mu_{0,1}^{(1)}} & \underset{(n-r)\times 1}{\mu_{1,1}^{(1)}} \\ \underset{r\times 1}{\mu_{0,1}^{(0)}} & \underset{r\times 1}{\mu_{1,1}^{(0)}} \end{pmatrix} \begin{pmatrix} d_{0,t}(b^*) \\ d_{1,t}(b^*) \end{pmatrix} + \begin{pmatrix} u_t^{(1)} \\ u_t^{(0)} \end{pmatrix} \tag{12}$$

where

$$\left(I_n - \begin{pmatrix} a_{1,1} I_{n-r} & 0 \\ 0 & a_{0,1} I_r \end{pmatrix} L \right) \left(I_n - \begin{pmatrix} a_2 I_{n-r} & 0 \\ 0 & a_2 I_r \end{pmatrix} L \right) \begin{pmatrix} u_t^{(1)} \\ u_t^{(0)} \end{pmatrix} = \begin{pmatrix} e_t^{(1)} \\ e_t^{(0)} \end{pmatrix} \tag{13}$$

where the superscript (1) denotes the $I(1)$ component under $H(r)$ and superscript (0) the $I(0)$ component. Here $|a_{0,1}| < 1$, $|a_2| < 1$, while $a_{1,1} = 1$ for $H(r)$ and $|a_{1,1}| < 1$ for $H(n)$. The disturbances are generated by $e_t^{(1)} \sim \text{i.i.d.}N(0, I_{n-r})$ and then, to allow for cross-correlation, we specify

$$e_t^{(0)} = \rho\kappa e_t^{(1)} + \sqrt{1-\rho^2}\varepsilon_t,\ \varepsilon_t \sim \text{i.i.d.}N(0, I_r)$$

where κ is an $r \times (n-r)$ matrix of ones. Here ρ controls the degree of cross-correlation (where relevant) between the $I(0)$ and $I(1)$ parts of the system. The deterministic specification we adopt sets $b^* = \lfloor \lambda^* T \rfloor$, for the set of trend break fractions $\lambda^* = 0.25, 0.50, 0.75$, and $\mu_{i,1}^{(j)} = c\iota$, $i, j = 0, 1$, where ι is a vector of ones and c is a scalar constant controlling the break magnitude. For simplicity,

[2]Explicit expressions for these eigenvalues are not required here, but are obtained in the proofs of part A2 of the theorem when a break is included, and can be found from the proofs of Theorem 11.1 of Johansen (1995) when the break is not included.

[3]This procedure starts with $r = 0$ and sequentially raises r by one until for $r = \hat{r}$ the trace test statistic does not exceed the ξ level critical value for the test.

this specification imposes the the same magnitudes for level and trend breaks, and in the $I(1)$ and $I(0)$ directions, with all breaks occurring at date b^*. The values $c = 0.8, 0.4, 0.2$ are used, along with $c = 0$, representing the case when no breaks of any kind occur.

The DGP in (12) corresponds directly to equation (1), while (13) is a special case of the VECM for u_t given in equation (2). With $a_{1,1} = 1$ the first $n - r$ components of u_t are $I(1)$ and the remaining r components are $I(0)$, implying r co-integrating vectors of the form $\beta = \left(0_{r\times(n-r)} : I_r\right)'$. The diagonal structure of (13) may appear restrictive but in fact is quite general because the DGP is invariant to taking orthogonal linear combinations of the columns of α and β. Moreover the statistical methods we describe are invariant to full rank linear combinations of the elements of y_t, and hence u_t, so that the appearance of a restrictive structure of r pure $I(0)$ variables and $n - r$ pure $I(1)$ variables is in fact quite general.

Tables 2 – 6 about here

Tables 2–5 give the empirical sizes and powers for the SC- procedures, based on our VAR(2) DGP for the case where the dimension of the system is $n = 2$ (additional results for the case of $n = 3$ can be found in the accompanying supplement, Harris *et al.*, 2015). We additionally include empirical sizes and powers for the VECM trace test which always includes the trend break with break fraction estimated under $H(r)$ (i.e. using $\hat{\lambda}_{r,\hat{p}_{1,r}}$ defined in SC-VECM), which we denote Break-VECM. The trace test which never includes a trend break (appropriate for $c = 0$) is also included and is simply denoted as VECM. Since the SC-VECM procedure selects between these two individual tests, they provide an informal benchmark for the performance of the SC- procedures. None of the tests assume a priori knowledge of p, but determine its value in the manner of section 3, assuming a maximum possible value of $\bar{p} = 4$. The simulation results are based on 10,000 Monte Carlo replications and we report results for the tests at the nominal (asymptotic) 0.05 level, for sample sizes of $T = 100$ and 200.

The tables of results given here are selected to illustrate the important features of the finite sample properties of the procedures. Nevertheless space constraints mean that results of the full experiment cannot be reported here, but they are made available in the accompanying supplement, Harris *et al.* (2015). Those results help to explain choices made in the reporting here. For example Tables 2-5 do not include results for SC-VAR because these tests were found to suffer from substantial size distortions when compared to SC-VECM and SC-DIFF, but results for SC-VAR are given in Harris *et al.* (2015). The supplement also provides finite sample evidence for the choice of 2 as the penalty for the break fraction parameter, rather than the usual 1.

5.2 Results for $r = 0$

Table 2 gives the results for the trace tests when testing the null hypothesis that $r = 0$ in the case where $p = 1$. Recall that the null (alternative) hypothesis is satisfied here when $a_{1,1} = 1$ ($a_{1,1} < 1$). The upper portion of Table 2 shows the results for $T = 100$. Starting with the Break-VECM benchmark test, we see that it has size that depends only modestly on λ^* and c, although it does appear slightly over-sized in general. It has power levels that increase with decreasing $a_{1,1}$, but which are also fairly insensitive to λ^* and c. Turning attention to the VECM benchmark test, we observe it being correctly sized for $c = 0$ and it is pertinent here that the VECM test is more powerful than the Break-VECM test. However, when $c > 0$, outside of the case $c = 0.2$, over-sizing becomes a very serious issue for the VECM test, to the extent that we cannot consider the rejection frequencies for $a_{1,1} < 1$ as representative of power in any meaningful sense. Examining SC-VECM when $c = 0$, it is immediately clear that it behaves very like the VECM test, both in terms of size and power. When $c > 0.2$ it clearly avoids the serious upward size distortion problems suffered by the VECM test, with size behaviour clearly rather more akin to that of the Break-VECM test. For $c = 0.8$, its power levels are very similar to those of the Break-VECM test. For $c = 0.4, 0.2$ its powers are similar to the rejection frequencies seen for the VECM test. A comparison of SC-VECM and SC-DIFF reveals very little difference between them.

The lower portion of Table 2 shows results for $T = 200$; here we employ larger values of $a_{1,1} < 1$ than for $T = 100$ in order to avoid too many high power entries. Here, the Break-VECM test is well size-controlled across all c; the modest over-sizing seen for $T = 100$ is no longer evident. However, the problems of significant over-sizing associated with the VECM test for $c > 0$ are even more readily apparent. SC-VECM is generally well sized controlled outside of $\lambda^* = 0.25$ and $c = 0.4$, where it appears slightly over-sized. It again inherits the power levels associated with the VECM test when $c = 0$ and those for the Break-VECM test when $c = 0.8$. For $c = 0.2$ its powers are once more similar to the rejection frequencies seen for the VECM test. SC-VECM and SC-DIFF again behave very similarly.

The behaviour of the SC-VECM tests depends both on that of the underlying Break-VECM and VECM tests (given in Table 2) and also the behaviour of the SC break selection criterion. In order to explicitly show how the break selection is working, Table 6 presents the empirical frequencies for which the SC selects a trend break, i.e. for which SC-VECM is set equal to Break-VECM.[4] The leftmost panels show the results for $r = 0$ and $p = 1$ corresponding to the results in Table 2. For $c \neq 0$ the correct decision is to include the break so the SC step is working best when the inclusions frequencies are close to one. Conversely, when $c = 0$ the correct decision is to omit the break, so inclusion frequencies near zero are better. The $r = 0$, $p = 1$ panels of Table 6 reveal the SC step in SC-VECM working close to perfectly for the largest break size ($c = 0.8$). Breaks of smaller magnitudes are more difficult to detect in this DGP, so as expected the inclusion frequencies are reduced as the break size is reduced through $c = 0.4$ and $c = 0.2$. When $c = 0$ the inclusion frequencies are close to zero as would be hoped. These findings hold generally for both $T = 100$ and $T = 200$, with the frequencies generally improved for $T = 200$, as expected.

The combination of the break inclusion frequencies in Table 6 with the size and power properties of the benchmark Break-VECM and VECM tests can often be used to attribute variations in the properties of the SC-VECM test. For example in Table 2 with $\lambda^* = 0.25$ and $c = 4$ the SC-VECM test

[4]Break inclusion frequencies can be computed for SC-DIFF as well, but we focus on SC-VECM here given its overall superior finite sample performance.

shows some moderate and surprising increases in size for both $T = 100$ and $T = 200$, and these can be seen to be the product of the interaction of the SC step with the size properties of the Break-VECM and VECM tests. In the presence of the trend break of size $c = 4$ at $\lambda^* = 0.25$, the VECM tests that ignore this break are predictably very over-sized (0.176 for $T = 100$ and 0.367 for $T = 200$). The SC step for SC-VECM correctly includes a break with frequency 0.266 ($T = 100$) and 0.469 ($T = 200$), implying that 73.4% ($T = 100$) and 53.1% ($T = 200$) of the time the SC-VECM procedure is using the incorrect and badly over-sized VECM test, producing the moderate over-sizing observed in Table 2. The explanation for why this over-sizing does not occur for $\lambda^* = 0.5$ or $\lambda^* = 0.75$ can be found in a similar way. For $\lambda^* = 0.5$ the over-sizing of the VECM test is very similar to $\lambda^* = 0.25$, but the SC criterion correctly includes the break more often (0.426 for $T = 100$ and 0.706 for $T = 200$), as may be expected for what is essentially a trend break pre-test, resulting in improved size properties for the SC-VECM test. The explanation for $\lambda^* = 0.75$ is the reverse, since the VECM test is less over-sized in this case than for $\lambda^* = 0.25$ or 0.5, so that even though the SC step reverts to its $\lambda^* = 0.25$ performance, the size distortions induced by carrying out the VECM test are reduced.

Throughout the tables it is possible to explain many variations in finite sample properties by similarly examining the interactions of the break selection and benchmark test properties. For example, the power of the SC-VECM test for $\lambda^* = 0.25$ and $c = 0.2$ appears to be unexpectedly large relative to the powers for $\lambda^* = 0.5$ or 0.75. This is due to the higher rejection frequencies for the VECM test for $\lambda^* = 0.25$ increasing the rejection frequencies for the SC-VECM test as well, while this does not occur for $\lambda^* = 0.5$ or 0.75. Similarly, the power of SC-VECM for $\lambda^* = 0.25$ and $c = 0.2$ appears to be unexpectedly large relative to the same break fraction with larger break sizes, especially for $T = 100$ when the SC-VECM power can even be slightly lower for larger break sizes. In this case it is due to variations in the SC break selection frequencies - the high power for $c = 0.2$ is actually mostly due to the SC step selecting the no break test which has higher power for $\lambda^* = 0.25$ at this point, while for $c = 0.4$ and 0.8 the SC step is selecting the correct test that includes a break. This is an unforeseen outcome in small samples that disappears as T increases, as is evident in the $T = 200$ results.

The presence of stationary autocorrelation in time series can make co-integration inference more difficult. Results for this situation are shown in Table 3 in which $r = 0$ but now $p = 2$ with $a_2 = 0.5$. Since estimation of the additional autoregressive components causes a significant reduction in power levels throughout, we consider some smaller values of $a_{1,1} < 1$ than the ones used in Table 2. For $T = 100$, across-the-board over-sizing is more of an issue for the Break-VECM test and also for the VECM test when $c = 0$ than in Table 2. This is a reflection of the well-known size issues that can arise for the trace test in the presence of stationary autocorrelation (see Cheung and Lai, 1993, and Johansen, 2002, among others) and also of the fact that it is now possible to under-specify the value of p. Not surprisingly, this also manifests itself in a slight general upward shift in the sizes of SC-VECM, although SC-VECM has noticeably good size properties relative to the Break-VECM and VECM tests on which it is based. Otherwise, we see the relationships between the four tests remain much as in Table 2. The second pair of panels in Table 6 reveals that this situation, with $r = 0$ and stationary second order autocorrelation present, is the most difficult in which to detect the break accurately. Relative to $a_2 = 0$ in the first panel, the correct break inclusion rate when $c \neq 0$ is lower and the incorrect break inclusion rate when $c = 0$ is higher. Nevertheless, even in this worst case, the SC-procedures provide co-integration tests with generally (albeit not uniformly) superior performance compared to the Break-VECM test that omits the break selection step.

5.3 Results for $r = 1$

Tables 4–5 give results for tests for $r = 1$. This case differs from $r = 0$ in at least two important ways. Under the null of $r = 1$ there is a mixture of $I(0)$ and $I(1)$ components in the model, which introduces additional nuisance parameters - the autocorrelation in the $I(0)$ component captured by $a_{0,1}$ in our data generating process, and the correlation between the $I(0)$ and $I(1)$ components captured by ρ. Also the presence of the $I(0)$ component can improve the properties of the break point estimation and SC break selection steps because inference on a trend break will be easier in $I(0)$ noise that it is in purely $I(1)$ noise. This will be demonstrated in the results that follow.

Table 4 looks at the case where $r = 1$, with $p = 1$, $a_{0,1} = 0$ and $\rho = 0$. The size of SC-VECM appears well-controlled everywhere, while its powers not only mirror those (superior) levels obtained from the VECM test when $c = 0$, but also those of the Break-VECM test for all $c > 0$ (not just $c = 0.8$). The correspondence is close for $T = 100$, and almost one-to-one for $T = 200$. This behaviour by SC-VECM is the clearest practical demonstration so far of how our procedures are intended to perform. To help explain this, the third panel of Table 6 provides the break selection frequencies for the SC-VECM procedure in this case, showing the SC step makes the correct selection in nearly every case for both $T = 100$ and $T = 200$. Relative to the frequencies in the first panel of Table 6 (in which $r = 0$), this also illustrates the advantage of the $I(0)$ component in detecting the trend break.

The results for SC-DIFF in Table 4 exhibit under-sizing for $c = 0.4$ and 0.2, something which appears to be inherited from the behaviour of the VECM test. In those cases where SC-DIFF is under-sized, corresponding powers are extremely low (as are those of the VECM test). This is the shortcoming of the SC-DIFF procedure - while it is computational convenient and performs well for $r = 0$, its imposition of $r = 0$ when this is not true has implications for the finite sample properties, despite its asymptotic validity. In this case the SC step of SC-DIFF omits a small to moderately sized break much too often, resulting in relatively poor properties. It is results of this nature that lead to our eventual suggestion that SC-VECM is the preferred procedure overall.

Table 5 gives results for $r = 1$, $p = 1$ and $a_{0,1} = 0.5$. Relative to $a_{0,1} = 0$ in Table 4, the properties of the benchmark Break-VECM and VECM tests are not dramatically affected by this autocorrelation. There is some deterioration in the ability of the SC step to detect a small break (see the fourth panel of Table 6), which translates into some power losses for the SC-VECM test relative to the Break-VECM test when the break size is small. The extent of this power loss depends considerably on the break fraction λ^*. When $\lambda^* = 0.25$ the power of SC-VECM remains good, which can be seen to be primarily due to the good properties of the VECM test, even though it is not the correct test in this case. By contrast the VECM test has very poor power properties for $\lambda^* = 0.5$ or 0.75, which adversely influences the resulting power properties of the SC-VECM test. These variations are especially marked for $T = 100$, with generally much better all round break selection, size and power properties evident for $T = 200$. This all illustrates again how the properties of the SC- tests are the result of the interaction of the SC step with the properties of the underlying Break-VECM and VECM tests.

Due to space constraints we omit results for the other nuisance parameters a_2 and ρ, but these are included in the supplement Harris *et al.* (2015). Briefly, the finite sample effect of $a_2 \neq 0$ when $r = 1$ is considerably less than it is when $r = 0$. The finite sample effect of $\rho \neq 0$ is also revealed to be relatively minor.

5.4 Summary

Drawing together all of our findings, what emerges is that while the SC-VECM and SC-DIFF tests behave similarly for $r = 0$, they can behave very differently for $r = 1$. Here, SC-DIFF can be prone to low size and very low power when $c > 0$. In contrast, SC-VECM is well size-controlled everywhere and frequently has the ability to secure close to the better levels of power available from the VECM and Break-VECM tests in the environments for which they are intended to operate. On this basis we recommend the SC-VECM procedure for practical use.

The ultimate properties of the SC-VECM procedures are subject to some variations according to various features of the data generating process. The presence of stationary autocorrelation introduces some size distortions and power losses into all co-integration tests (the Break-VECM and VECM tests here), and can also make the SC selection of the break more difficult. The size of the trend break generally affects the SC break selection in predictable ways, with larger breaks easier to detect. Variations in the break fraction λ^*, on the other hand, can produce unexpected effects on SC-VECM through its differing effects on the SC step and the benchmark Break-VECM and VECM tests - generally, as might be expected, a break in the middle of the sample is easiest for the SC to detect, while the rejection frequencies of the misspecified VECM test in particular can be considerably greater for earlier breaks than later ones, and the interactions of these effects produce variations in the performance of the SC-VECM procedure that may appear unexpected but turn out to be somewhat explicable in these terms.

6 Conclusions

We have focussed on the problem of testing for the co-integration rank in VAR processes of unknown lag order when a break in the deterministic trend component may be present at an unknown point in the sample. In order to simultaneously avoid the size and power problems which can result, even in large samples, from an un-modelled trend break and at the same time guard against the loss of finite sample efficiency which results from allowing for a trend break when no trend is present, we have outlined an approach based on the use of information criteria. These criteria are used to select the autoregressive lag length and to select between the trend break and no trend break models, using a consistent estimate of the break fraction in the former case. Two possible frameworks were considered depending on whether the deterministic component was included additively in a components representation or directly into the VAR equation, the latter referred to here as the SC-VECM procedure. In each case these procedures were shown to deliver asymptotically correctly sized and consistent tests of the co-integration rank regardless of whether a trend break is present in the data or not. By selecting the no break model when no trend break is present, these procedures were also shown to avoid the potentially large power losses associated with tests which assume that a trend break date is known to have occurred, when in fact no break is present. Monte Carlo simulation results were presented which suggest that the procedures generally performed well in practice with the SC-VECM procedure preferred overall.

We conclude with some suggestions for further research. First, we have focussed attention here on the case where a maximum of one break in the deterministic trend function can occur. In practice it might be useful to allow for the possibility that multiple trend breaks could exist. To do so, multiple break versions of the trend break estimators considered in section 3 would need to be developed; the estimators considered in Qu and Perron (2007) would seem to be a useful starting point for such an analysis. Combining such estimators with generalisations of the SC criteria in section 3, designed to choose between no trend break, one trend break, two trend breaks and so on, should then allow us to select the correct number of breaks in the limit, as is done in this paper for the no break against one break case. New tables of critical values would be needed for each number of breaks considered. Second, we have de-trended the data within the usual reduced rank regression framework. We chose to do this so as to produce a meaningful comparison across the procedures. It would also be possible to use pseudo-GLS de-trending in the context of tests from the components formulation (1)-(2) as in Saikkonen and Lütkepohl (2000), Lütkepohl *et al.* (2003) and Trenkler *et al.* (2007) and this might be expected to yield more powerful tests in both the trend break and no trend break environments. Finally, we have focussed here on "stochastic" rather than "deterministic" co-integration. In the former the deterministic trend is left unrestricted under co-integration, while for the latter the co-integrating vector also eliminates deterministic non-stationarity in the data. The latter case corresponds to imposing the restrictions that $\delta_{1,0} = 0$ and $\delta_{1,1} = 0$ in (3). An important and empirically relevant example which is therefore ruled out by these restrictions is one where the vector y_t contains some trend stationary (potentially around broken trends) time series. Versions of the co-integration rank tests proposed here which impose these restrictions could be used instead. This could potentially result in more powerful tests where those restrictions do in fact hold on (3), but would come at the expense of uncontrolled size where those restrictions did not hold. Indeed, for these reasons (and others) Perron and Campbell (1993, p.778) argue that stochastic co-integration is "... a more relevant concept of cointegration." Alternatively, one could develop a sequential procedure to jointly select the form of the deterministic component and the co-integrating rank, as is proposed for the linear trend case in Johansen (1992).

References

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821—56 (Corrigendum, 71, 395—7).

Bai, J. (1994). Least squares estimation of a shift in a linear process. *Journal of Time Series Analysis* 15, 453—472.

Carrion-i-Silvestre, J.L., D. Kim, and P. Perron (2009). GLS-based unit root tests with multiple structural breaks under both the null and the alternative hypotheses. *Econometric Theory* 25, 1754—1792.

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

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.

Davidson J. (1994). *Stochastic Limit Theory.* Oxford: Oxford University Press.

Harris, D., D.I. Harvey, S.J. Leybourne and A.M.R. Taylor (2009). Testing for a unit root in the presence of a possible break in trend. *Econometric Theory* 25, 1545—1588.

Harris, D., S.J. Leybourne and A.M.R. Taylor (2015). Supplement to 'Tests of the Co-integration Rank in VAR Models in the Presence of a Possible Break in Trend at an Unknown Point'.

Inoue, A. (1999). Tests of cointegrating rank with a trend-break, *Journal of Econometrics* 90, 215–237.

Johansen S. (1992). Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics* 54, 383—397.

Johansen, S. (2002). A sample correction of the test for cointegrating rank in the vector autoregressive model. *Econometrica* 70, 1929–1961.

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford: Oxford University Press.

Johansen, S., R. Mosconi and B. Nielsen (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal* 3, 216—49.

Kim, D. and P. Perron (2009). Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. *Journal of Econometrics* 148, 1–13.

Kim J-Y. (2012). Model selection in the presence of nonstationarity. *Journal of Econometrics* 169, 247-257.

Kurozumi, E. and P. Tuvaandorj (2011). Model selection criteria in multivariate models with multiple structural changes. *Journal of Econometrics* 164, 218–238.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Heidelberg: Springer-Verlag.

Lütkepohl, H. and P. Saikkonen (1999). Order selection in testing for the cointegrating rank of a VAR process, In: *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, R.F. Engle and H. White (eds.), Oxford: Oxford University Press, 168–199

Lütkepohl, H., P. Saikkonen and C. Trenkler (2004). Testing for the cointegrating rank of a VAR process with a structural shift at unknown time. *Econometrica*72, 647—62.

Perron, P. (1989). The Great Crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361—401.

Perron, P. (1997). Further evidence of breaking trend functions in macroeconomic variables. *Journal of Econometrics* 80, 355—385.

Perron, P. and J.Y. Campbell (1993). A note on Johansen's cointegration procedure when trends are present. *Empirical Economics* 18, 777–789.

Perron, P. and Z. Qu (2007). A modified information criterion for cointegration tests based on a VAR approximation. *Econometric Theory* 23, 638–685.

Perron, P. and X. Zhu (2005). Structural breaks with deterministic and stochastic trends. *Journal of Econometrics* 129, 65—119.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica* 75, 459—502.

Saikkonen, P. and H. Lütkepohl (2000). Testing for the cointegration rank of a VAR process with structural shifts. *Journal of Business & Economic Statistics* 18, 451—64.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* 6, 461–464.

Stock, J.H. and M.W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* 14, 11—30.

Stock, J.H. and M.W. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R.F. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pp. 1–44. Oxford University Press.

Stock, J. and M.W. Watson (2005). Implications of Dynamic Factor Analysis for VAR Models, *NBER Working paper* 11467.

Trenkler, C., P. Saikkonen and H. Lütkepohl (2007). Testing for the cointegrating rank of a VAR process with level shift and trend break. *Journal of Time Series Analysis* 29, 331–358.

Zhang and Siegmund (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* 63, 22–32.

Zivot, E. and D.W.K. Andrews (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics* 10, 251—270.

A Appendix

A.1 Preliminaries

For any X_0, X_1, the maximised log-likelihood and the trace co-integration test statistic are functions of the eigenvalues of the matrix

$$M_T\left(X_0, X_1\right) := \left(\frac{Z_0'\bar{P}_{X_0:Z_{\Delta,p}}Z_0}{T}\right)^{-1}\frac{Z_0'\bar{P}_{X_0:Z_{\Delta,p}}Z_1}{T}\left(\frac{Z_1'\bar{P}_{X_0:Z_{\Delta,p}}Z_1}{T}\right)^{-1}\frac{Z_1'\bar{P}_{X_0:Z_{\Delta,p}}Z_0}{T}. \tag{A.1}$$

Substitution of $Z_1 = (Y_1 : X_1)$ and working out the subsequent partitioned inverse gives

$$\begin{aligned} M_T\left(X_0, X_1\right) &= \left(\frac{Z_0'\bar{P}_{X_0:Z_{\Delta,p}}Z_0}{T}\right)^{-1}\left(\frac{Z_0'\bar{P}_{X_0:Z_{\Delta,p}}X_1}{T^2}\left(\frac{X_1'\bar{P}_{X_0:Z_{\Delta,p}}X_1}{T^3}\right)^{-1}\frac{X_1'\bar{P}_{X_0:Z_{\Delta,p}}Z_0}{T^2}\right. \\ &\quad \left. +\frac{Z_0'\bar{P}_{X_0:X_1:Z_{\Delta,p}}Y_1}{T}\left(\frac{Y_1'\bar{P}_{X_0:X_1:Z_{\Delta,p}}Y_1}{T}\right)^{-1}\frac{Y_1'\bar{P}_{X_0:X_1:Z_{\Delta,p}}Z_0}{T}\right). \end{aligned} \tag{A.2}$$

Consider a breakpoint estimator $\hat{b}$ for which $\hat{\lambda} = \hat{b}/T$ and $\hat{\lambda} - \lambda^* = O_p\left(T^{-1}\right)$. The existence of such estimators is argued in Theorem 1. We will demonstrate that the asymptotic behaviour of appropriately standardised components of $M_T\left(D_{0,\hat{\lambda}}, D_{1,\hat{\lambda}}\right)$ is the same as of $M_T\left(D_{0,\lambda^*}, D_{1,\lambda^*}\right)$, hence showing that statistics based on the likelihoods and likelihood ratios are asymptotically unaffected by the replacement of the unknown λ^* by an estimator $\hat{\lambda}$. The presence of the stationary lagged differences does not affect the substance of the derivations or results, so for simplicity we set $p = 1$ in what follows and consider

$$\begin{aligned} M_T\left(D_{0,\lambda}, D_{1,\lambda}\right) &= \left(\frac{Z_0'\bar{P}_{D_{0,\lambda}}Z_0}{T}\right)^{-1}\left(\frac{Z_0'\bar{P}_{D_{0,\lambda}}D_{1,\lambda}}{T^2}\left(\frac{D_{1,\lambda}'\bar{P}_{D_{0,\lambda}}D_{1,\lambda}}{T^3}\right)^{-1}\frac{D_{1,\lambda}\bar{P}_{D_{0,\lambda}}Z_0}{T^2}\right. \\ &\quad \left. +\frac{Z_0'\bar{P}_{D_\lambda}Y_1}{T}\left(\frac{Y_1'\bar{P}_{D_\lambda}Y_1}{T}\right)^{-1}\frac{Y_1'\bar{P}_{D_\lambda}Z_0}{T}\right), \end{aligned}$$

where $D_\lambda = (D_{0,\lambda} : D_{1,\lambda})$. It is not enough to simply show that $M_T\left(D_{0,\hat{\lambda}}, D_{1,\hat{\lambda}}\right) - M_T\left(D_{0,\lambda^*}, D_{1,\lambda^*}\right) \xrightarrow{p} 0$ since the trace test statistic depends on eigenvalues that are $O_p\left(T^{-1}\right)$. Therefore appropriately standardised components of $M\left(D_{0,\lambda^*}, D_{1,\lambda^*}\right)$ are considered, as set out in the following Lemma.

Lemma A.1 *For a break fraction estimator $\hat{\lambda}$ such that $\hat{\lambda}-\lambda^{*}=O_{p}\left(T^{-1}\right)$*

1. $T^{-1}\left(Z_0'\bar{P}_{D_{0,\hat{\lambda}}}Z_0 - Z_0'\bar{P}_{D_{0,\lambda^*}}Z_0\right) \xrightarrow{p} 0$

2. $T^{-2}\left(Z_0'\bar{P}_{D_{0,\hat{\lambda}}}D_{1,\hat{\lambda}} - Z_0'\bar{P}_{D_{0,\lambda^*}}D_{1,\lambda^*}\right) \xrightarrow{p} 0$

3. $T^{-3}\left(D_{1,\hat{\lambda}}\bar{P}_{D_{0,\hat{\lambda}}}D_{1,\hat{\lambda}} - D_{1,\lambda^*}'\bar{P}_{D_{0,\lambda^*}}D_{1,\lambda^*}\right) \xrightarrow{p} 0$

4. $T^{-1}\left(Z_0'\bar{P}_{D_{\hat{\lambda}}}Y_1\beta - Z_0'\bar{P}_{D_{\lambda^*}}Y_1\beta\right) \xrightarrow{p} 0,\ T^{-1}\left(\beta'Y_1'\bar{P}_{D_{\hat{\lambda}}}Y_1\beta - \beta'Y_1'\bar{P}_{D_{\lambda^*}}Y_1\beta\right) \xrightarrow{p} 0$

5. $T^{-1}\left(Z_0'\bar{P}_{D_{\hat{\lambda}}}Y_1\beta_\perp - Z_0'\bar{P}_{D_{\lambda^*}}Y_1\beta_\perp\right) \xrightarrow{p} 0,\ T^{-1}\left(\beta'Y_1'\bar{P}_{D_{\hat{\lambda}}}Y_1\beta_\perp - \beta'Y_1'\bar{P}_{D_{\lambda^*}}Y_1\beta_\perp\right) \xrightarrow{p} 0$

6. $T^{-2}\left(\beta_\perp'Y_1'\bar{P}_{D_{\hat{\lambda}}}Y_1\beta_\perp - \beta_\perp'Y_1'\bar{P}_{D_{\lambda^*}}Y_1\beta_\perp\right) \xrightarrow{p} 0$

Proof of Lemma A.1. The conclusions of the Lemma are to be expected in view of similar results in, for example, Corollary 1 of Qu and Perron (2007) for stationary multivariate regressions and Proposition 4 of Kim and Perron (2009) for univariate unit root regressions. Indicative details only are therefore provided of the proof to demonstrate the steps involved when both $I(1)$ and $I(0)$ components are involved.

In Part 1 of the Lemma, first suppose $r^*=0$ so that the DGP can be represented

$$Z_0 = D_{0,\lambda^*}\delta_0' + \mathcal{E}. \tag{A.3}$$

It follows that $\bar{P}_{D_{0,\lambda^*}}Z_0 = \bar{P}_{D_{0,\lambda^*}}\mathcal{E}$, $\bar{P}_{D_{0,\hat{\lambda}}}Z_0 = \bar{P}_{D_{0,\hat{\lambda}}}D_{0,\lambda^*}\delta_0' + \bar{P}_{D_{0,\hat{\lambda}}}\mathcal{E}$, and $T^{-1}(Z_0'\bar{P}_{D_{0,\lambda^*}}Z_0) = T^{-1}(\mathcal{E}'\bar{P}_{D_{0,\lambda^*}}\mathcal{E}) \xrightarrow{p} E\left(e_t e_t'\right)$, so the statistic is correctly standardised. Now

$$\frac{Z_0'\bar{P}_{D_{0,\hat{\lambda}}}Z_0}{T} = \frac{\mathcal{E}'\bar{P}_{D_{0,\hat{\lambda}}}\mathcal{E}}{T} + \frac{\mathcal{E}'\bar{P}_{D_{0,\hat{\lambda}}}D_{0,\lambda^*}}{T}\delta_0' + \delta_0\frac{D_{0,\lambda^*}\bar{P}_{D_{0,\hat{\lambda}}}\mathcal{E}}{T} + \delta_0\frac{D_{0,\lambda^*}\bar{P}_{D_{0,\hat{\lambda}}}D_{0,\lambda^*}}{T}\delta_0'$$

so that the difference between the statistics at the estimated and true breakpoints consists of

$$\begin{aligned}\frac{Z_0'\bar{P}_{D_{0,\hat{\lambda}}}Z_0}{T} - \frac{Z_0'\bar{P}_{D_{0,\lambda^*}}Z_0}{T} &= \frac{\mathcal{E}'\left(\bar{P}_{D_{0,\hat{\lambda}}} - \bar{P}_{D_{0,\lambda^*}}\right)\mathcal{E}}{T} + \frac{\mathcal{E}'\bar{P}_{D_{0,\hat{\lambda}}}D_{0,\lambda^*}}{T}\delta_0' + \delta_0\frac{D_{0,\lambda^*}\bar{P}_{D_{0,\hat{\lambda}}}\mathcal{E}}{T} \\ &\quad + \delta_0\frac{D_{0,\lambda^*}\bar{P}_{D_{0,\hat{\lambda}}}D_{0,\lambda^*}}{T}\delta_0'.\end{aligned} \tag{A.4}$$

each of which can in turn be represented

$$\frac{\mathcal{E}'\left(\bar{P}_{D_{0,\hat{\lambda}}} - \bar{P}_{D_{0,\lambda^*}}\right)\mathcal{E}}{T} = \frac{\mathcal{E}'D_{0,\lambda^*}}{T}\left(\frac{D_{0,\lambda^*}'D_{0,\lambda^*}}{T}\right)^{-1}\frac{D_{0,\lambda^*}'\mathcal{E}}{T} - \frac{\mathcal{E}'D_{0,\hat{\lambda}}'}{T}\left(\frac{D_{0,\hat{\lambda}}'D_{0,\hat{\lambda}}}{T}\right)^{-1}\frac{D_{0,\hat{\lambda}}'\mathcal{E}}{T} \tag{A.5}$$

$$\frac{\mathcal{E}'\bar{P}_{D_{0,\hat{\lambda}}}D_{0,\lambda^*}}{T} = \frac{\mathcal{E}'D_{0,\lambda^*}}{T} - \frac{\mathcal{E}'D_{0,\hat{\lambda}}}{T}\left(\frac{D_{0,\hat{\lambda}}'D_{0,\hat{\lambda}}}{T}\right)^{-1}\frac{D_{0,\hat{\lambda}}'D_{0,\lambda^*}}{T} \tag{A.6}$$

$$\frac{D_{0,\lambda^*}'\bar{P}_{D_{0,\hat{\lambda}}}D_{0,\lambda^*}}{T} = \frac{D_{0,\lambda^*}'D_{0,\lambda^*}}{T} - \frac{D_{0,\lambda^*}'D_{0,\hat{\lambda}}}{T}\left(\frac{D_{0,\hat{\lambda}}'D_{0,\hat{\lambda}}}{T}\right)^{-1}\frac{D_{0,\hat{\lambda}}'D_{0,\lambda^*}}{T}. \tag{A.7}$$

Using $D_{0,\lambda} = (\iota_0 : \iota_\lambda)$ gives

$$\frac{\left(D_{0,\hat{\lambda}} - D_{0,\lambda^*}\right)' D_{0,\lambda^*}}{T} = \frac{1}{T}\left(0 : \iota_{\hat{\lambda}} - \iota_{\lambda^*}\right)' (\iota_0 : \iota_{\lambda^*}) = \frac{1}{T}\begin{pmatrix} 0 & 0 \\ \lfloor \lambda^* T \rfloor - \left\lfloor \hat{\lambda} T \right\rfloor & \left(\lfloor \lambda^* T \rfloor - \left\lfloor \hat{\lambda} T \right\rfloor\right) \vee 0 \end{pmatrix} \xrightarrow{p} 0$$

since $\hat{\lambda} - \lambda^* = O_p\left(T^{-1}\right)$. Similarly

$$\frac{\left(D_{0,\hat{\lambda}} - D_{0,\lambda^*}\right)' \left(D_{0,\hat{\lambda}} - D_{0,\lambda^*}\right)}{T} = \frac{1}{T}\left(0 : \iota_{\hat{\lambda}} - \iota_{\lambda^*}\right)' \left(0 : \iota_{\hat{\lambda}} - \iota_{\lambda^*}\right) = \frac{1}{T}\begin{pmatrix} 0 & 0 \\ 0 & \left|\left\lfloor \hat{\lambda} T \right\rfloor - \lfloor \lambda^* T \rfloor\right| \end{pmatrix} \xrightarrow{p} 0.$$

Combining these gives $T^{-1}\left(D'_{0,\hat{\lambda}} D_{0,\hat{\lambda}}\right) - T^{-1}\left(D'_{0,\lambda^*} D_{0,\lambda^*}\right) \xrightarrow{p} 0$. Also

$$\frac{D'_{0,\hat{\lambda}}\mathcal{E}}{T} - \frac{D'_{0,\lambda^*}\mathcal{E}}{T} = T^{-1} \sum_{t=\left(\left\lfloor \hat{\lambda} T \right\rfloor \wedge \lfloor \lambda^* T \rfloor\right)+1}^{\left\lfloor \hat{\lambda} T \right\rfloor \vee \lfloor \lambda^* T \rfloor} e_t \xrightarrow{p} 0 \tag{A.8}$$

by Lemma A.1 of Qu and Perron (2007). Substituting these zero limits into (A.5)–(A.7) and then back into (A.4) shows part 1 under $r^* = 0$. If $r^* > 0$ then we use $Z_{1,\lambda^*}\gamma = Y_1\beta - D_{1,\lambda^*}\mu_1'\beta = D_{0,\lambda^*}(\mu_0'\beta) + U_1\beta$, to find the representation

$$Z_0 = D_{0,\lambda^*}\left(\mu_0'\beta + \delta_0'\right) + \left(U_1\beta\alpha' + \mathcal{E}\right), \tag{A.9}$$

which has the same form as (A.3) in the sense of containing the intercept and level shift and an $I(0)$ disturbance vector. The proof of $T^{-1}\left(Z_0'\bar{P}_{D_{0,\hat{\lambda}}} Z_0 - Z_0'\bar{P}_{D_{0,\lambda^*}} Z_0\right) \xrightarrow{p} 0$ therefore follows exactly the same arguments as when $r^* = 0$.

In Parts 2 and 3 of the lemma, those terms in $M(D_{0,\lambda}, D_{1,\lambda})$ involving $Z_0'\bar{P}_{D_{0,\lambda}} D_{1,\lambda}$ and $D'_{1,\lambda}\bar{P}_{D_{0,\lambda}} D_{1,\lambda}$, follow by the same arguments as $Z_0'\bar{P}_{D_{0,\lambda}} Z_0$ in Part 1.

We now consider the terms in $M(D_{0,\lambda}, D_{1,\lambda})$ involving the (partially) $I(1)$ matrix Y_1, which has representation

$$Y_1 = D_{\lambda^*}\mu' + U_1, \tag{A.10}$$

where $U_1 := (u_{t-1})_{t=2}^T$ and where $\bar{P}_{D_{\lambda^*}} Y_1 = \bar{P}_{D_{\lambda^*}} U_1$ and $\bar{P}_{D_{\hat{\lambda}}} Y_1 = \bar{P}_{D_{\hat{\lambda}}} D_{\lambda^*}\mu' + \bar{P}_{D_{\hat{\lambda}}} U_1$. In particular consider Part 6, which is correctly standardised since $T^{-2}\beta_\perp' Y_1'\bar{P}_{D_{\lambda^*}} Y_1\beta_\perp = T^{-2}\beta_\perp' U_1'\bar{P}_{D_{\lambda^*}} U_1\beta_\perp = O_p(1)$. The difference between estimated and true breakpoints is

$$\begin{aligned} \frac{\beta_\perp' Y_1'\bar{P}_{D_{\hat{\lambda}}} Y_1\beta_\perp}{T^2} - \frac{\beta_\perp' Y_1'\bar{P}_{D_{\lambda^*}} Y_1\beta_\perp}{T^2} &= \frac{\beta_\perp' U_1'\bar{P}_{D_{\hat{\lambda}}} U_1\beta_\perp}{T^2} - \frac{\beta_\perp' U_1'\bar{P}_{D_{\lambda^*}} U_1\beta_\perp}{T^2} \\ &\quad + \frac{\beta_\perp' U_1'\bar{P}_{D_{\hat{\lambda}}} D_{\lambda^*}}{T^2}\mu' + \mu\frac{D'_{\lambda^*}\bar{P}_{D_{\hat{\lambda}}} U_1\beta_\perp}{T^2} + \mu\frac{D'_{\lambda^*}\bar{P}_{D_{\hat{\lambda}}} D_{\lambda^*}}{T^2}\mu'. \end{aligned}$$

The terms involving $\bar{P}_{D_{\hat{\lambda}}} D_{\lambda^*}$ were addressed in Part 1, so we focus on the $I(1)$ sum, $T^{-2}\left(\left(D_{\hat{\lambda}} - D_{\lambda^*}\right)' U_1\beta_\perp\right)$. Let w_{t-1} be any single $I(1)$ element of $U_1\beta_\perp$, so the corresponding non-zero elements of $T^{-2}\left(D_{\hat{\lambda}} - D_{\lambda^*}\right)' U_1\beta_\perp$ can be written

$$\begin{pmatrix} T^{-2}\sum_{t=2}^T \left(d_{0,t}\left(\hat{b}\right) - d_{0,t}\left(b^*\right)\right) w_{t-1} \\ T^{-2}\sum_{t=2}^T \left(d_{1,t}\left(\hat{b}\right) - d_{1,t}\left(b^*\right)\right) w_{t-1} \end{pmatrix},$$

where $b^* := \lfloor \lambda^* T \rfloor$. Clearly if the second term (involving a broken linear trend) can be shown to disappear then the first term will as well. Using $d_{1,t}\left(\hat{b}\right) - d_{1,t}\left(b^*\right) = \left(\left(t-\hat{b}\right) \vee 0\right) - \left(\left(t-b^*\right) \vee 0\right) = \left(b^* \wedge t\right) - \left(\hat{b} \wedge t\right)$ gives

$$T^{-2}\sum_{t=2}^{T}\left(d_{1,t}\left(\hat{b}\right) - d_{1,t}\left(b^*\right)\right)w_t = T^{-2}\left(\sum_{t=\left(\hat{b}\wedge b^*\right)+1}^{\left(\hat{b}\vee b^*\right)}\left(\left(b^* \wedge t\right) - \left(\hat{b} \wedge t\right)\right)w_t + \left(b^* - \hat{b}\right)T^{-2}\sum_{t=\left(\hat{b}\vee b^*\right)+1}^{T} w_t\right).$$

Then

$$\left|T^{-2}\sum_{t=\left(\hat{b}\wedge b^*\right)+1}^{\left(\hat{b}\vee b^*\right)}\left(\left(b^* \wedge t\right) - \left(\hat{b} \wedge t\right)\right)w_t\right| \leq \left|\hat{b} - b^*\right| T^{-2}\sum_{t=\left(\hat{b}\wedge b^*\right)+1}^{\left(\hat{b}\vee b^*\right)} |w_t|$$

so it follows that

$$\begin{aligned}\left|T^{-2}\sum_{t=2}^{T}\left(d_{1,t}\left(\hat{b}\right) - d_{1,t}\left(b^*\right)\right)w_t\right| &\leq \left|\hat{b} - b^*\right| T^{-2}\sum_{t=\left(\hat{b}\wedge b^*\right)+1}^{T} |w_t| \\ &\leq \left|\hat{b} - b^*\right| T^{-2}\sum_{t=1}^{T} |w_t| = O_p\left(T^{-1/2}\right).\end{aligned}$$

Parts 4 and 5 then follow using the same arguments.

A.2 Proof of Theorem 1

A.2.1 (a) Break is present in DGP

The consistency of lag order selection in VAR and VECM models based on the SC is well known. See for example Proposition 8.1 of Lutkepohl (2005) for a textbook presentation of the consistency of the SC for the selection of p, a result that also justifies other criteria such as the Hannan-Quinn criterion (although not the Akaike criterion). The novelty in our results lies in the treatment of the trend breaks rather than the lag length, so we abstract from the selection of the lag length by assuming it known in these proofs, with the understanding that $\Pr\left(\hat{p}_{1,j} = p^*\right) \to 1$ for a lag length $\hat{p}_{1,0}, \hat{p}_{1,r}, \hat{p}_{1,n}$ selected by SC in any of our three procedures, with true lag length p^* that satisfies $p^* \leq \bar{p}$. In the models that include a break when the break is present, the lag length in the following is therefore treated as fixed.

A1. The break fraction estimator $\hat{\lambda}_{0,1}$ is a special case of the multivariate regression break estimators of Qu and Perron (2007), and their Lemma 1 would therefore imply $\hat{\lambda}_{0,1} - \lambda^* = O_p\left(T^{-1}\right)$. However when $r^* > 0$, the imposition of $r = 0$ in the estimator imposes some over-differencing into at least some of the series in y_t, which would imply a violation of their Assumption A4(c) for some directions e. Inspection of their proofs shows that this condition is necessary for the derivation of the asymptotic distribution of $\hat{\lambda}_{0,1}$, but not its rate of convergence, which will continue to hold in the current case. This is the multivariate extension of the argument put forward in Lemma 1(ii) of Harris et al (2009). The estimator $\hat{\lambda}_{r,p}$ is equal to $\hat{\lambda}_{0,1}$ if $p = 1$ and $r = 0$. For other values of p and r, $\hat{\lambda}_{r,p}$ is based on the correctly specified likelihood under the null while $\hat{\lambda}_{0,1}$ is not, implying the consistency and rate of convergence properties for $\hat{\lambda}_{r,p}$ are no worse than those of $\hat{\lambda}_{0,1}$. If the null is false, so that some co-integrating vectors are omitted from the log-likelihood, some stationary autocorrelation remains in the disturbances of the model, but Qu and Perron demonstrate that this does not affect the rate of convergence of the estimator, only its asymptotic distribution. Also $\hat{\lambda}_{r,p}$ is a multivariate version of $\hat{\lambda}_2^{AO}$ in Kim and Perron (2009) which converges at rate $O_p\left(T^{-1}\right)$.

A2. Consider SC-VECM. The strategy of the proof is to show that the standardised log-likelihood for the model with break converges in probability to a value greater than does the standardised log-likelihood for the model without break. When a break is present in the DGP the lag length $\hat{p}_0$ selected in the model excluding a break is not consistent. (While beyond the scope of this paper, we conjecture that in this case $\Pr\left(\hat{p}_0 = \bar{p}\right) \to 1$.) In what follows we argue that $\Pr\left(SC_1\left(p; r, \lambda^*\right) < SC_0\left(p; r\right)\right) \to 1$ for each $p = 1, \ldots, \bar{p}$, so that the trend break can be selected consistently by the SC without knowledge of the correct lag length, and hence it follows that $\Pr\left(\min_p SC_1\left(p; r, \lambda^*\right) < \min_p SC_0\left(p; r\right)\right) = \Pr\left(SC_1\left(\hat{p}_{1,r}; r, \lambda^*\right) < SC_0\left(\hat{p}_0; r\right)\right) \to 1$. The trend break is therefore correctly selected with probability approaching one, while the potentially misspecified lag length $\hat{p}_0$ is not used with probability approaching one.

The maximised standardised log-likelihood when the break is included in the model is

$$\max_{\delta_0,\delta_1,\alpha,\beta,\Gamma} T^{-1}\ell_T\left(\beta, \delta_1, \alpha, \delta_0, \Gamma, r; D_{0,\lambda^*}, D_{1,\lambda^*}\right) = T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\lambda^*}\right), \left(\tau_0 : \tau_{\lambda^*}\right), p\right),$$

but when the break is excluded we instead have the constrained maximum

$$\max_{\alpha,\beta,\Gamma} T^{-1}\ell_T\left(\beta, 0, \alpha, 0, \Gamma, r; D_{0,\lambda^*}, D_{1,\lambda^*}\right) = T^{-1}\hat{\ell}_T\left(r; \iota_0, \tau_0, p\right).$$

As a constrained maximisation, this obviously satisfies the relation

$$T^{-1}\hat{\ell}_T\left(r; \iota_0, \tau_0, p\right) \leq T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\lambda^*}\right), \left(\tau_0 : \tau_{\lambda^*}\right), p\right). \tag{A.11}$$

We will show that $T^{-1}\hat{\ell}_T\left(r; \iota_0, \tau_0, p\right) \xrightarrow{p} \ell_0$ and $T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\lambda^*}\right), \left(\tau_0 : \tau_{\lambda^*}\right), p\right) \xrightarrow{p} \ell_1$ with $\ell_0 \neq \ell_1$, and hence $\ell_0 < \ell_1$ for any λ^* and p. The SC decision rule can be expressed as

$$\begin{aligned} &T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\hat{\lambda}_r}\right), \left(\tau_0 : \tau_{\hat{\lambda}_r}\right), p\right) - T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\lambda^*}\right), \left(\tau_0 : \tau_{\lambda^*}\right), p\right) \\ &\quad + T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\lambda^*}\right), \left(\tau_0 : \tau_{\lambda^*}\right), p\right) - T^{-1}\hat{\ell}_T\left(r; \iota_0, \tau_0, p\right) > \frac{1}{2}\left(n + r + 2\right)\frac{\log T}{T}. \end{aligned}$$

Then, for any $\varepsilon > 0$, the results of Lemma A.1 imply that

$$\Pr\left(\left|T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\hat{\lambda}_{r,p}}\right), \left(\tau_0 : \tau_{\hat{\lambda}_{r,p}}\right), p\right) - T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\lambda^*}\right), \left(\tau_0 : \tau_{\lambda^*}\right), p\right)\right| > \varepsilon\right) \to 0$$

while $\ell_0 < \ell_1$ implies that there exists some $M > 0$ such that

$$\Pr\left(T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\hat{\lambda}_{r,p}}\right), \left(\tau_0 : \tau_{\hat{\lambda}_{r,p}}\right), p\right) - T^{-1}\hat{\ell}_T\left(r; \iota_0, \tau_0, p\right) > M\right) \to 1.$$

Since $T^{-1}\log T < M$ for large enough T, we conclude that

$$\Pr\left(T^{-1}\hat{\ell}_T\left(r; \left(\iota_0 : \iota_{\hat{\lambda}_{r,p}}\right), \left(\tau_0 : \tau_{\hat{\lambda}_{r,p}}\right), p\right) - T^{-1}\hat{\ell}_T\left(r; \iota_0, \tau_0, p\right) > \frac{1}{2}\left(n + r + 2\right)\frac{\log T}{T}\right) \to 1,$$

which shows that $\Pr\left(SC_1\left(p;r,\hat{\lambda}_{r,p}\right) \leq SC_0\left(p;r\right)\right) \rightarrow 1$, as required for the consistency of SC-VECM as argued above. The corresponding result for SC-VAR follows similarly with r set to n.

We now outline how $T^{-1}\hat{\ell}_T\left(r;\iota_0,\tau_0,p\right) \xrightarrow{p} \ell_0$ and $T^{-1}\hat{\ell}_T\left(r;\left(\iota_0:\iota_{\lambda^*}\right),\left(\tau_0:\tau_{\lambda^*}\right),p\right) \xrightarrow{p} \ell_1$. The DGP can be written (excluding $Z_{\Delta,p}$ since nothing of substance changes in the following arguments that rely only on the different orders of magnitude of $I\left(1\right)$ and $I\left(0\right)$ components, not on whether a specific p produces white noise disturbances)

$$Y_1 = \left(\iota_0 : \tau_0 : \iota_{\lambda^*} : \tau_{\lambda^*}\right)\begin{pmatrix} \mu_{0,0}' \\ \mu_{0,1}' \\ \mu_{1,0}' \\ \mu_{1,1}' \end{pmatrix} + U_1 \tag{A.12}$$

implying in the co-integrating direction

$$Y_1\beta = \left(\iota_0 : \tau_0 : \iota_{\lambda^*} : \tau_{\lambda^*}\right)\begin{pmatrix} \mu_{0,0}'\beta \\ \mu_{0,1}'\beta \\ \mu_{1,0}'\beta \\ \mu_{1,1}'\beta \end{pmatrix} + U_1\beta$$

and that

$$Y_1\beta - \left(\tau_0 : \tau_{\lambda^*}\right)\begin{pmatrix} \mu_{0,1}'\beta \\ \mu_{1,1}'\beta \end{pmatrix} = \left(\iota_0 : \iota_{\lambda^*}\right)\begin{pmatrix} \mu_{0,0}'\beta \\ \mu_{1,0}'\beta \end{pmatrix} + U_1\beta$$

behaves like a stationary process with a level shift. Thus in the VECM representation we find

$$\begin{aligned} Z_0 &= \left(Y_1\beta - \left(\tau_0 : \tau_{\lambda^*}\right)\begin{pmatrix} \mu_{0,1}'\beta \\ \mu_{1,1}'\beta \end{pmatrix}\right)\alpha' + \left(\iota_0 : \iota_{\lambda^*}\right)\begin{pmatrix} \delta_{0,0}' \\ \delta_{1,0}' \end{pmatrix} + \mathcal{E} \\ &= \left(\left(\iota_0 : \iota_{\lambda^*}\right)\begin{pmatrix} \delta_{0,1} \\ \delta_{1,1} \end{pmatrix} + U_1\beta\right)\alpha' + \left(\iota_0 : \iota_{\lambda^*}\right)\begin{pmatrix} \delta_{0,0}' \\ \delta_{1,0}' \end{pmatrix} + \mathcal{E} \\ &= \left(\iota_0 : \iota_{\lambda^*}\right)\begin{pmatrix} \eta_{0,0}' \\ \eta_{1,0}' \end{pmatrix} + V \end{aligned} \tag{A.13}$$

where $\eta_{0,0}' = \delta_{0,1}\alpha' + \delta_{0,0}'$, $\eta_{1,0}' = \delta_{1,1}\alpha' + \delta_{1,0}'$, and $V = U_1\beta\alpha' + \mathcal{E}$ is mean zero and $I\left(0\right)$ because of the co-integration. Thus Z_0 is represented in terms of a level, a level shift and an $I\left(0\right)$ disturbance.

The maximised log-likelihood at r with break included is

$$\hat{\ell}_T\left(r;\left(\iota_0:\iota_{\lambda^*}\right),\left(\tau_0:\tau_{\lambda^*}\right)\right) = -\frac{T}{2}\log\left|\frac{Z_0'\bar{P}_{\iota_0:\iota_{\lambda^*}}Z_0}{T}\right| - \frac{T}{2}\sum_{i=1}^{r}\log\left(1-\nu_i\left(M_T\left(\left(\iota_0:\iota_{\lambda^*}\right),\left(\tau_0:\tau_{\lambda^*}\right)\right)\right)\right)$$

where

$$\begin{aligned} M_T\left(X_0,X_1\right) &= \left(\frac{Z_0'\bar{P}_{X_0}Z_0}{T}\right)^{-1}\frac{Z_0'\bar{P}_{X_0}\left(Y_1:X_1\right)}{T}\left(\frac{\left(Y_1:X_1\right)'\bar{P}_{X_0}\left(Y_1:X_1\right)}{T}\right)^{-1}\frac{\left(Y_1:X_1\right)'\bar{P}_{X_0}Z_0}{T} \\ &= \left(\frac{Z_0'\bar{P}_{X_0}Z_0}{T}\right)^{-1}\left(\frac{Z_0'\bar{P}_{X_0}X_1}{T^2}\left(\frac{X_1'\bar{P}_{X_0}X_1}{T^3}\right)^{-1}\frac{X_1'\bar{P}_{X_0}Z_0}{T^2}\right. \\ &\qquad \left. + \frac{Z_0'\bar{P}_{X_0:X_1}Y_1}{T}\left(\frac{Y_1'\bar{P}_{X_0:X_1}Y_1}{T}\right)^{-1}\frac{Y_1'\bar{P}_{X_0:X_1}Z_0}{T}\right). \end{aligned}$$

The representation (A.13) implies the leading term in $M_T(X_0, X_1)$ has limit

$$\frac{Z_0'\bar{P}_{X_0}Z_0}{T} = \frac{Z_0'\bar{P}_{\iota_0:\iota_{\lambda^*}}Z_0}{T} = \frac{V'\bar{P}_{\iota_0:\iota_{\lambda^*}}V}{T} \xrightarrow{p} E\left(v_t v_t'\right) := \Sigma_{00},$$

since V_t is $I(0)$. Also

$$\frac{X_1'\bar{P}_{X_0}Z_0}{T^2} = \frac{(\tau_0 : \tau_{\lambda^*})'\bar{P}_{\iota_0:\iota_{\lambda^*}}V}{T^2} = O_p\left(T^{-1/2}\right),$$

since V is a zero-mean $I(0)$ process, while standard polynomial summation results imply

$$\frac{X_1'\bar{P}_{X_0}X_1}{T^3} = \frac{(\tau_0 : \tau_{\lambda^*})'\bar{P}_{\iota_0:\iota_{\lambda^*}}(\tau_0 : \tau_{\lambda^*})}{T^3} \to Q,$$

where Q is a fixed full rank matrix, so that the first term in the second factor in $M_T(X_0, X_1)$ disappears asymptotically: $(T^{-2}Z_0'\bar{P}_{X_0}X_1)\left(T^{-3}X_1'\bar{P}_{X_0}X_1\right)^{-1}(T^{-2}X_1'\bar{P}_{X_0}Z_0) = O_p\left(T^{-1}\right)$. In the second term we consider the partially $I(1)$ process Y_1 in its stationary and non-stationary directions as usual:

$$\begin{aligned}
&\frac{Z_0'\bar{P}_{X_0:X_1}Y_1}{T}\left(\frac{Y_1'\bar{P}_{X_0:X_1}Y_1}{T}\right)^{-1}\frac{Y_1'\bar{P}_{X_0:X_1}Z_0}{T} \\
= \ &\frac{Z_0'\bar{P}_{X_0:X_1}Y_1\left(\beta : T^{-1/2}\beta_\perp\right)}{T}\left(\frac{\left(\beta : T^{-1/2}\beta_\perp\right)'Y_1'\bar{P}_{X_0:X_1}Y_1\left(\beta : T^{-1/2}\beta_\perp\right)}{T}\right)^{-1}\frac{\left(\beta : T^{-1/2}\beta_\perp\right)'Y_1'\bar{P}_{X_0:X_1}Z_0}{T}.
\end{aligned}$$

From (A.12) we have $\bar{P}_{X_0:X_1}Y_1 = \bar{P}_{\iota_0:\tau_0:\iota_{\lambda^*}:\tau_{\lambda^*}}Y_1 = \bar{P}_{\iota_0:\tau_0:\iota_{\lambda^*}:\tau_{\lambda^*}}U_1$, and hence that $\bar{P}_{X_0:X_1}Y_1\beta$ is a zero-mean $I(0)$ process while $\bar{P}_{X_0:X_1}Y_1\beta_\perp$ is a de-trended $I(1)$ process. Standard $I(1)$ / $I(0)$ limit theory therefore implies that

$$\begin{aligned}
\frac{Z_0'\bar{P}_{X_0:X_1}Y_1\beta}{T}, \ \frac{\beta' Y_1\bar{P}_{X_0:X_1}Y_1\beta}{T}, \ \frac{\beta_\perp' Y_1\bar{P}_{X_0:X_1}Y_1\beta_\perp}{T^2} &= O_p(1) \\
\frac{Z_0'\bar{P}_{X_0:X_1}Y_1\beta_\perp}{T^{3/2}}, \ \frac{\beta' Y_1\bar{P}_{X_0:X_1}Y_1\beta_\perp}{T^{3/2}} &= O_p\left(T^{-1/2}\right).
\end{aligned}$$

and hence that

$$\begin{aligned}
\frac{Z_0'\bar{P}_{X_0:X_1}Y_1}{T}\left(\frac{Y_1'\bar{P}_{X_0:X_1}Y_1}{T}\right)^{-1}\frac{Y_1'\bar{P}_{X_0:X_1}Z_0}{T} &= \frac{Z_0'\bar{P}_{X_0:X_1}Y_1\beta}{T}\left(\frac{\beta' Y_1'\bar{P}_{X_0:X_1}Y_1\beta}{T}\right)^{-1}\frac{\beta' Y_1'\bar{P}_{X_0:X_1}Z_0}{T} + o_p(1) \\
&\xrightarrow{p} \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0}.
\end{aligned}$$

Taken together, these results imply that, for any $r \leq r_0$,

$$\frac{1}{T}\hat{\ell}_T\left(r; (\iota_0 : \iota_{\lambda^*}), (\tau_0 : \tau_{\lambda^*})\right) \xrightarrow{p} -\frac{1}{2}\log|\Sigma_{00}| - \frac{1}{2}\sum_{i=1}^{r}\log\left(1 - \nu_i\left(\Sigma_{00}^{-1}\Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0}\right)\right) =: \ell_1. \quad \text{(A.14)}$$

Now consider the maximised log-likelihood with the break excluded

$$\hat{\ell}_T(r; \iota_0, \tau_0) = -\frac{T}{2}\log\left|\frac{Z_0'\bar{P}_{\iota_0}Z_0}{T}\right| - \frac{T}{2}\sum_{i=1}^{r}\log\left(1 - \nu_i\left(M_T(\iota_0, \tau_0)\right)\right).$$

The limits in this case are different because the breaks are not being regressed out. It is not necessary to derive the complicated expression for this limiting log-likelihood, it is sufficient to demonstrate that it differs from the case where the break is included. To begin, (A.13) implies that

$$\bar{P}_{\iota_0}Z_0 = \bar{P}_{\iota_0}\iota_{\lambda^*}\eta_{1,0}' + \bar{P}_{\iota_0}V, \quad \text{(A.15)}$$

and, hence,

$$\begin{aligned}\frac{Z_0'\bar{P}_{\iota_0}Z_0}{T} &= \eta_{1,0}\frac{\iota_{\lambda^*}'\bar{P}_{\iota_0}\iota_{\lambda^*}}{T}\eta_{1,0}' + \eta_{1,0}\frac{\iota_{\lambda^*}'\bar{P}_{\iota_0}V}{T} + \frac{V'\bar{P}_{\iota_0}\iota_{\lambda^*}}{T}\eta_{1,0}' + \frac{V'\bar{P}_{\iota_0}V}{T} \\ &\xrightarrow{p} \lambda^*\left(1-\lambda^*\right)\eta_{1,0}\eta_{1,0}' + \Sigma_{00} =: \Omega_{00}.\end{aligned}$$

Similarly, $T^{-2}(X_1'\bar{P}_{X_0}Z_0) = T^{-2}(\tau_0'\bar{P}_{\iota_0}\iota_{\lambda^*})\eta_{1,0}' + o_p(1) \xrightarrow{p} \Omega_{10}$, while $T^{-3}(X_1'\bar{P}_{X_0}X_1) = T^{-3}(\tau_0'\bar{P}_{\iota_0}\tau_0) \to \Omega_{11}$. The exact forms of $\Omega_{00}, \Omega_{10}, \Omega_{11}$ are not important, only that the both the orders and limits of these terms is different here from the case with a break included. In the expression in $M_T(X_0, X_1)$ involving Y_1, we have the residuals

$$\begin{aligned}\bar{P}_{X_0:X_1}Y_1 = \bar{P}_{\iota_0:\tau_0}Y_1 &= \bar{P}_{\iota_0:\tau_0}\left(\iota_{\lambda^*}:\tau_{\lambda^*}\right)\begin{pmatrix}\mu_{1,0}' \\ \mu_{1,1}'\end{pmatrix} + \bar{P}_{\iota_0:\tau_0}U_1 \\ &= \bar{P}_{\iota_0:\tau_0}\tau_{\lambda^*}\mu_{1,1}' + \bar{P}_{\iota_0:\tau_0}\iota_{\lambda^*}\mu_{1,0}' + \bar{P}_{\iota_0:\tau_0}U_1 \end{aligned} \tag{A.16}$$

which retain the broken trend τ_{λ^*}. Define $A_T := \left(T^{-1}A_\mu : T^{-1/2}A_{\beta_\perp} : A_\beta\right)$ to be an $n \times n$ full rank matrix of mutually orthogonal elements such that $A_\mu = \mu_{1,1}$, and the columns of A_β and $A_{\beta_\perp}$ are spanned by the columns of β and $\beta_\perp$ respectively. These latter two matrices can be obtained as orthogonal bases for the vector spaces spanned by the columns of $\bar{P}_{\mu_{1,1}}\beta$ and $\bar{P}_{\mu_{1,1}}\beta_\perp$. Then using (A.15) and (A.16) we find

$$\begin{aligned}\frac{Z_0'\bar{P}_{\iota_0:\tau_0}Y_1A_\mu}{T^2} &= \frac{\left(\bar{P}_{\iota_0}\iota_{\lambda^*}\eta_{1,0}' + \bar{P}_{\iota_0}V\right)'\bar{P}_{\iota_0:\tau_0}\tau_{\lambda^*}}{T^2}\mu_{1,1}'\mu_{1,1} + o_p(1) \\ &= \eta_{1,0}\frac{\iota_{\lambda^*}'\bar{P}_{\iota_0:\tau_0}\tau_{\lambda^*}}{T^2}\mu_{1,1}'\mu_{1,1} + o_p(1) \xrightarrow{p} \Omega_{0\mu},\end{aligned}$$

$$\frac{Z_0'\bar{P}_{\iota_0:\tau_0}Y_1A_{\beta_\perp}}{T^{3/2}} = \frac{\left(\bar{P}_{\iota_0}\iota_{\lambda^*}\eta_{1,0}' + \bar{P}_{\iota_0}V\right)'\bar{P}_{\iota_0:\tau_0}U_1A_{\beta_\perp}}{T^{3/2}} + o_p(1) \xrightarrow{d} \int_0^1 U_{1,\lambda^*}(s)\,ds,$$

$$\frac{Z_0'\bar{P}_{\iota_0:\tau_0}Y_1A_\beta}{T} = \frac{\left(\bar{P}_{\iota_0}\iota_{\lambda^*}\eta_{1,0}' + \bar{P}_{\iota_0}V\right)'\bar{P}_{\iota_0:\tau_0}U_1A_\beta}{T} + o_p(1) \xrightarrow{p} \Omega_{0\beta}.$$

Also

$$\begin{aligned}\frac{A_\mu'Y_1'\bar{P}_{\iota_0:\tau_0}Y_1A_\mu}{T^3} &= \left(\mu_{1,1}'\mu_{1,1}\right)^2\frac{\tau_{\lambda^*}'\bar{P}_{\iota_0:\tau_0}\tau_{\lambda^*}}{T^3} + o_p(1) \xrightarrow{p} \Omega_{\mu\mu} \\ \frac{A_{\beta_\perp}'Y_1'\bar{P}_{\iota_0:\tau_0}Y_1A_\mu}{T^{5/2}} &= \frac{A_{\beta_\perp}'U_1\bar{P}_{\iota_0:\tau_0}\tau_{\lambda^*}}{T^{5/2}}\mu_{1,1}'\mu_{1,1} + o_p(1) \xrightarrow{d} \int_0^1 U_{2,\lambda^*}(s)\,ds \\ \frac{A_\beta'Y_1'\bar{P}_{\iota_0:\tau_0}Y_1A_\mu}{T^2} &= \frac{\left(\bar{P}_{\iota_0:\tau_0}\iota_{\lambda^*}\mu_{1,0}'A_\beta + \bar{P}_{\iota_0:\tau_0}U_1A_\beta\right)'\bar{P}_{\iota_0:\tau_0}\tau_{\lambda^*}}{T^2} + o_p(1) \xrightarrow{p} \Omega_{\beta\mu} \\ \frac{A_{\beta_\perp}'Y_1'\bar{P}_{\iota_0:\tau_0}Y_1A_{\beta_\perp}}{T^2} &= \frac{A_{\beta_\perp}'U_1\bar{P}_{\iota_0:\tau_0}U_1A_{\beta_\perp}}{T^2} + o_p(1) \xrightarrow{d} \int_0^1 U_3(s)\,ds \\ \frac{A_{\beta_\perp}'Y_1'\bar{P}_{\iota_0:\tau_0}Y_1A_\beta}{T^{3/2}} &= \frac{A_{\beta_\perp}'U_1\bar{P}_{\iota_0:\tau_0}\left(\bar{P}_{\iota_0:\tau_0}\iota_{\lambda^*}\mu_{1,0}'A_\beta + \bar{P}_{\iota_0:\tau_0}U_1A_\beta\right)}{T^{3/2}} + o_p(1) \xrightarrow{d} \int_0^1 U_4(s)\,ds \\ \frac{A_\beta'Y_1'\bar{P}_{\iota_0:\tau_0}Y_1A_\beta}{T} &= \frac{\left(\bar{P}_{\iota_0:\tau_0}\iota_{\lambda^*}\mu_{1,0}'A_\beta + \bar{P}_{\iota_0:\tau_0}U_1A_\beta\right)'\bar{P}_{\iota_0:\tau_0}\left(\bar{P}_{\iota_0:\tau_0}\iota_{\lambda^*}\mu_{1,0}'A_\beta + \bar{P}_{\iota_0:\tau_0}U_1A_\beta\right)}{T} \xrightarrow{p} \Omega_{\beta\beta}.\end{aligned}$$

All of these limits together imply that

$$
\begin{aligned}
& \frac{Z_0'\bar{P}_{\iota_0:\tau_0}Y_1}{T}\left(\frac{Y_1'\bar{P}_{\iota_0:\tau_0}Y_1}{T}\right)^{-1}\frac{Y_1'\bar{P}_{\iota_0:\tau_0}Z_0}{T} \\
= & \frac{Z_0'\bar{P}_{\iota_0:\tau_0}Y_1\left(T^{-1}A_\mu : T^{-1/2}A_{\beta_\perp} : A_\beta\right)}{T} \\
& \times\left(\frac{\left(T^{-1}A_\mu : T^{-1/2}A_{\beta_\perp} : A_\beta\right)'Y_1'\bar{P}_{\iota_0:\tau_0}Y_1\left(T^{-1}A_\mu : T^{-1/2}A_{\beta_\perp} : A_\beta\right)}{T}\right)^{-1} \\
& \times\frac{\left(T^{-1}A_\mu : T^{-1/2}A_{\beta_\perp} : A_\beta\right)'Y_1'\bar{P}_{\iota_0:\tau_0}Z_0}{T} \\
& \xrightarrow{d}\left(\Omega_{0\mu} : \int_0^1 U_{1,\lambda^*}(s)\,ds : \Omega_{0\beta}\right)\begin{pmatrix} \Omega_{\mu\mu} & \int_0^1 U_{2,\lambda^*}(s)'\,ds & \Omega_{\beta\mu}' \\ \int_0^1 U_{2,\lambda^*}(s)\,ds & \int_0^1 U_3(s)\,ds & \int_0^1 U_4(s)\,ds \\ \Omega_{\beta\mu} & \int_0^1 U_4(s)'\,ds & \Omega_{\beta\beta} \end{pmatrix}^{-1} \\
& \times\left(\Omega_{0\mu} : \int_0^1 U_{1,\lambda^*}(s)\,ds : \Omega_{0\beta}\right)'.
\end{aligned}
$$

Thus for $r \leq r_0$

$$
\begin{aligned}
\frac{1}{T}\hat{\ell}_T(r;\iota_0,\tau_0) &= -\frac{1}{2}\log\left|\frac{Z_0'\bar{P}_{\iota_0}Z_0}{T}\right| - \frac{1}{2}\sum_{i=1}^{r}\log\left(1-\nu_i\left(M_T(\iota_0,\tau_0)\right)\right) \\
&\xrightarrow{p} -\frac{1}{2}\log|\Omega_{00}| - \frac{1}{2}\sum_{i=1}^{r}\log\left(1-\nu_i\left(\Omega_{00}^{-1}\Omega_{0\beta}\Omega_{\beta\beta}^{-1}\Omega_{\beta 0}\right)\right) =: \ell_0, \qquad \text{(A.17)}
\end{aligned}
$$

which is clearly a different limit from that given in (A.14).

The consistency for SC-DIFF is similar, but may involve some over-differencing since it is implicitly setting $r = 0$ and $p = 1$ even when these are not true. Some relevant results are therefore added here. The SC-DIFF criterion to include the break can be expressed as

$$
\log\left|\frac{Z_0'\bar{P}_{\iota_0}Z_0}{T}\right| - \log\left|\frac{Z_0'\bar{P}_{\iota_0:\iota_{\hat{\lambda}}}Z_0}{T}\right| \geq (n+2)\frac{\log T}{T}.
$$

For general r and p we have, as an extension of (A.9), $Z_0 = D_{0,\lambda^*}\left(\mu_0'\beta + \delta_0'\right) + V$, where $V := U_1\beta\alpha' + Z_{\Delta,p}\Gamma' + \mathcal{E}$ is $I(0)$. First observe that part 1 of Lemma A.1 applies here since all that is required is that (A.8) holds, which is true even when V is autocorrelated in an $I(0)$ manner. A standard law of large numbers then implies that $T^{-1}(Z_0'\bar{P}_{\iota_0:\iota_{\lambda^*}}Z_0) = T^{-1}(V'\bar{P}_{\iota_0:\iota_{\lambda^*}}V) \xrightarrow{p} E(v_tv_t')$. Regressing only on the constant gives $\bar{P}_{\iota_0}Z_0 = \bar{P}_{\iota_0}\iota_{\lambda^*}\psi' + \bar{P}_{\iota_0}V$, where $\psi' := \mu_{0,1}'\beta + \delta_{0,1}' \neq 0$ and, hence, $T^{-1}(Z_0'\bar{P}_{\iota_0}Z_0) = T^{-1}(V'\bar{P}_{\iota_0}V) + T^{-1}(\psi(\iota_{\lambda^*}'\bar{P}_{\iota_0}\iota_{\lambda^*})\psi') + o_p(1) = E(v_tv_t') + \lambda^*(1-\lambda^*)\psi\psi'$. By the same logic as the SC-VECM, it therefore follows that

$$
\Pr\left(\log\left|\frac{Z_0'\bar{P}_{\iota_0}Z_0}{T}\right| - \log\left|\frac{Z_0'\bar{P}_{\iota_0:\iota_{\lambda^*}}Z_0}{T}\right| \geq (n+2)\frac{\log T}{T}\right) \to 1.
$$

A3. The asymptotic equivalence of using estimated and true break fractions follows immediately from the results of Lemma A.1.

A4. The asymptotic null distribution of $q_T\left(D_{0,\lambda^*}, D_{1,\lambda^*}\right)$ is given in Theorem 3.1 of Johansen *et al.* (2000), with the extension from i.i.d. to martingale difference disturbances as specified in Assumption 1 following from the results of Cavaliere, Rahbek and Taylor (2010).

A.2.2 (b) Break is absent from DGP

When the break is absent from the DGP all of the lag length estimators are consistent, i.e. satisfy $\Pr(\hat{p} = p^*) \to 1$. We therefore treat the lag length as fixed here.

B1. The SC-VECM is based on the likelihood ratio process, $\hat{\ell}_T(r; (\iota_0 : \iota_\lambda), (\tau_0 : \tau_\lambda)) - \hat{\ell}_T(r; \iota_0, \tau_0) = O_p(1)$, uniformly for λ in the compact set $[\lambda_L, \lambda_U] \subset [0,1]$. Since this is the likelihood ratio version of a Chow test statistic with a fixed value of r under null and alternative, standard asymptotic results apply for any given λ to give, $2\left(\hat{\ell}_T(r; (\iota_0 : \iota_\lambda), (\tau_0 : \tau_\lambda)) - \hat{\ell}_T(r; \iota_0, \tau_0)\right) \xrightarrow{d} \chi^2_{n+r}$. We can therefore conclude that $\Pr\left(\sup_\lambda 2\left(\hat{\ell}_T(r; (\iota_0 : \iota_\lambda), (\tau_0 : \tau_\lambda)) - \hat{\ell}_T(r; \iota_0, \tau_0)\right) > \frac{1}{2}(n+r+2)\log T\right) \to 0$. The exact form of the asymptotic distribution of the process of these likelihood ratios over λ is not required for the consistency property of the SC.

B2. The distribution for $q_T(\iota_0, \tau_0)$ follows immediately from Theorem 1 and Remark 3.2 of Cavaliere, Rahbek and Taylor (2010).

Table 1. Asymptotic 5% critical values for $q_T\left(D_{0,\lambda^*}, D_{1,\lambda^*}\right)$

	$n-r$							
λ^*	1	2	3	4	5	6	7	8
0.20	17.45	34.51	55.51	80.56	109.82	142.98	180.18	221.87
0.25	18.03	35.53	56.88	82.15	111.52	145.02	182.14	224.08
0.30	18.46	36.25	57.98	83.31	112.95	146.24	183.46	225.18
0.35	18.75	36.92	58.63	84.09	113.67	147.08	184.29	225.82
0.40	18.95	37.26	59.26	84.79	114.21	147.48	184.78	226.47
0.45	19.07	37.56	59.56	84.97	114.58	147.83	184.97	226.47
0.50	19.09	37.65	59.62	85.09	114.77	147.88	185.07	226.73
0.55	19.05	37.59	59.54	84.96	114.69	147.83	185.10	226.78
0.60	18.93	37.39	59.14	84.62	114.30	147.42	184.84	226.44
0.65	18.84	36.90	58.62	84.02	113.76	146.75	184.35	225.87
0.70	18.46	36.27	57.93	83.30	112.82	146.06	183.50	224.94
0.75	17.99	35.45	56.82	82.03	111.53	144.86	182.27	223.93
0.80	17.49	34.48	55.49	80.54	109.81	142.99	180.35	221.68

Table 2. Finite sample size and power; estimated lag length; $n = 2$, $r = 0$, $p = 1$

λ^*	c		SC-VECM				SC-DIFF				Break-VECM				VECM			
										$T = 100$								
		$a_{1,1}$:	1.00	0.90	0.80	0.70	1.00	0.90	0.80	0.70	1.00	0.90	0.80	0.70	1.00	0.90	0.80	0.70
0.25	0.8		0.079	0.159	0.502	0.894	0.066	0.151	0.495	0.891	0.065	0.139	0.484	0.887	0.676	0.796	0.951	0.996
0.50	0.8		0.073	0.122	0.410	0.839	0.063	0.112	0.397	0.832	0.072	0.122	0.410	0.839	0.795	0.875	0.968	0.997
0.75	0.8		0.071	0.136	0.481	0.891	0.064	0.133	0.476	0.889	0.076	0.144	0.488	0.896	0.459	0.417	0.654	0.892
0.25	0.4		0.093	0.193	0.506	0.868	0.090	0.195	0.509	0.870	0.080	0.136	0.401	0.759	0.176	0.205	0.517	0.874
0.50	0.4		0.066	0.089	0.274	0.634	0.064	0.092	0.276	0.637	0.075	0.121	0.362	0.739	0.173	0.102	0.275	0.634
0.75	0.4		0.046	0.045	0.148	0.443	0.046	0.046	0.148	0.444	0.075	0.124	0.395	0.764	0.074	0.039	0.133	0.427
0.25	0.2		0.059	0.151	0.537	0.920	0.059	0.151	0.540	0.923	0.079	0.137	0.397	0.774	0.073	0.152	0.541	0.925
0.50	0.2		0.056	0.074	0.242	0.628	0.055	0.074	0.243	0.631	0.080	0.135	0.377	0.764	0.068	0.073	0.242	0.629
0.75	0.2		0.044	0.071	0.275	0.693	0.045	0.071	0.276	0.696	0.080	0.127	0.377	0.747	0.049	0.070	0.275	0.695
0.00	0.0		0.049	0.165	0.708	0.987	0.050	0.165	0.711	0.989	0.083	0.148	0.463	0.877	0.051	0.165	0.713	0.991
										$T = 200$								
		$a_{1,1}$:	1.00	0.94	0.88	0.82	1.00	0.94	0.88	0.82	1.00	0.94	0.88	0.82	1.00	0.94	0.88	0.82
0.25	0.8		0.049	0.169	0.671	0.977	0.048	0.167	0.669	0.976	0.048	0.169	0.671	0.977	0.974	1.000	1.000	1.000
0.50	0.8		0.050	0.136	0.571	0.951	0.047	0.133	0.568	0.950	0.050	0.136	0.571	0.951	0.997	1.000	1.000	1.000
0.75	0.8		0.055	0.165	0.658	0.976	0.052	0.161	0.656	0.976	0.056	0.165	0.658	0.976	0.922	0.993	1.000	1.000
0.25	0.4		0.114	0.343	0.784	0.977	0.110	0.343	0.784	0.977	0.056	0.164	0.596	0.920	0.367	0.457	0.839	0.988
0.50	0.4		0.066	0.168	0.580	0.918	0.063	0.165	0.577	0.914	0.055	0.132	0.523	0.892	0.420	0.396	0.754	0.967
0.75	0.4		0.049	0.101	0.388	0.791	0.048	0.101	0.389	0.790	0.056	0.154	0.593	0.927	0.172	0.104	0.346	0.768
0.25	0.2		0.076	0.186	0.617	0.952	0.076	0.186	0.617	0.952	0.060	0.148	0.490	0.840	0.099	0.187	0.618	0.953
0.50	0.2		0.052	0.070	0.289	0.732	0.052	0.070	0.289	0.733	0.057	0.132	0.465	0.835	0.083	0.069	0.288	0.732
0.75	0.2		0.040	0.051	0.238	0.679	0.040	0.051	0.238	0.679	0.062	0.132	0.472	0.818	0.051	0.050	0.236	0.679
0.00	0.0		0.046	0.235	0.875	0.999	0.046	0.235	0.876	0.999	0.061	0.167	0.617	0.969	0.047	0.235	0.877	0.999

Table 3. Finite sample size and power; estimated lag length; $n = 2$, $r = 0$, $p = 2$, $a_2 = 0.5$

λ^*	c		SC-VECM				SC-DIFF				Break-VECM				VECM			
									$T = 100$									
		$a_{1,1}$:	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40
0.25	0.8		0.098	0.332	0.652	0.750	0.090	0.311	0.662	0.730	0.117	0.288	0.626	0.733	0.186	0.472	0.860	0.981
0.50	0.8		0.091	0.226	0.560	0.660	0.094	0.254	0.591	0.633	0.115	0.254	0.571	0.662	0.164	0.320	0.808	0.984
0.75	0.8		0.076	0.206	0.510	0.665	0.090	0.284	0.637	0.715	0.114	0.281	0.625	0.740	0.083	0.127	0.436	0.728
0.25	0.4		0.083	0.324	0.539	0.672	0.083	0.340	0.564	0.678	0.121	0.274	0.543	0.624	0.104	0.340	0.562	0.682
0.50	0.4		0.082	0.155	0.309	0.362	0.088	0.162	0.329	0.378	0.121	0.254	0.518	0.595	0.093	0.140	0.301	0.363
0.75	0.4		0.073	0.141	0.268	0.222	0.081	0.141	0.272	0.237	0.118	0.254	0.525	0.636	0.074	0.119	0.247	0.216
0.25	0.2		0.075	0.373	0.626	0.748	0.081	0.387	0.645	0.751	0.121	0.294	0.597	0.652	0.082	0.388	0.645	0.753
0.50	0.2		0.077	0.248	0.412	0.366	0.087	0.249	0.417	0.371	0.124	0.287	0.590	0.634	0.080	0.246	0.415	0.367
0.75	0.2		0.077	0.281	0.487	0.431	0.084	0.285	0.497	0.437	0.121	0.282	0.572	0.627	0.078	0.284	0.496	0.435
0.00	0.0		0.078	0.414	0.733	0.916	0.080	0.432	0.752	0.918	0.120	0.308	0.669	0.747	0.078	0.433	0.753	0.921
									$T = 200$									
		$a_{1,1}$:	1.00	0.90	0.80	0.70	1.00	0.90	0.80	0.70	1.00	0.90	0.80	0.70	1.00	0.90	0.80	0.70
0.25	0.8		0.089	0.402	0.875	0.984	0.061	0.361	0.896	0.993	0.073	0.340	0.860	0.983	0.311	0.690	0.978	0.999
0.50	0.8		0.070	0.276	0.797	0.970	0.065	0.283	0.831	0.984	0.073	0.277	0.797	0.970	0.320	0.592	0.959	0.999
0.75	0.8		0.058	0.287	0.828	0.983	0.065	0.348	0.885	0.993	0.071	0.327	0.850	0.985	0.137	0.253	0.766	0.974
0.25	0.4		0.075	0.365	0.827	0.967	0.062	0.371	0.842	0.976	0.073	0.278	0.706	0.902	0.111	0.377	0.848	0.979
0.50	0.4		0.062	0.172	0.614	0.893	0.059	0.217	0.677	0.918	0.072	0.250	0.692	0.896	0.097	0.162	0.606	0.902
0.75	0.4		0.049	0.123	0.468	0.800	0.057	0.167	0.516	0.817	0.071	0.259	0.697	0.893	0.059	0.102	0.431	0.787
0.25	0.2		0.065	0.424	0.894	0.985	0.063	0.430	0.903	0.989	0.078	0.295	0.754	0.940	0.073	0.431	0.903	0.989
0.50	0.2		0.058	0.224	0.645	0.892	0.059	0.224	0.639	0.891	0.076	0.281	0.746	0.935	0.067	0.220	0.637	0.891
0.75	0.2		0.059	0.275	0.714	0.923	0.060	0.277	0.717	0.924	0.077	0.282	0.729	0.920	0.062	0.274	0.715	0.925
0.00	0.0		0.057	0.523	0.980	0.999	0.061	0.530	0.986	0.999	0.078	0.333	0.862	0.990	0.057	0.531	0.986	0.999

Table 4. Finite sample size and power; estimated lag length; $n = 2$, $r = 1$, $p = 1$, $a_{0,1} = 0.0$, $\rho = 0.0$

λ^*	c		SC-VECM				SC-DIFF				Break-VECM				VECM			
										$T = 100$								
		$a_{1,1}$:	1.00	0.85	0.70	0.65	1.00	0.85	0.70	0.65	1.00	0.85	0.70	0.65	1.00	0.85	0.70	0.65
0.25	0.8		0.051	0.212	0.746	0.881	0.083	0.300	0.743	0.846	0.051	0.212	0.746	0.881	0.386	0.749	0.945	0.955
0.50	0.8		0.062	0.182	0.689	0.844	0.064	0.173	0.637	0.775	0.062	0.182	0.689	0.844	0.396	0.710	0.891	0.898
0.75	0.8		0.060	0.204	0.731	0.867	0.057	0.194	0.635	0.739	0.060	0.204	0.731	0.868	0.198	0.259	0.351	0.361
0.25	0.4		0.052	0.201	0.716	0.844	0.096	0.211	0.367	0.402	0.052	0.203	0.729	0.868	0.141	0.214	0.364	0.399
0.50	0.4		0.061	0.177	0.666	0.814	0.038	0.048	0.090	0.101	0.061	0.178	0.672	0.827	0.079	0.042	0.068	0.073
0.75	0.4		0.060	0.194	0.699	0.832	0.011	0.007	0.020	0.024	0.060	0.195	0.711	0.854	0.019	0.003	0.006	0.008
0.25	0.2		0.050	0.186	0.659	0.801	0.058	0.215	0.606	0.710	0.053	0.199	0.694	0.837	0.066	0.216	0.606	0.711
0.50	0.2		0.061	0.170	0.626	0.756	0.020	0.018	0.021	0.022	0.063	0.174	0.655	0.806	0.026	0.017	0.015	0.014
0.75	0.2		0.052	0.180	0.648	0.766	0.026	0.065	0.152	0.186	0.057	0.191	0.695	0.836	0.028	0.064	0.148	0.182
0.00	0.0		0.046	0.311	0.925	0.979	0.047	0.313	0.929	0.983	0.073	0.199	0.699	0.843	0.048	0.314	0.931	0.985
										$T = 200$								
		$a_{1,1}$:	1.00	0.94	0.88	0.82	1.00	0.94	0.88	0.82	1.00	0.94	0.88	0.82	1.00	0.94	0.88	0.82
0.25	0.8		0.050	0.140	0.493	0.886	0.046	0.133	0.471	0.857	0.050	0.140	0.493	0.886	0.302	0.641	0.945	0.997
0.50	0.8		0.060	0.124	0.437	0.847	0.060	0.122	0.425	0.817	0.060	0.124	0.437	0.847	0.350	0.647	0.943	0.995
0.75	0.8		0.049	0.132	0.471	0.876	0.052	0.136	0.470	0.863	0.049	0.132	0.471	0.876	0.278	0.567	0.914	0.991
0.25	0.4		0.048	0.141	0.493	0.886	0.166	0.463	0.805	0.924	0.048	0.141	0.493	0.886	0.261	0.499	0.837	0.938
0.50	0.4		0.055	0.121	0.430	0.839	0.124	0.327	0.636	0.811	0.055	0.121	0.430	0.839	0.258	0.451	0.746	0.860
0.75	0.4		0.047	0.133	0.465	0.870	0.053	0.108	0.220	0.296	0.047	0.133	0.465	0.870	0.090	0.112	0.208	0.271
0.25	0.2		0.048	0.142	0.482	0.881	0.069	0.112	0.239	0.314	0.048	0.142	0.482	0.881	0.080	0.113	0.239	0.313
0.50	0.2		0.055	0.123	0.426	0.830	0.021	0.007	0.014	0.023	0.055	0.123	0.427	0.830	0.031	0.006	0.012	0.018
0.75	0.2		0.050	0.133	0.463	0.867	0.005	0.001	0.002	0.003	0.050	0.133	0.463	0.867	0.007	0.000	0.000	0.000
0.00	0.0		0.049	0.200	0.720	0.984	0.049	0.200	0.720	0.985	0.063	0.130	0.446	0.850	0.049	0.200	0.721	0.985

Table 5. Finite sample size and power; estimated lag length; $n = 2$, $r = 1$, $p = 1$, $a_{0,1} = 0.5$, $\rho = 0.0$

λ^*	c		SC-VECM				SC-DIFF				Break-VECM				VECM			
										$T = 100$								
		$a_{1,1}$:	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40	1.00	0.80	0.60	0.40
0.25	0.8		0.049	0.338	0.897	0.964	0.054	0.351	0.876	0.932	0.048	0.330	0.892	0.970	0.386	0.860	0.961	0.961
0.50	0.8		0.062	0.297	0.877	0.962	0.053	0.274	0.850	0.929	0.062	0.297	0.879	0.967	0.389	0.817	0.915	0.902
0.75	0.8		0.059	0.324	0.847	0.902	0.051	0.309	0.838	0.868	0.060	0.330	0.902	0.976	0.172	0.178	0.157	0.137
0.25	0.4		0.047	0.207	0.419	0.500	0.072	0.195	0.293	0.351	0.047	0.278	0.730	0.849	0.122	0.195	0.286	0.343
0.50	0.4		0.048	0.160	0.344	0.395	0.019	0.024	0.055	0.066	0.057	0.252	0.742	0.874	0.067	0.005	0.008	0.020
0.75	0.4		0.041	0.155	0.293	0.334	0.010	0.011	0.026	0.027	0.052	0.282	0.779	0.887	0.015	0.000	0.001	0.001
0.25	0.2		0.050	0.294	0.749	0.894	0.055	0.330	0.755	0.895	0.055	0.272	0.750	0.886	0.061	0.332	0.757	0.896
0.50	0.2		0.035	0.107	0.200	0.255	0.018	0.013	0.015	0.020	0.057	0.248	0.741	0.875	0.024	0.010	0.007	0.010
0.75	0.2		0.031	0.130	0.268	0.365	0.024	0.074	0.131	0.213	0.055	0.270	0.747	0.877	0.025	0.072	0.125	0.208
0.00	0.0		0.047	0.544	0.989	0.996	0.046	0.551	0.995	1.000	0.071	0.327	0.908	0.987	0.047	0.553	0.996	1.000
										$T = 200$								
		$a_{1,1}$:	1.00	0.92	0.84	0.76	1.00	0.92	0.84	0.76	1.00	0.92	0.84	0.76	1.00	0.92	0.84	0.76
0.25	0.8		0.051	0.239	0.790	0.993	0.045	0.225	0.768	0.980	0.051	0.239	0.790	0.993	0.511	0.945	1.000	1.000
0.50	0.8		0.057	0.201	0.739	0.986	0.056	0.193	0.718	0.971	0.057	0.201	0.739	0.986	0.550	0.939	0.999	1.000
0.75	0.8		0.050	0.221	0.774	0.993	0.050	0.223	0.763	0.983	0.050	0.221	0.774	0.993	0.446	0.886	0.995	1.000
0.25	0.4		0.048	0.229	0.770	0.946	0.124	0.389	0.577	0.634	0.048	0.230	0.777	0.982	0.289	0.436	0.581	0.620
0.50	0.4		0.057	0.197	0.711	0.949	0.043	0.130	0.295	0.363	0.057	0.198	0.714	0.978	0.231	0.170	0.145	0.126
0.75	0.4		0.050	0.216	0.748	0.909	0.022	0.035	0.089	0.101	0.050	0.216	0.759	0.985	0.062	0.004	0.004	0.006
0.25	0.2		0.045	0.207	0.679	0.803	0.069	0.167	0.331	0.466	0.048	0.221	0.748	0.962	0.082	0.167	0.328	0.464
0.50	0.2		0.056	0.184	0.642	0.743	0.019	0.001	0.006	0.007	0.057	0.190	0.694	0.957	0.032	0.000	0.000	0.000
0.75	0.2		0.049	0.197	0.671	0.712	0.011	0.005	0.003	0.005	0.051	0.209	0.741	0.965	0.015	0.003	0.000	0.000
0.00	0.0		0.049	0.360	0.951	1.000	0.049	0.360	0.951	1.000	0.062	0.211	0.743	0.987	0.050	0.361	0.952	1.000

Table 6. Break inclusion frequency for SC-VECM, $n = 2$

λ^*	c		$r=0, p=1$				$r=0, p=2, a_2=.5$				$r=1, p=1, a_{0,1}=0, \rho=0$				$r=1, p=1, a_{0,1}=.5, \rho=0$			
										$T=100$								
		$a_{1,1}$:	1.00	0.90	0.80	0.70	1.00	0.80	0.60	0.40	1.00	0.85	0.70	0.65	1.00	0.80	0.60	0.40
0.25	0.8		0.897	0.942	0.943	0.922	0.557	0.548	0.648	0.716	1.000	1.000	1.000	0.999	0.990	0.971	0.889	0.862
0.50	0.8		0.980	0.993	0.993	0.992	0.710	0.791	0.895	0.881	1.000	1.000	1.000	1.000	0.999	0.998	0.989	0.983
0.75	0.8		0.926	0.961	0.965	0.962	0.635	0.666	0.747	0.765	1.000	1.000	1.000	1.000	0.998	0.990	0.940	0.923
0.25	0.4		0.266	0.090	0.058	0.049	0.346	0.172	0.125	0.060	0.999	0.998	0.982	0.967	0.832	0.589	0.310	0.323
0.50	0.4		0.426	0.220	0.144	0.116	0.410	0.250	0.190	0.106	1.000	0.999	0.994	0.987	0.950	0.761	0.454	0.438
0.75	0.4		0.297	0.128	0.076	0.063	0.368	0.198	0.142	0.066	1.000	0.999	0.988	0.977	0.929	0.686	0.369	0.365
0.25	0.2		0.095	0.020	0.017	0.019	0.300	0.120	0.084	0.032	0.960	0.956	0.903	0.870	0.414	0.312	0.191	0.217
0.50	0.2		0.130	0.027	0.022	0.023	0.316	0.136	0.087	0.030	0.994	0.993	0.968	0.946	0.771	0.548	0.267	0.278
0.75	0.2		0.102	0.023	0.019	0.019	0.305	0.125	0.085	0.027	0.982	0.980	0.944	0.919	0.635	0.457	0.241	0.255
0.00	0.0		0.055	0.014	0.014	0.015	0.287	0.108	0.080	0.035	0.035	0.018	0.020	0.022	0.051	0.034	0.045	0.044
										$T=200$								
		$a_{1,1}$:	1.00	0.94	0.88	0.82	1.00	0.90	0.80	0.70	1.00	0.94	0.88	0.82	1.00	0.92	0.84	0.76
0.25	0.8		0.998	1.000	1.000	1.000	0.673	0.737	0.867	0.949	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.50	0.8		1.000	1.000	1.000	1.000	0.845	0.946	0.982	0.991	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.75	0.8		0.999	1.000	1.000	1.000	0.731	0.814	0.921	0.971	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.4		0.469	0.282	0.194	0.153	0.274	0.117	0.127	0.152	1.000	1.000	1.000	1.000	1.000	0.999	0.985	0.908
0.50	0.4		0.706	0.689	0.621	0.554	0.369	0.211	0.248	0.314	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.967
0.75	0.4		0.502	0.351	0.261	0.202	0.292	0.138	0.150	0.184	1.000	1.000	1.000	1.000	1.000	0.999	0.988	0.921
0.25	0.2		0.087	0.011	0.009	0.010	0.179	0.056	0.054	0.055	1.000	1.000	1.000	1.000	0.988	0.973	0.892	0.695
0.50	0.2		0.142	0.023	0.014	0.014	0.206	0.067	0.067	0.073	1.000	1.000	1.000	1.000	0.998	0.992	0.942	0.773
0.75	0.2		0.096	0.014	0.009	0.008	0.186	0.059	0.059	0.061	1.000	1.000	1.000	1.000	0.995	0.984	0.922	0.730
0.00	0.0		0.027	0.005	0.004	0.005	0.156	0.046	0.043	0.043	0.010	0.005	0.004	0.004	0.013	0.006	0.005	0.006