

Partial cointegrated vector autoregressive models with structural breaks in deterministic terms

Takamitsu Kurita* and Bent Nielsen†

22 October 2018

Abstract

This paper proposes a class of partial cointegrated models allowing for structural breaks in their deterministic terms. Details of the proposed models and their moving-average representations are examined. It is then shown that, under the assumption of martingale difference innovations, the limit distributions of partial quasi-likelihood ratio tests for cointegrating rank have a close connection to those for standard full models. This connection facilitates a response surface analysis which is required to extract critical information about moments from large-scale simulation studies. An empirical illustration of the proposed methodology is also provided. This paper renders partial cointegrated models more flexible and reliable devices for the study of non-stationary time series data with structural breaks.

JEL classification: C12, C32, C50.

Keywords: Partial cointegrated vector autoregressive models, Structural breaks, Deterministic terms, Weak exogeneity, Cointegrating rank, Response surface.

1 Introduction

Partial cointegration models with structural shifts in level or linear trends are quite common in practice; yet, no formal analysis is available for these models. The likelihood analysis of the partial models with such breaks is based on reduced rank regression, just like standard full cointegrated vector autoregressive models introduced by Johansen (1988, 1995). The main difference lies in the fact that likelihood-based tests for cointegrating rank in the partial models involve a set of new asymptotic distributions which reflect the combination of weakly exogenous regressors and broken deterministic terms. We generalise the standard assumption of normal innovations (see Johansen, 1995) to a flexible class of heterogeneous martingale difference innovations. We then derive the asymptotic distributions of the test statistics in question and also simulate their response surfaces that are applicable to a wide range of situations.

The presented models combine two widely used extensions of Johansen's original model. The first extension was a partial cointegrated system investigated by Harbo,

*Faculty of Economics, Fukuoka University, 8-19-1, Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan. E-mail: tkurita@fukuoka-u.ac.jp.

†Nuffield College, Oxford, OX1 1NF, UK; and Department of Economics and Program for Economic Modelling, University of Oxford. E-mail: bent.nielsen@nuffield.ox.ac.uk.

Johansen, Nielsen and Rahbek (1998), referred to as HJNR henceforth. This partial system is a conditional vector autoregressive model for a vector of variables, Y_t , given another vector of variables, Z_t , as well as lags of both variables. They also presented simulated tables for asymptotic rank test distributions based on the partial system. Boswijk (1995) and Ericsson and MacKinnon (2002) explored the use of conditional autoregressive models. Recently Cavaliere, De Angelis and Fanelli (2018) have explored how to turn these HJNR test statistics into information criteria. The second extension was a full cointegrated system allowing for structural breaks in a constant level or linear trend, a model explored by Johansen, Mosconi and Nielsen (2000), referred to as JMN hereafter. This full model is a multivariate extension of model C of Perron (1989), where both level and linear trend slope are changed at the time of the break, as opposed to his models A and B, in which only one of the two is changing. Deterministic breaks in cointegrated systems have also been explored by Inoue (1999) and Hendry and Massmann (2007).

Each of the above two extensions has proved to be useful in empirical applications; furthermore, subsequent practical work has shown that we frequently require both of the two extensions simultaneously. As an example, Bårdsen, Eitrheim, Jansen and Nymoen (2005) built a large scale model of the Norwegian economy by combining a number of smaller partial cointegration models. Each of these sub-systems is regarded as a partial model subject to structural shifts, and these types of models are useful in a practical sense for empirical macroeconomic research. As it stands, however, the exact asymptotic properties of likelihood-based test statistics derived from the partial models with structural breaks are unknown, so that a formal econometric study based on these models is unfeasible. This paper, therefore, conducts both analytical and simulation-based investigations into the unknown asymptotic properties, in order that researchers can perform a formal analysis using the partial models with structural breaks. Another example of these partial models is a trade model for the UK by Schreiber (2015), which we are going to use as an empirical illustration later in this paper.

This paper shows that the asymptotic distributions of the proposed likelihood-based test statistics are dependent on information about the dimension of the variables Y_t and Z_t , cointegrating rank, the number of breaks and their locations; but the distributions themselves are free of any unknown parameters. Hence, the limit distributions can be simulated given the above information, as in a manner similar to Johansen (1995, §15), HJNR or MacKinnon, Haug and Michelis (1999). The Granger-Johansen representation for the full model in JMN is also reexamined as a basis for the required asymptotic study, and this reexamination can be viewed as a useful clarification of roles of a set of starting values in the workings of the system. It should be noted that a condition for weak exogeneity reviewed in Section 2 is assumed to be satisfied when exploring the properties of the test statistics; the violation of this condition can give rise to a class of limit results unfavourable in applications, as discussed by Johansen (1992a). This assumption is testable by following an ex-post testing procedure suggested by Johansen (1992a) and others. We demonstrate this procedure in the empirical illustration in Section 4.

In deriving the asymptotic distributions of the test statistics, the assumption of normal innovations in Johansen (1995), HJNR and JMN is relaxed to the assumption of martingale difference innovations, with a view to widening the scope of applications of the proposed models. This means we have to be careful in developing asymptotic arguments required for the quasi-likelihood ratio test statistics. We use martingale limit results of Anderson and Kunitomo (1992) and Brown (1971) for approximately stationary

components and for non-stationary components, respectively.

Furthermore, it is shown that the derived asymptotic distributions can be approximated by gamma distributions, a class of common statistical distributions identifiable only by the first two moments; the validity of this gamma-distribution approximation method in various other existing models was documented by Nielsen (1997), Doornik (1998) and JMN. In demonstrating the applicability of this method to the proposed quasi-likelihood ratio tests, this study utilises the following property: the parameters of gamma distributions approximated to the limit distributions for the proposed partial models are expressible in terms of the corresponding parameters for full models and estimable covariance terms. As a result, it is feasible to apply the gamma approximation method to simulation results based on the *full* models, in order to obtain precise limit quantiles of the test statistics for the proposed *partial* models.

This approximation via the full models saves a great deal of computational time. It also contributes to a concise presentation of the simulation outcomes. Hence, we are justified in conducting comprehensive simulations in the full-model framework, the results of which are applied to response surface regression involved in the gamma approximation method. The outcomes of the regression analysis are tabulated in two tables, the accuracy of which is verified by moving back to the partial-model framework. The tables provide information of practical use, allowing researchers to conduct formal applied studies with the proposed partial models. A brief empirical study is also conducted to illustrate the advantage of having these tables available for applied research.

Overall, this paper makes a valuable addition to literature on time series econometrics and applied macroeconomics. As a result, the partial cointegrated models will be recognised as more flexible and practical devices for modelling and analysing non-stationary time series data containing structural breaks. Since the idea of JMN was extended to modelling time series data integrated of order 2, denoted as $I(2)$, by Kurita, Nielsen and Rahbek (2011), the same type of extension of the present paper is also conceivable in order to study partial $I(2)$ cointegrated models subject to deterministic breaks. This may be one of the important research directions to follow in the future.

The rest of this paper consists of four sections. Section 2 introduces partial cointegrated models subject to deterministic breaks and their moving-average representations. Section 3 derives partial quasi likelihood-based tests for cointegrating rank allowing for the breaks, and explores the limit distributions of the test statistics. In this section, a response surface analysis is performed by using simulated distributions and then the results of the analysis are summarised as a set of statistical tables. An empirical illustration of the proposed methodology is provided in Section 4. Finally, Section 5 gives concluding remarks. This study used *Ox* (Doornik, 2013) and *PcGive* (Doornik and Hendry, 2013) to conduct the simulations and the empirical study, respectively.

2 Models and representations

This section introduces partial cointegrated vector autoregressive models in the presence of deterministic breaks. Section 2.1 reviews the existing standard models known in the literature, while Sections 2.2 - 2.4 provide details of the proposed new models.

2.1 Previous models

The *cointegrated vector autoregressive model* was proposed by Johansen (1988, 1995). Suppose that we observe a p -variate vector time series X_t integrated of order 1, denoted as $I(1)$ hereafter. In the presence of 2 lags, a constant and restricted linear trend, the model equation for X_t , given cointegrating rank $r \leq p$, is

$$\Delta X_t = \alpha(\beta', \gamma) \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \Gamma \Delta X_{t-1} + \mu + \varepsilon_t \quad \text{for } t = 3, \dots, T, \quad (1)$$

where the initial values X_1 and X_2 are fixed while p -vector innovations $\varepsilon_3, \dots, \varepsilon_T$ are distributed as independent normal, denoted by $\mathbf{N}_p(0, \Omega)$. The parameters in equation (1) are all variation free, defined as $\alpha, \beta \in \mathbf{R}^{p \times r}$, $\gamma \in \mathbf{R}^r$, $\mu \in \mathbf{R}^p$ and $\Gamma, \Omega \in \mathbf{R}^{p \times p}$ with $\text{rank}(\alpha) = \text{rank}(\beta) = r$ and with Ω being positive definite. This model is interpreted in terms of its Granger-Johansen representation. The likelihood function is maximised through reduced rank regression of ΔX_t on the vector of $X_{t-1}, 1$ corrected for ΔX_{t-1} . The cointegrating rank r can be determined through a sequence of rank test statistics, which have Dickey-Fuller type limit distributions depending on the number of common trends, $p - r$ in this case, and on the specification of deterministic terms. Once the rank is determined, asymptotic inference for the cointegrating vectors β and the adjustment vectors α can be based on χ^2 distributions.

The *partial model* is derived from equation (1) to adjust to a situation where there are exogenous regressors that are not necessarily analysed in the model equation. With a view to setting up the partial model, let us introduce an integer m satisfying $0 \leq r \leq m < p$, so that we can decompose X_t into an m -vector Y_t and a vector Z_t of dimension $p - m$. Decompose the parameters of equation (1) accordingly, which is referred to as *the full model* henceforth, and also decompose its error sequence similarly to find

$$\alpha = \begin{pmatrix} \alpha_y \\ \alpha_z \end{pmatrix}, \Gamma = \begin{pmatrix} \Gamma_y \\ \Gamma_z \end{pmatrix}, \mu = \begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix}, \varepsilon_t = \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{pmatrix}.$$

We also define the population regression coefficient $\omega = \Omega_{yz} \Omega_{zz}^{-1}$, which leads to a class of conditional coefficients $\alpha_{y \cdot z} = \alpha_y - \omega \alpha_z$, $\Gamma_{y \cdot z} = \Gamma_y - \omega \Gamma_z$ and $\mu_{y \cdot z} = \mu_y - \omega \mu_z$. The partial or conditional model for Y_t given Z_t is then presented as

$$\Delta Y_t = \omega \Delta Z_t + \alpha_{y \cdot z}(\beta', \gamma) \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \Gamma_{y \cdot z} \Delta X_{t-1} + \mu_{y \cdot z} + \varepsilon_{y \cdot z, t}, \quad (2)$$

where the conditional innovation sequence $\varepsilon_{y \cdot z, t} = \varepsilon_{y, t} - \omega \varepsilon_{z, t}$ is $\mathbf{N}_m(0, \Omega_{yy \cdot z})$ distributed, so $\varepsilon_{y \cdot z, t}$ is independent of Z_t and the overall past series, while its variance is

$$\Omega_{yy \cdot z} = \Omega_{yy} - \Omega_{yz} \Omega_{zz}^{-1} \Omega_{zy}. \quad (3)$$

The marginal model for Z_t is simply given as

$$\Delta Z_t = \alpha_z(\beta', \gamma) \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \Gamma_z \Delta X_{t-1} + \mu_z + \varepsilon_{z, t}. \quad (4)$$

Due to the conditioning of Y_t on Z_t , the innovations $\varepsilon_{y \cdot z, t}$ and $\varepsilon_{z, t}$ are independent. Even so, the cointegrating relationships $\beta' X_{t-1} + \gamma t$ form cross equation restrictions, so that

maximum likelihood estimation needs to involve a joint analysis of (2) and (4). The rank can be determined from a partial analysis using information criteria albeit without size control as argued by Cavaliere, De Angelis and Fanelli (2018).

Weak exogeneity arises when $\alpha_z = 0$. In this case, the partial model and the marginal model are unrelated and Z_t is weakly exogenous for a class of parameters of interest, α_y , β and γ , in the sense of Engle, Hendry and Richard (1983). See also Johansen (1992 a,b; 1995, §8) and HJNR. Maximum likelihood estimation can be performed by analysing the two models separately, *i.e.* the partial model is estimated by reduced rank regression while the marginal model is by least squares regression. The maintained assumption is that the joint vector X_t has r cointegrating relations and hence $p - r$ common trends, with the cointegrating relations being in the partial model for Y_t . A notable feature of the setup is that it is left unspecified whether or not Z_t is cointegrated. In a one-lag model Z_t will not be cointegrated, but with further lags Z_t could be cointegrated since the short-run dynamics are determined by both α and Γ ; see HJNR (p. 390) for an example of these models. HJNR explored an asymptotic theory for likelihood-based rank testing in the partial model (2). The asymptotic distribution of HJNR's rank test statistic is of the Dickey-Fuller type, now depending on both $m - r$ and $p - r$, which are the dimensions of common trends for Y_t and X_t , respectively.

Structural breaks in deterministic terms were included in the full model by JMN. The idea is to consider, say, two sub-samples starting at time T_0 and T_1 , respectively, for $0 = T_0 < T_1 < T_2 = T$. The dynamic parameters in the model are the same for both sub-samples, while the parameters for deterministic terms can differ. In the model with lag-length $k = 2$ the observations $T_{j-1} + 1, T_{j-2} + 2$ for $j = 1, 2$ are held back as initial observations. Thus, the transition from one regime to the next is not modelled. A similar idea was recently used in a structural time series model by Harvey and Thiele (2017).

2.2 The partial model with structural breaks

We are in a position to introduce a new model, a partial cointegrated model allowing for structural breaks in its deterministic terms. Following JMN, we introduce a pre-specified number of sub-sample periods, q say, determined by the sub-sample structure $0 = T_0 < T_1 < \dots < T_q = T$. The dynamic parameters are common across the sub-sample periods, whereas the parameters for deterministic terms vary. For each period j and time point t under $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$, respectively, we extend the partial model for Y_t under $\alpha_z = 0$ to the one with $q - 1$ breaks and k lags:

$$\Delta Y_t = \omega \Delta Z_t + \alpha_y (\beta', \gamma_j) \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_{y \cdot z, i} \Delta X_{t-i} + \mu_{y \cdot z, j} + \varepsilon_{y \cdot z, t}, \quad (5)$$

where $\gamma_j \in \mathbf{R}^r$, $\mu_j = (\mu'_{y \cdot j}, \mu'_{z \cdot j})' \in \mathbf{R}^p$ and $\mu_{y \cdot z, j} = \mu_{y \cdot j} - \omega \mu_{z \cdot j}$ for $j = 1, \dots, q$, along with $\Gamma_i = (\Gamma'_{y \cdot i}, \Gamma'_{z \cdot i})' \in \mathbf{R}^{p \times p}$ and $\Gamma_{y \cdot z, i} = \Gamma_{y \cdot i} - \omega \Gamma_{z \cdot i}$ for $i = 1, \dots, k - 1$, and all the other parameters were defined in the previous sub-section. Note that the parameters for deterministic terms depend on j , indicating the presence of parameter shifts according to regime changes. A class of initial observations $X_{T_{j-1}+1}, \dots, X_{T_{j-1}+k}$ plays the dual role of capturing the transition from the previous regime, $j - 1$ and of serving as the initial observations for the regime j . In some applications the transition between the regimes may be longer than k observations, in which case more observations could be classified as

initial observations. The marginal model for Z_t under $\alpha_z = 0$ is

$$\Delta Z_t = \sum_{i=1}^{k-1} \Gamma_z \Delta X_{t-i} + \mu_{z,j} + \varepsilon_{z,t}. \quad (6)$$

As before, we can form a full model equation as in (1) for each sub-sample period. This yields the model of JMN with the condition of weak exogeneity $\alpha_z = 0$ implicitly placed on α . This model will be presented in the next sub-section.

The partial model can be formulated as a single equation for the full sample period in terms of the following notation. Following JMN, we define impulse dummy variables as

$$D_{j,t} = \begin{cases} 1 & \text{for } t = T_{j-1}, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for } j = 1, \dots, q \quad \text{and} \quad t = 1, \dots, T,$$

so that $D_{j,t-i} = 1$ if $t = T_{j-1} + i$, and also define indicators for the effective samples as

$$E_{j,t} = \sum_{i=k+1}^{T_j - T_{j-1}} D_{j,t-i} = \begin{cases} 1 & \text{for } T_{j-1} + k < t \leq T_j, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad E_t = (E_{1,t}, \dots, E_{q,t})'.$$

The whole-sample model equation then has the form

$$\begin{aligned} \Delta Y_t = \omega \Delta Z_t + \alpha_y (\beta', \gamma) \begin{pmatrix} X_{t-1} \\ t E_t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_{y,z,i} \Delta X_{t-i} + \mu_{y,z} E_t \\ + \sum_{i=1}^k \sum_{j=2}^q \varphi_{j,i} D_{j,t-i} + \varepsilon_{y,z,t} \quad \text{for } t = k+1, \dots, T, \end{aligned} \quad (7)$$

where $\varphi_{j,i} \in \mathbf{R}^m$ represents a class of parameters for $D_{j,t-i}$ for $i = 1, \dots, k$ and $j = 2, \dots, q$, while the parameters γ and $\mu_{y,z}$ are now redefined in a manner allowing for breaks as

$$\gamma = (\gamma_1, \dots, \gamma_q) \in \mathbf{R}^{r \times q} \quad \text{and} \quad \mu_{y,z} = (\mu_{y,z,1}, \dots, \mu_{y,z,q}) \in \mathbf{R}^{m \times q},$$

which are used in the rest of this study. Equation (5), or its whole-sample form (7), is referred to as the partial model with a broken linear trend term.

2.3 Representations

Various properties of the proposed partial model (5) will be analysed using the Granger-Johansen representation of an $I(1)$ process, which is formulated based on the full model for X_t ; thus, the representation is the same as that in JMN (2000, Theorem 2.1). In JMN each sub-sample period is analysed conditionally on its initial observations. As a result, the representation for each sub-sample period is the same as that in Johansen (1995, Theorem 4.2). The initial values for each sub-sample can be large and thus be influential even in the asymptotic context, but, when following the underlying argument of JMN, one can see that such initial values do not play critical roles in the required asymptotic analysis. We are going to clarify this point by demonstrating further details of the representation than JMN. Following Kurita and Nielsen (2009), we implement this clarification in two steps: first, we analyse a homogeneous equation, and then consider

the roles of deterministic terms by moving to a non-homogeneous equation. For further details, see the proof of Theorem 2.2 below.

For $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$, the full model for X_t with deterministic breaks is defined as a joint system of (5) and (6) in the following manner:

$$\Delta X_t = \alpha(\beta', \gamma_j) \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \mu_j + \varepsilon_t, \quad (8)$$

while the corresponding homogeneous equation is

$$\Delta \tilde{X}_t = \alpha \beta' \tilde{X}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \tilde{X}_{t-i} + \varepsilon_t, \quad (9)$$

where \tilde{X}_t denotes a p -variate *mean-zero* vector time series. We then set up a companion vector based on (9) and analyse a companion form of this equation. Several choices are conceivable with respect to a companion form for (9) and we use the choice that appears, for instance, in Hansen (2005). For the purpose of studying details of the representation, we need an assumption for the parameters, Assumption 2.1 below, which is applicable to both (8) and (9). Some additional notation is required. Johansen (1995, §3)'s notation β_\perp signifies the $p \times (p-r)$ dimensional orthogonal complement to β of full column rank r , so that (β, β_\perp) is invertible and $\beta'_\perp \beta = 0$, along with the normalization $\bar{\beta} = \beta(\beta'\beta)^{-1}$; the same notational conventions are applied to α of full column rank r . It is also necessary to define $\Psi = I_p - \sum_{i=1}^{k-1} \Gamma_i$.

Assumption 2.1 *Assume that the roots of the characteristic polynomial,*

$$A(z) = (1-z)I_p - \alpha\beta'z - \sum_{i=1}^{k-1} \Gamma_i(1-z)z^i,$$

are outside the complex unit circle or at unity; furthermore, assume that the matrices α and β have full column rank r and that the square matrix $\alpha'_\perp \Psi \beta_\perp$ has full rank $p-r$.

Given Assumption 2.1, we also need to define $C = \beta_\perp(\alpha'_\perp \Psi \beta_\perp)^{-1} \alpha'_\perp$ for future reference, which is often referred to as the impact matrix in cointegration literature; see Paruolo (1997) for inference on this matrix.

We are approaching the stage where the Granger-Johansen representation for each sub-sample period is presented. For the homogeneous equation (9), let us define

$$\alpha = \begin{pmatrix} \alpha & \Gamma_1 & \cdots & \Gamma_{k-1} \\ 0 & I_p & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & I_p \end{pmatrix}, \quad \Lambda = \begin{pmatrix} I_p & 0 & \cdots & 0 \\ I_p & -I_p & & \vdots \\ 0 & & \ddots & \\ \vdots & & & 0 \\ 0 & \cdots & 0 & I_p & -I_p \end{pmatrix}, \quad (10)$$

$$\beta' = \begin{pmatrix} \beta' & 0 \\ 0 & I_{p(k-1)} \end{pmatrix} \Lambda, \quad \tilde{\mathbf{X}}_{t-1} = \begin{pmatrix} X_{t-1} \\ \vdots \\ X_{t-k} \end{pmatrix}, \quad (11)$$

and $\iota = (I_p, 0, \dots, 0)'$ together with $\mathbf{r} = r + p(k-1)$. The representation is then given in the theorem below, the proof of which is provided in Appendix B.

Theorem 2.2 *Suppose that Assumption 2.1 is fulfilled. Then, for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$, an \mathbf{r} -variate process $\beta' \tilde{\mathbf{X}}_t$ derived from (9) satisfies*

$$\beta' \tilde{\mathbf{X}}_t = (I_{\mathbf{r}} + \beta' \alpha) \beta' \tilde{\mathbf{X}}_{t-1} + \beta' \boldsymbol{\varepsilon}_t \quad \text{with} \quad |\text{eigen}(I_{\mathbf{r}} + \beta' \alpha)| < 1, \quad (12)$$

which is a stable first-order vector autoregression. Moreover, for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$, the solution to (9) is given as

$$\begin{aligned} \tilde{X}_t = C \sum_{s=T_{j-1}+k+1}^t \varepsilon_s + \{(I - C\Psi)\bar{\beta}, C\Upsilon_1, C\Upsilon_2, \dots, C\Upsilon_{k-1}\} \beta' \tilde{\mathbf{X}}_t \\ - C(\Psi, \Upsilon_1, \Upsilon_2, \dots, \Upsilon_{k-1}) \Lambda \tilde{\mathbf{X}}_{T_{j-1}+k}, \end{aligned} \quad (13)$$

where $\Upsilon_i = -\Gamma_i - \dots - \Gamma_{k-1}$ for $1 \leq i \leq k-1$. Thus, for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$, the variable X_t in (8) satisfies

$$\begin{aligned} X_t = C \sum_{s=T_{j-1}+k+1}^t \varepsilon_s + (I - C\Psi)\bar{\beta} \beta' \tilde{X}_t - C \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{t-\ell} \\ - C\Psi \tilde{X}_{T_{j-1}+k} + C \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{T_{j-1}+k-\ell} + \tau_{c,j} + \tau_{\ell,j} t, \end{aligned} \quad (14)$$

for $\tilde{X}_t = X_t - \tau_{c,j} - \tau_{\ell,j} t$ with the parameters $\tau_{c,j}$ and $\tau_{\ell,j}$ satisfying

$$\Psi \tau_{l,j} = \alpha \beta' (\tau_{c,j} - \tau_{l,j}) + \mu_j \quad \text{and} \quad \beta' \tau_{l,j} + \gamma_j = 0.$$

Note that the initial observations for the j -th sub-sample in (14) are expressed in terms of linear combinations of the mean-zero values $\tilde{X}_{T_{j-1}+1}, \dots, \tilde{X}_{T_{j-1}+k}$, so that we can in general argue that the starting values for each sub-sample period do not play critical roles in asymptotic analysis. This property was not explicitly examined in JMN. Thus, Theorem 2.2 can be seen as a useful clarification of roles of the initial values in the full cointegrated model subject to deterministic breaks. The Granger-Johansen representation is utilised in proofs of asymptotic theorems in Section 3 under Assumption 2.1.

As an alternative to the above sub-sample representation, one can derive a joint representation for the whole sample. For this purpose we need a full system equation for X_t over the entire sample period. This equation is derived from a combination of (8) over $j = 1, \dots, q$ augmented with dummies $D_{j,t-i}$ and $E_{j,t}$, as in (7); that is,

$$\Delta X_t = \alpha (\beta', \gamma) \begin{pmatrix} X_{t-1} \\ tE_t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \mu E_t + \sum_{i=1}^k \sum_{j=2}^q \kappa_{j,i} D_{j,t-i} + \varepsilon_t, \quad (15)$$

where $\kappa_{j,i} \in \mathbf{R}^p$ for $i = 1, \dots, k$ and $j = 2, \dots, q$, and $\mu = (\mu_1, \dots, \mu_q) \in \mathbf{R}^{p \times q}$; see equation (2.6) in JMN. We then replace the innovations ε_t with $\varepsilon_t^D = \varepsilon_t + \alpha \gamma' t E_t + \mu E_t + \sum_{i=1}^k \sum_{j=2}^q \kappa_{j,i} D_{j,t-i}$ to reach a whole-sample representation such as

$$X_t \approx C \sum_{s=k+1}^t \varepsilon_s^D + C_1(L) \varepsilon_t^D + A \quad \text{for} \quad k < t \leq T, \quad (16)$$

where $C_1(L)\varepsilon_t^D$ denotes a moving-average process whose coefficients decrease exponentially fast, and A depends on initial observations X_1, \dots, X_k , satisfying $\beta' A = 0$. As before the deterministic parts of the common trends $C \sum_{s=k+1}^T \varepsilon_s^D$ will be piecewise constant since $C\alpha = 0$, so that each constant fails to cumulate to a linear trend. A similar approach was adopted in $I(2)$ cointegration analysis by Kurita, Nielsen and Rahbek (2011). The representation (16) is clear and concise, but the transition from one regime to another is considered to be explicitly autoregressive, which may leave less flexibility to represent regime transitions of some persistent and messy nature. For the asymptotic study conducted below, we follow JMN by using the sub-sample representation (14).

2.4 The partial model with shifts in the level

In some applications it suffices to exclude the broken linear trend term and just include shifts in the level (a broken constant). By restricting the broken constant term within the cointegrating space, equation (7) for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$ is reduced to

$$\Delta Y_t = \omega \Delta Z_t + \alpha_y (\beta', \gamma) \begin{pmatrix} X_{t-1} \\ E_t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_{y,z,i} \Delta X_{t-i} + \sum_{i=1}^k \sum_{j=2}^q \varphi_{j,i} D_{j,t-i} + \varepsilon_{y,z,t}. \quad (17)$$

The Granger-Johansen representation has the same form as (14) but is subject to

$$\beta' \tau_{c,j} + \gamma'_j = 0 \quad \text{and} \quad \tau_{\ell,j} = 0.$$

3 Testing for cointegrating rank in the partial models

This section addresses the issue of testing for cointegrating rank in the suggested partial models with deterministic shifts. Section 3.1 introduces a partial likelihood ratio test for the choice of rank based on the broken linear-trend model and Section 3.2 derives its limit distribution. Section 3.3 then turns to the broken constant model and examines the test statistic based upon it. Finally, Section 3.4 derives a class of approximations to the limit distributions by means of computer simulations and response surface regression.

3.1 Rank test statistic

For each sub-sample period, the partial model (5) or (7) is seen as equivalent to that in HJNR, given the presence of structural breaks in its deterministic terms. This model can thus be analysed by reduced rank regression in a manner similar to the original cointegration model in Johansen (1988, 1995). We express below that reduced rank regression is applicable to the proposed partial model by following three different approaches.

The first approach is based on a full-sample reduced rank regression. Regress each of the vectors ΔY_t and $X_{t-1}^\circ = (X'_{t-1}, tE'_t)'$ on a vector H_t consisting of the variables ΔZ_t , the lagged differences $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$, the intercepts E_t and the impulse dummies $D_{j,t-i}$ for $i = 1, \dots, k$ and $j = 2, \dots, q$, so that H_t has dimension $pk - m + q + k(q - 1)$. This gives residuals $R_{0,t}$ and $R_{1,t}$:

$$\begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix} = \begin{pmatrix} \Delta Y_t \\ X_{t-1}^\circ \end{pmatrix} - \sum_{s=k+1}^T \begin{pmatrix} \Delta Y_s \\ X_{s-1}^\circ \end{pmatrix} H'_s \left(\sum_{s=k+1}^T H_s H'_s \right)^{-1} H_t \quad \text{for } k < t \leq T.$$

The second approach is viewed as a sub-sample approach. We note that the impulse dummies result in a perfect fit for each transitional period in between two connecting regimes; thus $R_{0,t}$ and $R_{1,t}$ are zero for all transitional periods. We can, therefore, compute the residuals $R_{0,t}$ and $R_{1,t}$ by analysing the effective sub-sample periods only; see also Doornik, Hendry and Nielsen (1998, Sec. 12.2). For this purpose, let us form regressors P_t from the variables ΔZ_t , the lagged differences $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$ and the intercepts E_t , so that P_t is a vector of dimension $pk - m + q$. The residuals $R_{0,t}$ and $R_{1,t}$ then satisfy

$$\begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix} = \begin{pmatrix} \Delta Y_t \\ X_{t-1}^\circ \end{pmatrix} - \sum_{j=1}^q \sum_{s=T_{j-1}+k+1}^{T_j} \begin{pmatrix} \Delta Y_s \\ X_{s-1}^\circ \end{pmatrix} P'_s \left(\sum_{j=1}^q \sum_{s=T_{j-1}+k+1}^{T_j} P_s P'_s \right)^{-1} P_t,$$

for $T_{j-1} + k < t \leq T_j$ with $1 \leq j \leq q$, while $R_{0,t}$ and $R_{1,t}$ are zero otherwise.

The third approach is recognised as a two-step approach, in which we first demean the observed time series and then partial out influences from the lagged differences. In the first step, we analyse two vectors ΔY_t and X_{t-1}° , along with a vector V_t consisting of the variables ΔZ_t and the lagged differences $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$. These three vectors are demeaned within each sub-sample period, yielding $\mathcal{Z}_{0,t}$, $\mathcal{Z}_{1,t}$ and $\mathcal{Z}_{2,t}$ defined as

$$\begin{pmatrix} \mathcal{Z}_{0,t} \\ \mathcal{Z}_{1,t} \\ \mathcal{Z}_{2,t} \end{pmatrix} = \begin{pmatrix} \Delta Y_t \\ X_{t-1}^\circ \\ V_t \end{pmatrix} - \frac{1}{T_j - T_{j-1} + k} \sum_{s=T_{j-1}+k+1}^{T_j} \begin{pmatrix} \Delta Y_s \\ X_{s-1}^\circ \\ V_s \end{pmatrix}, \quad (18)$$

for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$ and zero otherwise. In the second step we compute

$$\begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix} = \begin{pmatrix} \mathcal{Z}_{0,t} \\ \mathcal{Z}_{1,t} \end{pmatrix} - \sum_{s=k+1}^T \begin{pmatrix} \mathcal{Z}_{0,s} \\ \mathcal{Z}_{1,s} \end{pmatrix} \mathcal{Z}'_{2,s} \left(\sum_{s=k+1}^T \mathcal{Z}_{2,s} \mathcal{Z}'_{2,s} \right)^{-1} \mathcal{Z}_{2,t}.$$

Since $\mathcal{Z}_{0,t}$, $\mathcal{Z}_{1,t}$ are zero within the transitional periods, so are the residuals $R_{0,t}$, $R_{1,t}$.

With the residuals $R_{0,t}$ and $R_{1,t}$ in hand, we can compute the product moments

$$\begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix} = \frac{1}{T-k} \sum_{t=k+1}^T \begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix} \begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix}', \quad (19)$$

and a set of squared canonical correlations $1 \geq \hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_m \geq 0$ by solving the eigenvalue problem

$$0 = \det(\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}).$$

Hence, the log partial likelihood ratio (*PLR*) test statistic for the null hypothesis of cointegrating rank r , $\mathbf{H}_\ell(r)$, against the hypothesis $\mathbf{H}_\ell(m)$ is

$$PLR\{\mathbf{H}_\ell(r)|\mathbf{H}_\ell(m)\} = -(T-k) \sum_{i=r+1}^m \log(1 - \hat{\lambda}_i). \quad (20)$$

3.2 Asymptotic distribution of the test statistic

We derive the asymptotic distribution of the rank test statistic in a setting where the relative break points satisfy $T_j/T \rightarrow v_j$ for $j = 0, \dots, q$ while T goes to infinity. The

relative break points satisfy $0 = v_0 < v_1 < \dots < v_q = 1$. Define a q -variate vector of indicator functions as

$$e_u = \{1_{(v_0 < u < v_1)}, \dots, 1_{(v_{q-1} < u < v_q)}\}', \quad (21)$$

which is viewed as the limit of $E_{int(Tu)}$ on $u \in [0, 1]$ as $T \rightarrow \infty$ with $v_j = T_j/T$ fixed.

In the standard framework developed by Johansen (1995), the innovation sequence ε_t is assumed to be independent and identically Gaussian distributed, and this assumption was adopted by HJNR and JMN as reviewed in Section 2.1 above. We relax this normality assumption to a martingale difference assumption. If the innovations ε_t are not normal, the model equations lead to a quasi-likelihood function rather than a likelihood function. Weak exogeneity is preserved as it is a property of the likelihood rather than the distribution of the innovations as such. The partial innovation $\varepsilon_{y,z,t} = \varepsilon_{y,t} - \omega \varepsilon_{z,t}$ and the marginal innovation $\varepsilon_{z,t}$ are uncorrelated, but they will not be independent in general when moving away from the normality assumption. We can no longer appeal to the conditional-distribution argument, as implied in equation (2). Thus, the conditional-distribution argument is replaced with a regression argument, see Appendix C.2. The martingale difference assumption is summarised as Assumption 3.1 below.

Assumption 3.1 *Assume that ε_t is a martingale difference sequence with respect to a filtration \mathcal{F}_t such that $\mathbf{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ almost surely (a.s.). Let Ω be a positive definite matrix. Suppose that*

- (i) $T^{-1} \sum_{t=1}^T \mathbf{E}(\varepsilon_t \varepsilon_t') \rightarrow \Omega$;
- (ii) $T^{-1} \sum_{t=1}^T \mathbf{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) \xrightarrow{P} \Omega$;
- (iii) *either of the following boundedness conditions*
 - (a) $\sup_{t \in \mathbb{N}} \mathbf{E}\{\varepsilon_t' \varepsilon_t 1_{(\varepsilon_t' \varepsilon_t > a)} | \mathcal{F}_{t-1}\} \xrightarrow{P} 0$ as $a \rightarrow \infty$;
 - (b) $\sup_{t \in \mathbb{N}} \mathbf{E}|\varepsilon_t|^4 < \infty$.

The boundedness conditions in part (iii) are not nested. Part (a) can be satisfied without the existence of fourth moments as in part (b). Conversely, bounded fourth moments in part (b) do not necessarily imply part (a); see Remark C.4 in Appendix C.1.

Under Assumption 3.1, we are able to apply the results of Brown (1971) to analyse the random walk components of the process. For this we require a Lindeberg condition, which is established in Lemma C.2 in Appendix C.1 under Assumption 3.1. Brown's result is for univariate martingale difference sequences and requires that the ratio of the sum of conditional variances to that of unconditional variances should converge to unity. For the multivariate case we can apply the Cramér-Wold device and form linear combinations of the present multivariate martingale differences. Using parts (i), (ii) we can then show that Brown's ratios converge to unity.

Under Assumption 3.1, we can also analyze the (approximately) stationary components of the process. Under part (iii.a) we can apply the results of Anderson and Kunitomo (1992), which exploit a truncation argument. Under part (iii.b) we can apply the same ideas as in Anderson and Kunitomo (1992) but without the truncation argument.

Cointegration models with heteroscedasticity have previously been analyzed by for instance Cavaliere, Rahbek and Taylor (2010) and Boswijk, Cavaliere, Rahbek and Taylor (2016). The former paper is concerned with rank testing in a full system. For the analysis of the (approximately) stationary components it relies on Hannan and Heyde (1972) with slightly different conditions from here. See Remark C.5 in Appendix C.1. The latter paper is concerned with testing on the cointegrating vectors in a full system with an elaborate,

deterministic structure for the variances of the innovations. We note that Assumption 3.1 does not require any particular structure for the individual variances for the innovations.

Before proceeding to the main results, we present a set of stronger assumptions requiring constant conditional variance, see Assumption 3.2 and Lemma 3.3 below. These assumptions were used by Lai and Wei (1982, 1985) as well as Chan and Wei (1988). They have two advantages. First, they are easier to check for practitioners than the convergence results in Assumption 3.1. Second, the assumptions can also be used to derive a variety of almost sure convergence results for least squares estimators, as explored by Lai and Wei (1982, 1985) and Nielsen (2005), although we will not exploit those properties here.

Assumption 3.2 *Assume that ε_t is a martingale difference sequence with respect to a filtration \mathcal{F}_t such that $\mathbf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s. and*

(i) *$\text{Var}(\varepsilon_t|\mathcal{F}_{t-1}) = \Omega$ a.s., where Ω is positive definite;*

(ii) *$\sup_{t \in \mathbb{N}} \mathbf{E}(|\varepsilon_t|^{2+\xi}|\mathcal{F}_{t-1}) < \infty$ a.s. for some $\xi > 0$.*

Lemma 3.3 *Assumption 3.2 implies Assumption 3.1.*

We are in a position to present the limit distribution of the *PLR* statistic (20), which should now be viewed as a log partial quasi-likelihood ratio test statistic, although it will continue to be denoted as the *PLR* test statistic. For this purpose, let \xrightarrow{D} signify weak convergence, while let B_u represent a $(p - r)$ -dimensional standard Brownian motion process on $u \in [0, 1]$ and let $B_u^{(m-r)}$ be the first $m - r$ coordinates of B_u . The limit distribution is given in the next theorem, which is proved in Appendix C.3.

Theorem 3.4 *Suppose that Assumptions 2.1 and 3.1 are satisfied along with $\alpha_z = 0$, so that Z_t is weakly exogenous with respect to α_y , β and γ . As $T \rightarrow \infty$, with relative break point satisfying $T_j/T \rightarrow v_j$ for $0 = v_0 < v_1 < \dots < v_q = 1$, the *PLR* test statistic (20) under $\mathbf{H}_\ell(r)$ satisfies*

$$PLR\{\mathbf{H}_\ell(r)|\mathbf{H}_\ell(m)\} \xrightarrow{D} \text{DF}_\ell(m - r, p - r; v), \quad (22)$$

where, with e_u defined in (21),

$$\begin{aligned} \text{DF}_\ell(m - r, p - r; v) &= \text{tr} \left\{ \int_0^1 dB_u^{(m-r)} G'_u \left(\int_0^1 G_u G'_u du \right)^{-1} \int_0^1 G_u dB_u^{(m-r)'} \right\}, \\ G_u &= \begin{pmatrix} B_u \\ ue_u \end{pmatrix} - \int_0^1 \begin{pmatrix} B_s \\ se_s \end{pmatrix} e_s' ds \left(\int_0^1 e_s e_s' ds \right)^{-1} e_u. \end{aligned}$$

Note that, when $p = m$, the result in Theorem 3.4 corresponds to Theorem 3.1 in JMN. A direct simulation of (22) is rather laborious. By exploiting some analytic properties of the distributions, we are able to simplify this simulation task. The next Theorem 3.5 describes these properties by linking the moments of the limit distribution in Theorem 3.4 to those for the full model. Theorem 3.5 provides a basis for simulation in Section 3.4. The proof of this theorem, given in Appendix C.3, is based on a slight modification of results in Doornik (1998, Sec. 9); see also Boswijk and Doornik (2005).

Theorem 3.5 Let $B_{i,u}$ be the i -th coordinate of the Brownian motion B_u . Let

$$\mathbb{T}_i = \int_0^1 dB_{i,u} G'_u \left(\int_0^1 G_u G'_u du \right)^{-1} \int_0^1 G_u dB'_{i,u} \quad \text{for } i = 1, \dots, p-r.$$

Then $\mathbb{T}_1, \dots, \mathbb{T}_{p-r}$ are identically distributed and any pairs $\mathbb{T}_j, \mathbb{T}_k$ are also identically distributed. Moreover, the limiting statistic (21) satisfies $\text{DF}_\ell(m-r, p-r; v) = \sum_{i=1}^{m-r} \mathbb{T}_i$ with expectation and variance given by

$$\begin{aligned} \mathbb{E}\{\text{DF}_\ell(m-r, p-r; v)\} &= \left(\frac{m-r}{p-r} \right) \mathbb{E} \left(\sum_{i=1}^{p-r} \mathbb{T}_i \right), \\ \text{Var}\{\text{DF}_\ell(m-r, p-r; v)\} &= \left(\frac{m-r}{p-r} \right) \text{Var} \left(\sum_{i=1}^{p-r} \mathbb{T}_i \right) - (m-r)(p-m) \text{Cov}(\mathbb{T}_1, \mathbb{T}_2). \end{aligned}$$

Since the above Theorem 3.5 links the moments of the statistics for partial systems and for full system, we can now proceed by simulating distributions for full systems only. JMN simulated response surfaces for the mean and variance of $\sum_{i=1}^{p-r} \mathbb{T}_i$. As we will also need a response surface for $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$ we have to redo their simulation exercise. For this purpose we quote a result from JMN.

Theorem 3.6 (JMN, Theorem 3.2) Let $B^{[1]}, \dots, B^{[q]}$ be independent $(p-r)$ -dimensional standard Brownian motions and define

$$\begin{aligned} J_j &= \left\{ \int_0^1 (u|1)^2 du \right\}^{-1/2} \int_0^1 (u|1) \{dB_u^{[j](m-r)}\}, \\ K_j &= \int_0^1 (B_u^{[j]} | u, 1) \{dB_u^{[j](m-r)}\}', \\ L_j &= \int_0^1 (B_u^{[j]} | u, 1) (B_u^{[j]} | u, 1)' du. \end{aligned}$$

Then the limiting variable (21) for a full sample with $p = m$ satisfies

$$\text{DF}_\ell(p-r, p-r; v) = \text{tr} \left[\left(\sum_{j=1}^q K_j \Delta v_j \right)' \left\{ \sum_{j=1}^q L_j (\Delta v_j)^2 \right\}^{-1} \left(\sum_{j=1}^q K_j \Delta v_j \right) \right] + \sum_{j=1}^q J'_j J_j$$

where $\Delta v_j = v_j - v_{j-1}$. Here the two summands are independent and $\sum_{j=1}^q J'_j J_j$ is distributed as $\chi^2\{q(m-r)\}$. Moreover, let $J_j^{[i]}$ and $K_j^{[i]}$ denote the i th coordinate of J_j and K_j so that

$$\mathbb{T}_i = \left(\sum_{j=1}^q K_j^{[i]} \Delta v_j \right)' \left\{ \sum_{j=1}^q L_j (\Delta v_j)^2 \right\}^{-1} \left(\sum_{j=1}^q K_j^{[i]} \Delta v_j \right) + \sum_{j=1}^q (J_j^{[i]})^2.$$

As in JMN we note the Theorem 3.6 implies a simple relation between the limiting statistics for models with q and with $q-1$ sub-sample periods, that is

$$\lim_{\Delta v_q \rightarrow 0} \text{DF}_\ell(p-r, p-r; v_1, \dots, v_{q-1}, v_q) = \text{DF}_\ell(p-r, p-r; v_1, \dots, v_{q-1}) + J'_q J_q, \quad (23)$$

where $\text{DF}_\ell(p-r, p-r; v_1, \dots, v_{q-1})$ and $J'_q J_q$ are independent and $J'_q J_q$ is $\chi^2(p-r)$.

3.3 Asymptotic distribution for the broken constant case

The model investigated in the previous sub-sections has a broken linear trend. A variant of this model is free of such a linear trend but with a broken constant; see equation (17). Equation (5) then reduces to

$$\Delta Y_t = \omega \Delta Z_t + \alpha_y(\beta', \gamma_j) \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_{y,z,i} \Delta X_{t-i} + \varepsilon_{y,z,t},$$

for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$. As before the partial quasi-likelihood is maximized by reduced rank regression. We follow the third approach (18) in Section 3.1, in which the broken linear trend is now replaced with the broken constant, so that we consider the vectors ΔY_t and $X_{t-1}^\circ = (X'_{t-1}, E'_t)'$, together with the vector V_t composed of the variables ΔZ_t and the lagged differences $\Delta X_{t-1}, \dots, \Delta X_{t-k+1}$. Equation (18) then reduces to

$$\begin{pmatrix} Z_{0,t} \\ Z_{1,t} \\ Z_{2,t} \end{pmatrix} = \begin{pmatrix} \Delta Y_t \\ X_{t-1}^\circ \\ V_t \end{pmatrix}.$$

The limit distribution of the log partial quasi-likelihood ratio test statistic for cointegrating rank r , denoted by $PLR\{H_c(r)|H_c(m)\}$, is given in Theorem 3.7 below. Its proof is based on a set of modifications of the proofs for the limit theorems in the previous sub-sections.

Theorem 3.7 *Suppose that Assumptions 2.1 and 3.1 are satisfied along with $\alpha_z = 0$, so that Z_t is weakly exogenous with respect to α_y , β and γ . As $T \rightarrow \infty$, with relative break point satisfying $T_j/T \rightarrow v_j$ for $0 = v_0 < v_1 < \dots < v_q = 1$, the PLR test statistic (20) under $H_c(r)$ satisfies*

$$PLR\{H_c(r)|H_c(m)\} \xrightarrow{D} DF_c(m-r, p-r; v),$$

where DF_c is defined as in Theorem 3.4 with the difference that

$$G_u = \begin{pmatrix} B_u \\ e_u \end{pmatrix}.$$

The results in Theorems 3.5, 3.6 also apply with the present choice of G_u .

3.4 Approximations of the asymptotic distributions

The limit distributions of the cointegrating rank test statistics are non-standard, as shown in the previous sub-sections; however, given the existing results in literature, the distributions can be closely approximated by a gamma distribution identified by the first two moments, mean and variance. In addition, the theorems and corollary in the previous sub-sections will simplify the application of this approximation method to the rank test statistics in the proposed partial models.

The quality of the gamma-distribution approximation method has been documented in several papers. Using analytic methodology, Nielsen (1997) showed a very good agreement between limit distributions and approximate gamma distributions in tests for unit roots. Doornik (1998) then conducted detailed simulation studies to demonstrate a similar agreement for standard full-system cointegration rank test statistics; see also Doornik

(2003) for various tables of asymptotic quantiles produced by the gamma-distribution approximations. Furthermore, JMN employed this method to test for cointegrating rank in the full system subject to deterministic breaks. The gamma-distribution approximation method has thus facilitated testing for cointegrating rank in various model settings.

In order to apply the gamma approximation method we first define parameters for shape, λ_F , and scale, δ_F , for the full system statistic $\text{DF}_\ell(p-r, p-r; v) = \sum_{i=1}^{p-r} \mathbf{T}_i$ where $m = p$ and $Y_t = X_t$ so Z_t is absent. These parameters are

$$\frac{1}{\lambda_F} = \frac{\text{Var}(\sum_{i=1}^{p-r} \mathbf{T}_i)}{\{\mathbf{E}(\sum_{i=1}^{p-r} \mathbf{T}_i)\}^2}, \quad \delta_F = \frac{\text{Var}(\sum_{i=1}^{p-r} \mathbf{T}_i)}{\mathbf{E}(\sum_{i=1}^{p-r} \mathbf{T}_i)}.$$

Below we present response surface approximations to λ_F and δ_F based on an extensive simulation study. We also show that the distribution of $\text{DF}_\ell(p-r, p-r; v)$ can be approximated well by a gamma distribution with shape λ_F and scale δ_F .

For the partial system statistic $\text{DF}_\ell(m-r, p-r; v) = \sum_{i=1}^{m-r} \mathbf{T}_i$ we defined shape λ_P and scale δ_P in the same way. Replace $\text{Var}(\sum_{i=1}^{m-r} \mathbf{T}_i)$ and $\mathbf{E}(\sum_{i=1}^{m-r} \mathbf{T}_i)$ with the expressions in Theorem 3.5 and then replace $\text{Var}(\sum_{i=1}^{p-r} \mathbf{T}_i) = \delta_F^2 \lambda_F$ and $\mathbf{E}(\sum_{i=1}^{p-r} \mathbf{T}_i) = \delta_F \lambda_F$ to get

$$\frac{1}{\lambda_P} = \frac{\text{Var}(\sum_{i=1}^{m-r} \mathbf{T}_i)}{\{\mathbf{E}(\sum_{i=1}^{m-r} \mathbf{T}_i)\}^2} = \frac{p-r}{\lambda_F(m-r)} \left\{ 1 - \frac{(p-r)(p-m)}{\lambda_F \delta_F^2} \text{Cov}(\mathbf{T}_1, \mathbf{T}_2) \right\}, \quad (24)$$

$$\delta_P = \frac{\text{Var}(\sum_{i=1}^{m-r} \mathbf{T}_i)}{\mathbf{E}(\sum_{i=1}^{m-r} \mathbf{T}_i)} = \delta_F - \frac{(p-r)(p-m)}{\lambda_F \delta_F} \text{Cov}(\mathbf{T}_1, \mathbf{T}_2). \quad (25)$$

Thus, by simulating $\text{Cov}(\mathbf{T}_1, \mathbf{T}_2)$ in addition to λ_F and δ_F we can compute the shape and scale for the $\text{DF}_\ell(m-r, p-r; v)$ distribution. Through simulation we will show that the $\text{DF}_\ell(m-r, p-r; v)$ distribution can also be approximated well in numerical terms by a gamma distribution with shape λ_P and scale δ_P .

This indirect method, by means of equations (24) and (25), saves a lot of computational costs in comparison with approximating λ_P and δ_P directly. Such direct approximations require a huge number of simulations of partial-system-based tests, depending on numerous combinations of $m-r$, $p-r$ and relative break points. In addition, the direct approximations will generate another problem: difficulty in presenting the results neatly. Presenting the results obtained by the direct method will require huge tables including, again, numerous combinations of $m-r$, $p-r$ and relative break points. These tables may be too cumbersome to use in practical applications.

We require a set of precise approximates for λ_F , δ_F and $\text{Cov}(\mathbf{T}_1, \mathbf{T}_2)$, in order to conduct reliable inference using the equations (24) and (25). Following JMN and Doornik (1998) we simulate these quantities for various data generating processes and conduct a response surface analysis. We take logs of the dependent variables λ_F , δ_F to alleviate residual heteroscedasticity. Compared to JMN we increase the maximum number of observations from $T = 500$ to $T = 2000$. It is then found that the large-sample ($T \geq 1,000$) approximates of the mean and variance in small dimensions ($p-r \leq 3$) tend to be rather different from those when T is small, in comparison with a difference between those moment approximates for larger dimensions ($p-r > 3$). This finding is consistent with Doornik (1998), who introduced a set of indicator variables being assigned 1 for $p-r = 2$ and $p-r = 1$ and assigned 0 otherwise; these indicators put residual heteroscedasticity under control even in the presence of influential values for $p-r = 2$ and $p-r = 1$.

$p - r$	$m - r$	a	b	q_{95}	q_{95}^*	$ q_{95}/q_{95}^* - 1 $
2	1	0.0	0.0	15.45	15.33	0.0078
2	1	0.0	0.3	21.25	21.25	0.0000
2	1	0.1	0.4	25.63	25.76	0.0050
2	1	0.2	0.3	27.23	27.11	0.0044
2	1	0.3	0.3	27.74	27.62	0.0043
4	3	0.0	0.0	50.29	50.08	0.0042
4	3	0.0	0.3	65.09	64.97	0.0018
4	3	0.1	0.4	77.01	76.84	0.0022
4	3	0.2	0.3	80.25	80.11	0.0017
4	3	0.3	0.3	81.92	81.84	0.0010
5	3	0.0	0.0	57.35	57.32	0.0005
5	3	0.0	0.3	72.27	72.03	0.0033
5	3	0.1	0.4	84.00	83.98	0.0002
5	3	0.2	0.3	87.23	87.10	0.0015
5	3	0.3	0.3	88.44	88.47	0.0003
7	4	0.0	0.0	91.64	91.79	0.0016
7	4	0.0	0.3	110.97	110.81	0.0014
7	4	0.1	0.4	126.33	126.34	0.0001
7	4	0.2	0.3	130.53	130.07	0.0035
7	4	0.3	0.3	131.26	131.45	0.0014

Table 1: A comparative analysis of 95% limit quantiles: broken linear-trend models

With a view to addressing this problem, our response surface study also employs such indicator variables. Moreover, it adopts the strategy of using a set of simulated approximates for $\log \lambda_F$, $\log \delta_F$ and $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$ as the response variables, instead of the logged means and variance as in JMN. It turns out that the use of these response variables ($\log \lambda_F$, in particular) mitigates the residual heteroscedasticity problem, hence resulting in the reduction in the number of indicator variables required for $p - r = 2$ and $p - r = 1$. Note that $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$ needs to be included in the set of response variables in any response surface study, in order to make use of equations (24) and (25). In addition, note that taking the log of $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$ is not permissible, since covariance is not always positive.

All the response variables for the full model were simulated for a set of given $p - r$, T and relative break points. Following JMN, we chose $q = 3$ as the maximum number of sub-samples, with a and b representing the smallest and the second smallest of relative sample lengths, respectively. For example, if $q = 2$ along with $v_1 < 1 - v_1$, we then have $a = 0$ and $b = v_1$. The decimals a and b were selected in the same way as those for Figure 1 in JMN e.g., $(a, b) = (0, 0), (0, 0.05), (0, 0.1), \dots$, so that they were subject to the constraints of $a \leq b$ and $b < (1 - a - b)$ and the total number of their combinations was 20, along with the selection of non-stationary components $p - r = 1, \dots, 8$. For the overall sample sizes or T s, JMN used 10 integers derived from $500/i$ for $i = 1, \dots, 10$, but we quadrupled them in order to improve approximations to the underlying limit distributions of the response variables. Thus, we obtained a new set of 10 sample sizes, T s, ranging from 200 to 2,000. For $\log \lambda_F$ and $\log \delta_F$, this simulation design led to 1,600 ($= 20 \times 8 \times 10$) cases, while the number of cases was reduced to 1,400 for $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$ as a result of missing values corresponding to $p - r = 1$.

$p - r$	$m - r$	a	b	q_{95}	q_{95}^*	$ q_{95}/q_{95}^* - 1 $
2	1	0.0	0.0	12.21	12.28	0.0057
2	1	0.0	0.3	15.51	15.55	0.0026
2	1	0.1	0.4	18.24	18.35	0.0060
2	1	0.2	0.3	18.71	18.75	0.0021
2	1	0.3	0.3	18.81	18.87	0.0032
4	3	0.0	0.0	42.76	42.60	0.0038
4	3	0.0	0.3	50.66	50.71	0.0010
4	3	0.1	0.4	57.40	57.67	0.0047
4	3	0.2	0.3	58.63	58.69	0.0010
4	3	0.3	0.3	58.83	58.90	0.0012
5	3	0.0	0.0	50.06	49.96	0.0020
5	3	0.0	0.3	57.88	57.73	0.0026
5	3	0.1	0.4	64.64	64.73	0.0014
5	3	0.2	0.3	65.66	65.50	0.0024
5	3	0.3	0.3	65.62	65.66	0.0006
7	4	0.0	0.0	82.47	82.35	0.0015
7	4	0.0	0.3	92.22	92.26	0.0004
7	4	0.1	0.4	101.46	101.56	0.0010
7	4	0.2	0.3	102.01	102.04	0.0003
7	4	0.3	0.3	101.81	102.16	0.0034

Table 2: A comparative analysis of 95% limit quantiles: broken constant models

The computational algorithm used in our study was based on Theorem 3.6 in JMN. These asymptotic results justify simulating three sets of T -step random walks for broken linear-trend and constant cases and scaling them according to the pre-specified relative sample lengths. The number of simulation replications N was set at 100,000.

We have reached a stage to explain further details of response surface regression equations. We can regard the simulated response variables as functions of a , b , $p - r$ and T ; however, they are unknown in functional forms. Response surface regression enables us to approximate these unknown functions. We adopted a class of modified versions of regression equation (3.11) in JMN as the baseline functional forms for our three response variables, $\log \lambda_F$, $\log \delta_F$ and $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$. Following Doornik (1998), we also added to the equation a set of indicator variables as explanatory variables, each of which is 1 for a selected value of dimension $p - r$ and is 0 otherwise. By performing a series of regression analyses and carefully removing insignificant explanatory variables, we arrived at parsimonious response-surface functions for $\log \lambda_F$, $\log \delta_F$ and $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$; these functions are henceforth denoted $f_F^z(p - r, a, b, T)$ with z taking values λ , δ and cov , respectively.

Tables 6 and 7 in Appendix A record rounded values of coefficients for a , b , $p - r$ and their variants in the response surface regression for the broken linear trend case and the broken constant case, respectively. The inverse of the observation number, T^{-1} , and its variants such as T^{-2} , also play critical roles in the response surface regression, but all of them are irrelevant asymptotically and thus disregarded when calculating the limit approximates based on these tables.

It should also be noted that a response-surface regression analysis of $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$ was technically difficult in terms of residual diagnostic tests. Doornik (1998) used the average

of estimates for $\text{Cov}(T_1, T_2)$ when performing a response surface analysis for partial systems with no break. We adhered to the regression approach, rather than simply taking the average of the covariance estimates, by assigning importance to various significant influences of a , b and $p - r$ on the behaviour of $\text{Cov}(T_1, T_2)$. This regression analysis indeed bore fruit and clarified the highly complex structure of the dependence of $\text{Cov}(T_1, T_2)$ on a , b , $p - r$ and its variants, as shown in the third column of each of Tables 6 and 7. These findings about $\text{Cov}(T_1, T_2)$ are not known in the literature, thus giving added value to the response surface study conducted in this paper.

With regard to Tables 6 and 7, it would be useful to give an instruction for how to use them. First, construct a regression equation for each of the three response variables by referring to the coefficients reported in either Table 6 or Table 7 (according to the specification of deterministic terms). Note that the explanatory variables for each equation consist of linear and non-linear functions of a , b and $p - r$. Second, substitute a set of pre-determined values for a , b and $p - r$ into the explanatory variables of each regression equation, with the result that the approximate values of $f_F^z(p - r, a, b, \infty)$ for $z = \lambda, \delta, cov$ are calculated given a , b and $p - r$.

These approximate values are computed for a full system under the assumption of 2 breaks. Values for one or no breaks can be computed using formula (23), see also (3.12) and (3.13) in JMN. Thus, we approximate the moments of $\text{DF}_\ell(m - r, p - r; v)$ as follows

$$\begin{aligned} \text{Mean} &\approx \exp \{f_F^\lambda(\cdot)\} \exp \{f_F^\delta(\cdot)\} - (3 - q)(p - r), \\ \text{Variance} &\approx \exp \{f_F^\lambda(\cdot)\} [\exp \{f_F^\delta(\cdot)\}]^2 - 2(3 - q)(p - r), \end{aligned}$$

where (\cdot) stands for $(p - r, a, b, \infty)$; note that the second terms on the right-hand side vanish when $q = 3$. In turn, we calculate the shape and scale parameters:

$$f_F^{\lambda*}(\cdot) = \log \left\{ \frac{(\text{Mean})^2}{\text{Variance}} \right\} \quad \text{and} \quad f_F^{\delta*}(\cdot) = \log \left\{ \frac{\text{Variance}}{\text{Mean}} \right\}.$$

The values for $\exp \{f_F^{\lambda*}(\cdot)\}$, $\exp \{f_F^{\delta*}(\cdot)\}$ and $f_F^{cov}(\cdot)$ are then inserted into λ_F , δ_F and $\text{Cov}(T_1, T_2)$ respectively in equations (24) and (25), which yield a set of approximates to λ_P and δ_P . Finally, noting that a gamma distribution function can be specified only by shape and scale parameters, we depend on the approximates to λ_P and δ_P to identify the underlying gamma distribution, which enables us to calculate both a quantile of interest and a p -value of the observed PLR statistic.

Tables 1 and 2 display a set of examples demonstrating the accuracy of the response surface regression results. A class of approximate 95% limit quantiles is presented in each of the tables for various conceivable combinations of a , b , $p - r$ and $m - r$, when either broken-linear-trend or broken-constant specifications are adopted in analysis. Approximate quantiles in the fifth column (q_{95}) in Tables 1 and 2 are derived from Table 6 and 7, respectively; that is, they are from the full-system-based response surface analysis, combined with the mappings (24) and (25).

By contrast, approximate quantiles recorded in the sixth column (q_{95}^*) of each table, except those for $a = b = 0$, were obtained directly from auxiliary response surface regression based on partial-system simulations with the same T s and N as above. Each of auxiliary regression equations employed a simulated 95% quantile as a response variable and involved a constant, T^{-1} and its powers if necessary, as explanatory variables. The regression equations vary in specification for the purpose of capturing the underlying

smooth response surfaces of various simulated quantiles; the graph of each regression's actual and fitted values was checked to ensure the capturing of the underlying smoothness. Estimated constants in these regression equations are then recorded in the columns for q_{95}^* as approximate 95% limit quantiles. The limit quantiles in q_{95}^* for $a = b = 0$ (that is, no break cases) were taken from Doornik (2003).

Tables 1 and 2 show that the quantiles in q_{95} almost coincide with those in q_{95}^* regardless of specifications of the deterministic terms; see the seventh column of each table for $|q_{95}/q_{95}^* - 1|$, a series of absolute relative errors, all of which are very small. This correspondence can be seen as strong evidence supporting the validity of the proposed approximation method based on the full model. As a caveat, let us recall that our response surface regression was conducted by using a class of realistic number of non-stationary variables, $p - r = 1, \dots, 8$, which are presumably assumed in most applied research. Thus, an empirical study using a partial system of large dimension may require careful examinations of the underlying cointegrating rank, in addition to the application of the proposed *PLR* tests to the data under study, as discussed by Juselius (2006, §8).

4 Empirical illustration

As empirical illustration we analyse a set of quarterly time series data from Schreiber (2015), who attained an econometric system for the exchange rate and bilateral trade between the UK and Germany. She decomposed the UK-Germany economic system into two blocks, a foreign exchange block and bilateral trade block, in order to obtain a data-congruent representation useful for forecasting and policy analysis. Various econometric studies were conducted by Schreiber (2015), and one of them was the analysis of a partial model for the bilateral trade block with a structural break. The methodology developed in the above sections enables us to conduct formal tests for cointegrating rank that underlies such a partial system subject to a break. This partial system analysis may also be encouraged in terms of local power advantage of partial-system-based tests over those based on a full system under weak exogeneity, as demonstrated by Doornik, Hendry and Nielsen (1998) as well as Kurita (2011).

Figure 1 presents an overview of the quarterly data spanning the sample period of the first quarter in 1991 - the second quarter in 2014, denoted as 1991.1 - 2014.2 hereafter. The variable tb_t is the trade balance between the UK and Germany, *i.e.* the difference between the log of exports of UK goods to Germany and the log of imports of German goods to the UK; $dulc_t$ represents the unit labour cost differential between the two countries; y_t and y_t^* denote the logs of the UK and German gross domestic products, respectively; ppp_t represents the terms of trade in logarithm. See Schreiber (2015) for further details of the data. The figure indicates the presence of a structural break around 2008-2009 attributable to a global economic recession over this period.

In this empirical illustration, we analyse the data using a bivariate partial autoregressive model for tb_t and $dulc_t$, with y_t , y_t^* and ppp_t assumed to be weakly exogenous for the class of parameters of interest such as cointegrating vectors; that is, $p = 5$ and $m = 2$. This assumption is based on Schreiber's study, suggesting that modelling the bilateral trade block centering on tb_t and $dulc_t$ appears to be conformable to the underlying data structure. The lag-length $k = 2$ is selected for our bivariate partial autoregressive model.

With regard to the issue on a structural break, we adopt a broken trend specification; that is, the presence of a shift in the restricted trend as well as the unrestricted constant.

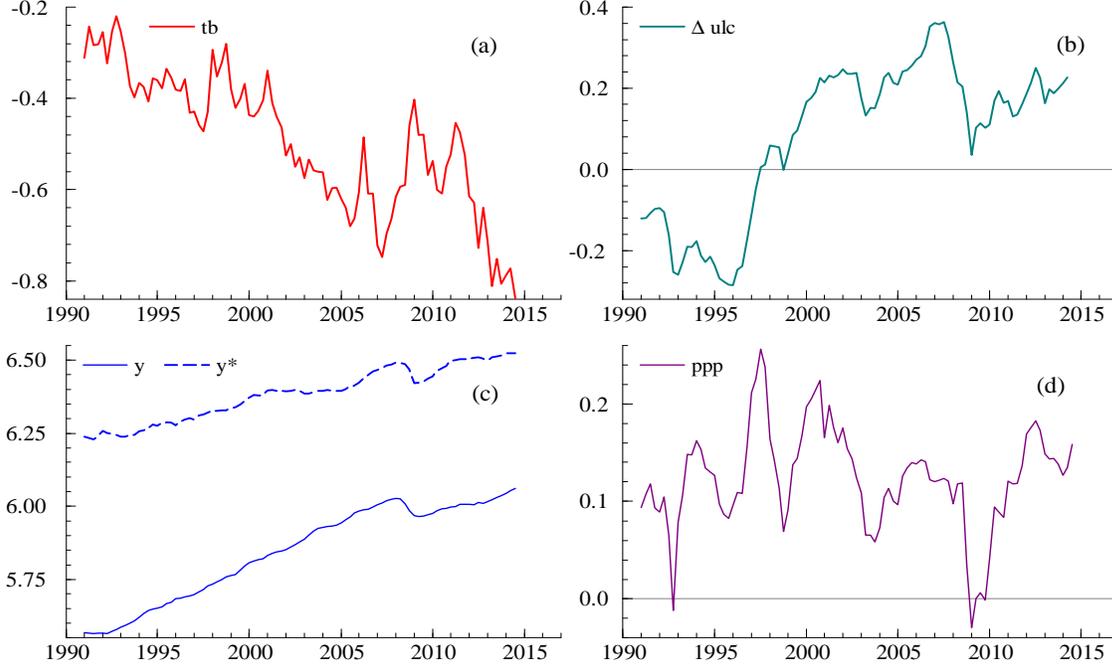


Figure 1: Data overview: tb_t , $dulc_t$, y_t , y_t^* and ppp_t

The second sub-sample period starts in 2008.3, corresponding to the observation point in T_{q-1} for $q = 2$, which results in the selection of relative break points $a = 0$ and $b = 0.255$. According to (7), our bivariate model with $k = 2$ requires a set of two impulse dummy variables for the initial values of the second sub-sample periods. In addition, a pair of impulse dummy variables, $Dp1998(1)$ and $Dp2006(2)$, is employed in our model to capture outliers in the data, as in Schreiber (2015); the former variable is 1 in 1998.1 and zero otherwise for an outlier due to the Asian financial crisis, while the latter is 1 in 2006.2 and zero otherwise, corresponding to an outlier attributable to an increase in oil prices.

A set of residual diagnostic tests for the partial system is reported in Table 3. Most of the test statistics are given in the form $F_j(df1, df2)$, which denotes an approximate F test (with relevant degrees of freedom $df1$ and $df2$) against the alternative hypothesis j . The alternative hypotheses are specified as: 5th-order serial correlation (F_{AR5} : see Godfrey 1978), 4th-order autoregressive conditional heteroscedasticity (F_{ARCH4} : see Engle, 1982), heteroscedasticity (F_{HET} : see White, 1980). Chi-squared tests for normality (χ_{ND}^2 : see Doornik and Hansen, 2008) are also recorded in the table. We also note the following

Single-eq. tests	tb_t	$dulc_t$	Vector tests	
$F_{AR5}(5,66)$	0.946[0.457]	0.777[0.570]	$F_{AR5}(20,120)$	0.542[0.943]
$F_{ARCH4}(4,84)$	0.469[0.758]	0.511[0.728]	$F_{HET}(93,162)$	0.838[0.825]
$F_{HET}(31,56)$	0.726[0.831]	1.016[0.468]	$\chi_{ND}^2(4)$	2.233[0.693]
$\chi_{ND}^2(2)$	1.341[0.512]	0.426[0.808]		

Note. Figures in square brackets are p -values.

Table 3: Diagnostic test statistics for the estimated partial system

	$r = 0$	$r \leq 1$
$PLR\{H_\ell(r) H_\ell(2)\}$	56.610[0.014]*	21.964[0.148]
95% limit quantiles	50.864	26.334

Notes. Figures in square brackets are p -values.

* denotes significance at the 5% level.

Table 4: Testing for cointegrating rank

caveats based on recent advances in the field of mis-specification tests: Nielsen (2006) demonstrated that F_{AR5} is a valid test in the presence of unit roots; Berenguer-Rico and Wilms (2018) showed that F_{HET} is valid after eliminating outliers from the observations, while χ_{ND}^2 is not necessarily valid after the removal of outliers, which was demonstrated by Berenguer-Rico and Nielsen (2017). In any case, no evidence is found in Table 3 suggesting significant mis-specification problems. We can thus judge this partial system is formulated sufficiently well to be subjected to PLR tests for cointegrating rank.

Table 4 presents a class of PLR test statistics for the determination of cointegrating rank, along with the corresponding p -values and approximate 95% limit quantiles calculated from the response surface outcomes in the previous section. We used Table 6 in Appendix A to calculate approximates to $\log \delta_F$, $\log \lambda_F$ and $\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$, and then applied them to the mappings (24) and (25) adjusted for extra χ^2 terms, so that the gamma-distribution approximation method yielded the p -values. Table 4 shows that, at the 5% level, the null hypothesis $r = 0$ is rejected while the hypothesis $r \leq 1$ fails to be rejected. Hence, this formal analysis enables us to reach the conclusion of $r = 1$, which supports the informal analysis of Schreiber (2015).

The estimated cointegrating relationship under some additional restrictions is

$$tb = \underset{(0.121)}{0.259}dulc_t - \underset{(0.3)}{0.726}ppp_t + \underset{(0.323)}{2.34}(y_t^* - y_t) - \underset{(0.004)}{0.019}t1_{(\geq 2009:1)} + v_t, \quad (26)$$

where a figure in brackets under each coefficient is a standard error and v_t represents a stationary error. The signs of the coefficients in (26) are the same as those in Schreiber (2015)'s cointegrating equation except for y_t^* . The German income y_t^* was insignificant in her cointegrating relationship and thus removed from it, while, in (26), y_t^* plays a significant role, along with y_t . As a result of checking a set of unrestricted estimates for the cointegrating vector, we have arrived at equation (26), where y_t and y_t^* are restricted such that they have the same coefficient but opposite sign, while a zero-restriction is placed on the coefficient for $t1_{(\leq 2008:2)}$; that is, a linear trend is present only in the second sub-sample period. The PLR test statistic for these restrictions is 3.571[0.168], in which the figure in square brackets is a p -value according to $\chi^2(2)$. Thus, the hypothesis of the overall restrictions cannot be rejected at the 5% level.

There are several interesting aspects of equation (26) that are worth discussing here. The real income difference between Germany and the UK, $y_t^* - y_t$, has a positive coefficient, implying that a spread in the income difference leads to an improvement in the UK trade balance with Germany. This finding is interpretable in the context of an income effect from each of the two countries. The coefficient for the terms of trade, ppp_t , should also be noted. It is negative, thus indicating a relative price effect on the trade balance in a theory-consistent manner; that is, a decrease in exports prices relative to import prices leads to trade balance improvement, so that the well-known elasticity approach to trade balance

appears to be empirically valid for the two countries. Furthermore, the linear trend t is significant solely in the second sub-sample period, suggesting long-lasting influences of the global recession on the two countries' trade balance and other economic variables.

y_t	y_t^*	ppp_t
0.004[0.951]	1.183[0.277]	1.954[0.162]

Note. Figures in square brackets are p -values.

Table 5: Checking weak exogeneity

Finally, we will check that the three variables, y_t , y_t^* and ppp_t , are indeed weakly exogenous for the class of parameters of interest. We follow the testing procedure suggested by Johansen (1992a), Boswijk (1992) and HJNR. First, the restricted cointegrating combination is added as a regressor to a marginal system (6) for $Z_t = (y_t, y_t^*, ppp_t)'$. Second, a standard regression analysis is performed to test for the significance of the cointegrating combination in each equation. Table 5 reports a class of LR test statistics for the exclusion of the empirical cointegrating linkage from each equation in the marginal system. Judging from the reported p -values according to $\chi^2(1)$, none of the test statistics indicate evidence against the assumption of weak exogeneity; thus, the preceding partial-system analysis of cointegrating rank has been justified.

5 Conclusion

This study has explored partial cointegrated vector autoregressive models subject to structural breaks in deterministic terms, a linear trend and constant. The Granger-Johansen representation of the full model in JMN has been reexamined, leading to a useful clarification of roles of the initial values in asymptotic analysis. A class of log likelihood ratio test statistics for cointegrating rank has then been introduced in the proposed partial-model framework. We have investigated asymptotic theory under a general class of innovation distributions allowing martingale difference sequences with conditional heteroscedasticity. The derived limit distributions of the statistics are closely related to those for the full models investigated by JMN. This relationship allows us to perform a response surface analysis in a simplified full-system framework, instead of relying on laborious partial-system-based simulations. The outcomes of the analysis are summarised as a set of two statistical tables providing valuable information for inference on the underlying cointegrating rank. Lastly, an empirical analysis of real-life data from the UK and Germany has demonstrated the practicality of these tables in applied economic research. As a result of this study, the partial cointegrated models have become more flexible and reliable devices for modelling time series data subject to various structural breaks.

Acknowledgement

The results reported in this paper were presented at Norwegian University of Science and Technology (NTNU), Statistics Norway and the University of Copenhagen. We would like to thank Gunnar Bårdsen, Pål Boug, Søren Johansen and various other seminar participants for helpful comments and suggestions. T. Kurita gratefully acknowledges financial support to this work from JSPS KAKENHI (Grant Number: 15KK0141).

A Tables for response surfaces

$\log \lambda_F$		$\log \delta_F$		$\text{Cov}(\mathbb{T}_1, \mathbb{T}_2)$	
<i>const.</i>	4.14	<i>const.</i>	0.5987	<i>const.</i>	-1.298
$(p-r)^{-1}$	-6.301	$p-r$	-0.0538	$1_{(2)}$	0.03616
$(p-r)^{-2}$	5.8842	a	-1.039	$1_{(4)}$	-0.027
$(p-r)^{-3}$	-2.32576	b	-0.39	$(p-r)^{-3}$	-2.022
$p-r$	0.17	$(p-r)^2$	0.00686	a	-8.689
a	2.6165	a^2	5.547	b	2.225
b	2.5245	ab	2.331	a^2	59.77
$(p-r)a$	-0.0572	b^2	1.841	ab	24.31
$(p-r)b$	-0.0971	$(p-r)^3$	-0.00033	b^2	-5.156
a^2	-7.550	a^3	-10.42	a^3	-133.5
ab	-5.323	ab^2	-4.325	ab^2	-59.05
b^2	-7.412	b^3	-2.553	$a(p-r)^{-1}$	-29.55
$(p-r)^3$	-0.000124	$a(p-r)^{-1}$	9.905	$b(p-r)^{-1}$	-66.58
$(p-r)ab$	0.161	$b(p-r)^{-1}$	1.862	$b^2(p-r)^{-1}$	255.3
$(p-r)b^2$	0.179	$a^2(p-r)^{-1}$	-61.09	$a^3(p-r)^{-1}$	280.5
a^3	10.40	$ab(p-r)^{-1}$	-17.09	$ab^2(p-r)^{-1}$	155.3
ab^2	6.096	$b^2(p-r)^{-1}$	-11.48	$b^3(p-r)^{-1}$	-240
b^3	5.851	$a^3(p-r)^{-1}$	117.68	$a(p-r)^{-2}$	21.32
$a(p-r)^{-1}$	-8.860	$ab^2(p-r)^{-1}$	35.19	$b(p-r)^{-2}$	71.68
$b(p-r)^{-1}$	-4.948	$b^3(p-r)^{-1}$	18.6	$b^2(p-r)^{-2}$	-305.7
$a^2(p-r)^{-1}$	46.15	$a(p-r)^{-2}$	-8.836	$a^2b(p-r)^{-2}$	-321.1
$ab(p-r)^{-1}$	31.85	$b(p-r)^{-2}$	1.033	$b^3(p-r)^{-2}$	332.1
$b^2(p-r)^{-1}$	26.12	$a^2(p-r)^{-2}$	66.94	$(p-r)1_{(3)}$	0.038
$a^3(p-r)^{-1}$	-86.58	$ab(p-r)^{-2}$	10.84	$b^21_{(3)}$	-0.184
$ab^2(p-r)^{-1}$	-50.50	$a^3(p-r)^{-2}$	-140.88		
$b^3(p-r)^{-1}$	-28.78	$ab^2(p-r)^{-2}$	-30.16		
$a(p-r)^{-2}$	5.296	$b^3(p-r)^{-2}$	-10.05		
$b(p-r)^{-2}$	2.386	$a1_{(1)}$	2.107		
$a^2(p-r)^{-2}$	-29.03	$b1_{(1)}$	-1.029		
$ab(p-r)^{-2}$	-19.46	$a^21_{(1)}$	-20.63		
$b^2(p-r)^{-2}$	-13.42	$b^21_{(1)}$	3.511		
$a^3(p-r)^{-2}$	62.00	$a^31_{(1)}$	45.85		
$a^2b(p-r)^{-2}$	-5.880	$ab^21_{(1)}$	4.267		
$ab^2(p-r)^{-2}$	34.59	$(p-r)b^21_{(2)}$	0.062		
$b^3(p-r)^{-2}$	15.93				

Table 6: Response surfaces for broken trend models

Note: $1_{(x)}$ is 1 when $p-r = x$ and zero otherwise.

$\log \lambda_F$		$\log \delta_F$		$\text{Cov}(\mathbf{T}_1, \mathbf{T}_2)$	
<i>const.</i>	4.95486	<i>const.</i>	0.4472	<i>const.</i>	-1.531
$(p-r)^{-1}$	-9.263	$(p-r)^{-2}$	1.17564	$(p-r)^{-1}$	0.9029
$(p-r)^{-2}$	9.162	$(p-r)^{-3}$	-1.5294	<i>a</i>	4.164
$(p-r)^{-3}$	-3.662	<i>b</i>	0.8286	$(p-r)^2$	0.01579
<i>a</i>	3.05	$(p-r)b$	-0.0646	$(p-r)b$	0.3388
<i>b</i>	0.3315	<i>ab</i>	1.75	<i>ab</i>	-27.16
$(p-r)^2$	0.01738	$(p-r)b^2$	0.04051	b^2	-14.15
$(p-r)a$	-0.128	a^3	-2.084	$(p-r)^3$	-0.0013
a^2	-14.61	ab^2	-3.698	$(p-r)^2b$	-0.0167
<i>ab</i>	-4.14	b^3	-0.788	a^3	-19.65
b^2	-2.419	$a(p-r)^{-1}$	-4.819	a^2b	14.03
$(p-r)^3$	-0.00084	$b(p-r)^{-1}$	-3.897	ab^2	42.2
$(p-r)a^2$	0.3264	$a^2(p-r)^{-1}$	30.49	b^3	17.43
$(p-r)ab$	0.1302	$ab(p-r)^{-1}$	-5.108	$a(p-r)^{-1}$	-77.72
$(p-r)b^2$	0.0266	$b^2(p-r)^{-1}$	2.273	$b(p-r)^{-1}$	-20.52
a^3	21.56	$a^3(p-r)^{-1}$	-40.9	$a^2(p-r)^{-1}$	278.7
ab^2	5.56	$ab^2(p-r)^{-1}$	13.37	$ab(p-r)^{-1}$	313.6
b^3	3.03	$a(p-r)^{-2}$	16	$b^2(p-r)^{-1}$	169.1
$a(p-r)^{-1}$	-5.742	$b(p-r)^{-2}$	3.795	$a^3(p-r)^{-1}$	-461.7
$b(p-r)^{-1}$	3.339	$a^2(p-r)^{-2}$	-110.5	$ab^2(p-r)^{-1}$	-562.9
$a^2(p-r)^{-1}$	44.2	$a^3(p-r)^{-2}$	184.8	$b^3(p-r)^{-1}$	-221.2
$ab(p-r)^{-1}$	9.66	$ab^2(p-r)^{-2}$	-4.478	$a(p-r)^{-2}$	81.64
$b^2(p-r)^{-1}$	-4.44	$(p-r)1_{(1)}$	0.5014	$a^2(p-r)^{-2}$	-315
$a^3(p-r)^{-1}$	-81.67	$a1_{(1)}$	-9.833	$ab(p-r)^{-2}$	-384.8
$ab^2(p-r)^{-1}$	-15.2	$a^21_{(1)}$	73.02	$b^2(p-r)^{-2}$	-114.6
$a(p-r)^{-2}$	2.41	$b^21_{(1)}$	-5.835	$a^3(p-r)^{-2}$	804
$b(p-r)^{-2}$	-3.44	$a^31_{(1)}$	-130.2	$a^2b(p-r)^{-2}$	-290
$a^2(p-r)^{-2}$	-24.23	$b^31_{(1)}$	4.743	$ab^2(p-r)^{-2}$	860.7
$b^2(p-r)^{-2}$	9.6	$(p-r)^2a1_{(2)}$	-0.2472	$b^3(p-r)^{-2}$	205.2
$a^3(p-r)^{-2}$	47.34	$(p-r)^2b1_{(2)}$	0.06919	$b^21_{(2)}$	0.18
$b^3(p-r)^{-2}$	-7.22	$(p-r)a^21_{(2)}$	3.765	$(p-r)^31_{(2)}$	-0.00017
		$(p-r)b^21_{(2)}$	-0.884	$(p-r)a1_{(3)}$	1.337
		$a^31_{(2)}$	-14.06	$(p-r)b1_{(3)}$	-0.0215
		$b^31_{(2)}$	1.944	$(p-r)^2a1_{(3)}$	-0.408

Table 7: Response surfaces for broken constant models

Note: $1_{(x)}$ is 1 when $p-r = x$ and zero otherwise.

B Proof of the Granger-Johansen representation

This section provides a proof of Theorem 2.2, in which the Granger-Johansen representation of the full model with deterministic breaks is presented.

Proof of Theorem 2.2. The companion form of the homogenous equation (9) for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$ is

$$\Delta \tilde{\mathbf{X}}_{t-1} = \boldsymbol{\alpha} \boldsymbol{\beta}' \tilde{\mathbf{X}}_{t-1} + \boldsymbol{\nu} \varepsilon_t.$$

As shown by Hansen (2005, Lemma A.1), $\text{rank}(\boldsymbol{\alpha}'_{\perp} \Psi \boldsymbol{\beta}_{\perp}) = p - r$ stated in Assumption 2.1 implies that the above homogenous equation is an $I(1)$ system satisfying

$$\boldsymbol{\beta}' \tilde{\mathbf{X}}_t = (I_r + \boldsymbol{\beta}' \boldsymbol{\alpha}) \boldsymbol{\beta}' \tilde{\mathbf{X}}_{t-1} + \boldsymbol{\beta}' \boldsymbol{\nu} \varepsilon_t \quad \text{with} \quad |\text{eigen}(I_r + \boldsymbol{\beta}' \boldsymbol{\alpha})| < 1,$$

which is a stable equation (see Lai and Wei, 1985).

We then follow Kurita and Nielsen (2009) in the analysis of non-stationary components. Start by the homogenous equation (9) for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$:

$$\Delta \tilde{X}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \tilde{X}_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \tilde{X}_{t-i} + \varepsilon_t.$$

Pre-multiplying the above equation by $\boldsymbol{\alpha}'_{\perp}$ and replacing $\Delta \tilde{X}_{t-i} = \Delta \tilde{X}_t - \sum_{\ell=0}^{i-1} \Delta^2 \tilde{X}_{t-\ell}$, we collect repeated terms $\Delta \tilde{X}_t$ on the left hand side to find

$$\boldsymbol{\alpha}'_{\perp} \Psi \Delta \tilde{X}_t = -\boldsymbol{\alpha}'_{\perp} \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta^2 \tilde{X}_{t-\ell} + \boldsymbol{\alpha}'_{\perp} \varepsilon_t,$$

by recalling $\Psi = I_p - \sum_{i=1}^{k-1} \Gamma_i$. Summing $\boldsymbol{\alpha}'_{\perp} \Psi \Delta \tilde{X}_s$ over $s = T_{j-1} + k + 1, \dots, t$ yields

$$\boldsymbol{\alpha}'_{\perp} \Psi \tilde{X}_t = \boldsymbol{\alpha}'_{\perp} \sum_{s=T_{j-1}+k+1}^t \varepsilon_s - \boldsymbol{\alpha}'_{\perp} \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{t-\ell} - \boldsymbol{\alpha}'_{\perp} \Psi \tilde{X}_{T_{j-1}+k} + \boldsymbol{\alpha}'_{\perp} \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{T_{j-1}+k-\ell}.$$

Apply the orthogonal projection identity $\boldsymbol{\alpha}'_{\perp} \Psi \tilde{X}_t = \boldsymbol{\alpha}'_{\perp} \Psi \boldsymbol{\beta}_{\perp} \bar{\boldsymbol{\beta}}'_{\perp} \tilde{X}_t + \boldsymbol{\alpha}'_{\perp} \Psi \bar{\boldsymbol{\beta}} \boldsymbol{\beta}' \tilde{X}_t$ to the left hand side and then pre-multiply both sides by $\boldsymbol{\beta}_{\perp} (\boldsymbol{\alpha}'_{\perp} \Psi \boldsymbol{\beta}_{\perp})^{-1}$ to find the C matrix. Shifting $C \Psi \bar{\boldsymbol{\beta}} \boldsymbol{\beta}' \tilde{X}_t$ to the right hand side, we arrive at

$$\begin{aligned} \boldsymbol{\beta}_{\perp} \bar{\boldsymbol{\beta}}'_{\perp} \tilde{X}_t &= C \sum_{s=T_{j-1}+k+1}^t \varepsilon_s - C \Psi \bar{\boldsymbol{\beta}} \boldsymbol{\beta}' \tilde{X}_t - C \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{t-\ell} \\ &\quad - C \Psi \tilde{X}_{T_{j-1}+k} + C \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{T_{j-1}+k-\ell}. \end{aligned}$$

Adding $\bar{\boldsymbol{\beta}} \boldsymbol{\beta}' \tilde{X}_t$ on both sides results in

$$\begin{aligned} \tilde{X}_t &= C \sum_{s=T_{j-1}+k+1}^t \varepsilon_s + (I - C \Psi) \bar{\boldsymbol{\beta}} \boldsymbol{\beta}' \tilde{X}_t - C \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{t-\ell} \\ &\quad - C \Psi \tilde{X}_{T_{j-1}+k} + C \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{X}_{T_{j-1}+k-\ell}. \quad (27) \end{aligned}$$

Using the notation $\Upsilon_i = -\Gamma_i - \dots - \Gamma_{k-1}$ for $1 \leq i \leq k-1$ as well as the matrix Λ defined in (10) leads to the first desired result (13).

Next, we move on to the non-homogenous formulation (8), in which μ_j and γ_j are distinct from zero. Replace X_t in (8) with $\tilde{X}_t + \tau_{c,j} + \tau_{l,j}t$ and refer to the proof of Theorem 2.1 in JMN to find

$$\Psi\tau_{l,j} = \alpha\beta'(\tau_{c,j} - \tau_{l,j}) + \mu_j \quad \text{and} \quad \beta'\tau_{l,j} + \gamma'_j = 0. \quad (28)$$

Applying (28) to (8) recovers the homogenous equation (9), so the above results derived for (9) are all applicable to (8) under (28). Substituting $\tilde{X}_t = X_t - \tau_{c,j} - \tau_{l,j}t$ into (27) yields the desired representation (14). \blacksquare

C Proofs of asymptotic results

In this section we present a high-level assumption which overrides Assumptions 3.1, 3.2 in the subsequent arguments. We then provide some specific lemmas required for proofs of the limit theorems in Section 3. Finally, we proceed to the proofs of Theorems 3.4, 3.5.

We introduce some notation. For a vector v let the outer product be $v^{\otimes 2} = vv'$. For a matrix m the spectral norm is $\|m\|^2 = \max \text{eigen}(m'm)$. Note that $\|m\|^2 \leq \text{tr}(m'm)$.

C.1 A high level assumption

In order to give proofs of the theorems introduced in this paper, we need a Law of Large Numbers for the approximately stationary components of the full model, while we require a Functional Central Limit Theorem and a convergence to a stochastic integral for the non-stationary components of the full model. We formulate these as the following high level assumption and then prove that it is satisfied under Assumptions 2.1 and 3.1.

Assumption C.1 *Let ε_t be a p -dimensional random variables and suppose that Assumption 2.1 is satisfied. Let \tilde{X}_t satisfy the homogenous equation (9) and define \mathbf{U}_{t-1} as*

$$\mathbf{U}_{t-1} = \begin{pmatrix} \beta'\tilde{X}_{t-1} \\ \Delta\tilde{X}_{t-1} \\ \vdots \\ \Delta\tilde{X}_{t-k+1} \end{pmatrix}.$$

Suppose that

$$T^{-1/2} \max_{1 \leq t \leq T} |\mathbf{U}_t| = o_{\mathbb{P}}(1) \quad (29)$$

and

$$T^{-1} \sum_{t=1}^T \begin{pmatrix} \varepsilon_t \\ \mathbf{U}_{t-1} \\ 1 \end{pmatrix}^{\otimes 2} \xrightarrow{\mathbb{P}} \begin{pmatrix} \Omega & 0 & 0 \\ 0 & \Sigma_u & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (30)$$

where Ω and Σ_u are positive definite matrices. Furthermore, let W_u be a p -dimensional Brownian motion with variance Ω . Suppose that, for $0 \leq u \leq 1$,

$$T^{-1/2} \sum_{t=1}^{\text{int}(Tu)} \varepsilon_t \xrightarrow{\mathbb{D}} W_u, \quad (31)$$

as a process on $(D[0,1])^p$ endowed with the Skorokhod metric with common distortion. Finally,

$$T^{-1} \sum_{t=1}^T \sum_{s=1}^{t-1} \varepsilon_s \varepsilon'_t \xrightarrow{D} \int_0^1 W_u dW'_u. \quad (32)$$

The next result explores the conditions of for instance Brown (1971). Subsequently we use this to show that Assumptions 2.1, 3.1 imply Assumption C.1.

Lemma C.2 *Suppose Assumption 3.1 is satisfied. Then*

- (a) $T^{-1} \sum_{t=1}^T \varepsilon_t \varepsilon'_t \xrightarrow{P} \Omega$;
- (b) $T^{-1} \sum_{t=1}^T \mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > \delta T)} | \mathcal{F}_{t-1}\} \xrightarrow{P} 0$ for all $\delta > 0$;
- (c) $T^{-1} \sum_{t=1}^T \mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > \delta T)}\} \rightarrow 0$ for all $\delta > 0$;
- (d) $\max_{1 \leq t \leq T} |\varepsilon_t^2|/T \xrightarrow{P} 0$.

Proof of Lemma C.2: (a) Since $T^{-1} \sum_{t=1}^T \mathbf{E}(\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1}) \xrightarrow{P} \Omega$ by Assumption 3.1(i) it suffices to show that $T^{-1} \sum_{t=1}^T m_t \xrightarrow{P} 0$ where $m_t = \varepsilon_t \varepsilon'_t - \mathbf{E}(\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1})$. Using the spectral norm and the Chebychev inequality

$$\mathcal{P} = \mathbf{P}(\|T^{-1} \sum_{t=1}^T m_t\| > \epsilon) \leq \frac{1}{\epsilon^2} \mathbf{E} \|T^{-1} \sum_{t=1}^T m_t\|^2 \leq \frac{1}{T^2 \epsilon^2} \text{tr} \mathbf{E} \sum_{s=1}^T \sum_{t=1}^T m_s m_t.$$

Since m_t is a martingale difference then $\mathbf{E} m_s m'_t = 0$ for $s \neq t$ while $\mathbf{E} m_t m_t \leq \mathbf{E} \varepsilon_t \varepsilon'_t \varepsilon_t \varepsilon'_t$ so that by Assumption 3.1(ii) we get

$$\mathcal{P} \leq \frac{1}{T^2 \epsilon^2} \text{tr} \mathbf{E} \sum_{t=1}^T m_t m_t \leq \frac{1}{T^2 \epsilon^2} \mathbf{E} \sum_{t=1}^T |\varepsilon_t|^4 \leq \frac{1}{T \epsilon^2} \sup_{t \in \mathbb{N}} \mathbf{E} |\varepsilon_t|^4 \leq \frac{C}{T \epsilon^2} \rightarrow 0.$$

(b, c) Brown (1971, Lemma 2) shows that the conditional Lindeberg condition (b) and the marginal Lindeberg condition (c) are equivalent under Assumption 3.1(i, ii).

First, under Assumption 3.1(iii.a), so that $\sup_{t \in \mathbb{N}} \mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > a)} | \mathcal{F}_{t-1}\} = o_{\mathbf{P}}(1)$ as $a \rightarrow \infty$. Thus, $\forall \xi > 0$, $\exists a_0$ and $\forall a \geq a_0$, it follows that

$$\mathbf{P}[\sup_{t \in \mathbb{N}} \mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > a)} | \mathcal{F}_{t-1}\} > \xi] < \xi.$$

Thus, given $\delta > 0$ and for $\forall T > a_0/\delta$, we find

$$\mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > \delta T)} | \mathcal{F}_{t-1}\} \leq \mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > a_0)} | \mathcal{F}_{t-1}\} < \xi,$$

so that the conditional Lindeberg condition (b) follows.

Second, suppose Assumption 3.1(iii.b) holds, so that $\sup_{t \in \mathbb{N}} \mathbf{E} |\varepsilon_t|^4 < \infty$. The Chebychev inequality gives

$$\mathbf{E}\{1_{(\varepsilon'_t \varepsilon_t > \delta T)}\} = \mathbf{P}(\varepsilon'_t \varepsilon_t > \delta T) \leq \frac{1}{\delta^2 T^2} \mathbf{E} |\varepsilon_t|^4.$$

Next, by the Cauchy-Schwarz inequality and Assumption 3.1(iii.b) we get

$$\mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > \delta T)}\} \leq [\mathbf{E} |\varepsilon_t|^4 \mathbf{E}\{1_{(\varepsilon'_t \varepsilon_t > \delta T)}\}]^{1/2} \leq \delta^{-1} T^{-1} \mathbf{E} |\varepsilon_t|^4 \leq \delta^{-1} T^{-1} \sup_{t \in \mathbb{N}} \mathbf{E} |\varepsilon_t|^4 \leq \delta^{-1} T^{-1} C.$$

As a consequence $T^{-1} \sum_{t=1}^T \mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(|\varepsilon'_t \varepsilon_t| > \delta T)}\} \leq \delta^{-1} T^{-1} C \rightarrow 0$ so that the marginal Lindeberg condition (c) holds.

(d) We show $\mathcal{P}_T = \mathbf{P}(\max_{1 \leq t \leq T} |\varepsilon_t|^2 > \delta T) \rightarrow 0$ for all $\delta > 0$. Note that $\mathcal{P}_T = \bigcup_{t=1}^T \mathbf{P}(|\varepsilon_t|^2 > \delta T)$. Boole's inequality gives $\mathcal{P}_T \leq \sum_{t=1}^T \mathbf{P}(|\varepsilon_t|^2 > \delta T) = \sum_{t=1}^T \mathbf{E}1_{(|\varepsilon_t|^2 > \delta T)}$. On the set $(|\varepsilon_t|^2 > \delta T)$ we get the further bound $\mathcal{P}_T \leq \delta^{-1} T^{-1} \sum_{t=1}^T \mathbf{E}|\varepsilon_t|^2 1_{(|\varepsilon_t|^2 > \delta T)}$, which vanishes by part (c). \blacksquare

We then prove that Assumption C.1 is satisfied under Assumptions 2.1 and 3.1.

Lemma C.3 *Suppose that X_t and ε_t satisfy Assumptions 2.1 and 3.1, respectively, while \tilde{X}_t solves the homogenous equation (9). Then, Assumption C.1 is satisfied.*

Proof of Lemma C.3. Note that the process \mathbf{U}_t equals $\beta' \tilde{\mathbf{X}}_t$, which is studied in Theorem 2.2. It satisfies equation (12), which is of the form $\mathbf{U}_t = \Phi \mathbf{U}_{t-1} + F \varepsilon_t$ with $\Phi = (I_r + \beta' \alpha)$ for $\mathbf{r} = r + p(k-1)$ and $F = \beta' \iota$, where Φ has spectral radius less than unity as verified in Theorem 2.2.

For (29), we apply Lemma 1 in Anderson and Kunitomo (1992). This requires that $\max_{1 \leq t \leq T} |\varepsilon_t|^2 = o_{\mathbf{P}}(T)$, which is proved in Lemma C.2 using Assumption 3.1.

For (30) using Assumption 3.1(iii.a). We apply Lemma 2 in Anderson and Kunitomo (1992) to show the convergence of the product moment matrix, which requires Assumption 3.1(ii, iii.a). Assumption 3.1 states that Ω is a positive definite matrix, which results in the positive definiteness of Σ_u by Lemma 3 in Anderson and Kunitomo (1992).

For (30) using Assumption 3.1(iii.b). We follow Lemma 2 in Anderson and Kunitomo (1992) but avoid their truncation argument. We first argue that $\sum_{t=1}^T \mathbf{U}_{t-1} F \varepsilon'_t = o_{\mathbf{P}}(T)$. Since $\mathbf{U}_{t-1} F \varepsilon'_t$ is a martingale difference sequence with second moments due to Assumption 3.1(ii) and the spectral norm is bounded by the trace, we obtain

$$\mathcal{E} = \mathbf{E} \left\| \sum_{t=1}^T \mathbf{U}_{t-1} F \varepsilon_t \right\|^2 \leq \mathbf{E} \text{tr} \sum_{s=1}^T \sum_{t=1}^T \mathbf{U}_{t-1} F \varepsilon'_t \varepsilon_s F' \mathbf{U}'_{s-1} = \mathbf{E} \text{tr} \sum_{t=1}^T \mathbf{U}_{t-1} F \varepsilon'_t \varepsilon_t F' \mathbf{U}'_{t-1}.$$

Applying iterated expectations and using that $\max_{1 \leq t \leq T} \mathbf{E}|\varepsilon_t|^2$ is bounded by Assumption 3.1(iii.b) gives $\mathcal{E} \leq C \sum_{t=1}^T \mathbf{E}|\mathbf{U}_{t-1}|^2$. Noting that $\mathbf{U}_t = \sum_{j=0}^{t-1} \Phi^j F \varepsilon_{t-j} + \Phi^t \mathbf{U}_0$ and using that ε_t is a martingale difference array, we arrive at

$$\mathbf{E} \mathbf{U}'_{t-1} \mathbf{U}_{t-1} = \mathbf{E} \sum_{j=0}^{t-2} \varepsilon'_{t-1-j} F' (\Phi^j)' \Phi^j F \varepsilon_{t-1-j} = \sum_{j=0}^{t-2} \text{tr} \{F' (\Phi^j)' \Phi^j F\} \mathbf{E} |\varepsilon_{t-1-j}|^2.$$

Using that $\max_{1 \leq t \leq T} \mathbf{E}|\varepsilon_t|^2$ is bounded and Φ has spectral radius less than unity

$$\mathbf{E} |\mathbf{U}_{t-1}|^2 \leq \sum_{j=0}^{\infty} \text{tr} \{F' (\Phi^j)' \Phi^j F\} \max_{1 \leq t \leq T} \mathbf{E} |\varepsilon_t|^2 \leq C.$$

As a consequence $\mathcal{E} = O(T)$ and by the Markov inequality $\sum_{t=1}^T \mathbf{U}_{t-1} F \varepsilon'_t = o_{\mathbf{P}}(T)$. Next, we show $T^{-1} \sum_{t=1}^T \mathbf{U}_{t-1} \mathbf{U}'_{t-1} \rightarrow \Sigma_U$ in probability. Since $\mathbf{U}_{t-1} = \Phi \mathbf{U}_{t-2} + F \varepsilon_{t-1}$ then

$$\begin{aligned} T^{-1} \sum_{t=1}^T \mathbf{U}_{t-1} \mathbf{U}'_{t-1} &= \Phi T^{-1} \sum_{t=1}^T \mathbf{U}_{t-2} \mathbf{U}'_{t-2} \Phi' + T^{-1} \sum_{t=1}^T F \varepsilon_{t-1} \varepsilon'_{t-1} F' \\ &\quad + \Phi T^{-1} \sum_{t=1}^T \mathbf{U}_{t-2} \varepsilon'_{t-1} F' + T^{-1} \sum_{t=1}^T F \varepsilon_{t-1} \mathbf{U}'_{t-2} \Phi'. \end{aligned}$$

Here, the second term converges to $F\Omega F'$ by Assumption 3.1(i) while the last two term vanishes. Since $\max_{1 \leq t \leq T} |\mathbf{U}_t| = o_P(T^{1/2})$ we therefore get

$$T^{-1} \sum_{t=1}^T \mathbf{U}_{t-1} \mathbf{U}'_{t-1} = \Phi T^{-1} \sum_{t=1}^T \mathbf{U}_{t-1} \mathbf{U}'_{t-1} \Phi' + F\Omega F' + o_P(1).$$

This is a linear equation in $T^{-1} \sum_{t=1}^T \mathbf{U}_{t-1} \mathbf{U}'_{t-1}$ so that $T^{-1} \sum_{t=1}^T \mathbf{U}_{t-1} \mathbf{U}'_{t-1} \rightarrow \Sigma_U$ in probability where Σ_U solves $\Sigma_U = \Phi \Sigma_U \Phi' + F\Omega F'$. The invertibility of Σ_U follows by Lemma 3 of Anderson and Kunitomo (1992).

For (31), the Functional Central Limit Theorem follows from Brown (1971, Theorem 3), equipped with Cramér-Wold device (see Billingsley 1968, Theorem 7.7). Brown's result applies under Assumption 3.1(i, ii) and either of the Lindeberg condition established in Lemma C.2(b, c) under Assumption 3.1.

For (32), the convergence to a stochastic integral for the univariate case is based on the results of Jakubowki, Ménin and Pages (1989), which was referred to by Kurtz and Protter (1996), while the convergence to a stochastic integral for the multivariate case is based on the results of Kurtz and Protter (1991). For the univariate case, Kurtz and Protter (1996, Theorem 7.1) show that we need to check that the martingale array $M_t^T = T^{-1/2} \sum_{s=1}^t \varepsilon_s$ is uniformly tight, as required by Jakubowski, Ménin and Pages (1989) or, equivalently, it has uniformly controlled variations. We use Theorem 2.2 in Kurtz and Protter (1991), which applies to the multivariate case; see also Hansen (1992, Theorem 2.1). Choose $\delta = \infty$ so that $M_t^{T, \delta} = M_t^T$ in Kurtz and Protter's notation. For each $\alpha > 0$, $T \geq 1$, choose stopping times $\tau^{T, \alpha} = \infty$ so that $\mathbf{P}(\tau^{T, \alpha} \leq \alpha) = 0 \leq 1/\alpha$. Then, we obtain a quadratic variation processes

$$[M^{T, \delta}]_{t \wedge \tau^{T, \alpha}} = T^{-1} \sum_{s=1}^{t \wedge \tau^{T, \alpha}} \varepsilon_s \varepsilon'_s \leq T^{-1} \sum_{t=1}^T \varepsilon_t \varepsilon'_t = [M^{T, \delta}]_T,$$

so that $\mathbf{E}[M^{T, \delta}]_{t \wedge \tau^{T, \alpha}} \leq \mathbf{E}[M^{T, \delta}]_T$. From Assumption 3.1(i) it follows that $\mathbf{E}[M^{T, \delta}]_T \rightarrow \Omega$. Consequently, we have $\sup_T \|\mathbf{E}[M^{T, \delta}]_T\| < \infty$. In turn $\sup_T \|\mathbf{E}[M^{T, \delta}]_{t \wedge \tau^{T, \alpha}}\| < \infty$ for each t , so that M_t^T has uniformly controlled variations. ■

Proof of Lemma 3.3. Under Assumption 3.2, we have $\mathbf{E}(\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1}) = \Omega$ *a.s.*, so that $T^{-1} \sum_{t=1}^T \mathbf{E}(\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1}) = \Omega$ *a.s.* follows and Assumption 3.1(ii) holds. Taking iterated expectations, we obtain $\mathbf{E}(\varepsilon_t \varepsilon'_t) = \Omega$, which leads to $T^{-1} \sum_{t=1}^T \mathbf{E}(\varepsilon_t \varepsilon'_t) = \Omega$ and thus Assumption 3.1(i) is satisfied. Lastly, we show that Assumption 3.1(iii.a) is implied by Assumption 3.2(ii). Using Hölder's inequality, we find, for $\eta = \xi/2 > 0$,

$$\mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > a)} | \mathcal{F}_{t-1}\} \leq [\mathbf{E}(|\varepsilon_t|^{2+2\eta} | \mathcal{F}_{t-1})]^{1/(1+\eta)} [\mathbf{E}\{1_{(\varepsilon'_t \varepsilon_t > a)}^{(1+\eta)/\eta} | \mathcal{F}_{t-1}\}]^{\eta/(1+\eta)},$$

in which we note the equality $1_{(\varepsilon'_t \varepsilon_t > a)}^{(1+\eta)/\eta} = 1_{(\varepsilon'_t \varepsilon_t > a)}$. Hence, writing the expectation of the indicator as a probability, and also using Markov's inequality, we arrive at

$$\mathbf{E}\{1_{(\varepsilon'_t \varepsilon_t > a)}^{(1+\eta)/\eta} | \mathcal{F}_{t-1}\} = \mathbf{P}(\varepsilon'_t \varepsilon_t > a | \mathcal{F}_{t-1}) \leq \frac{1}{a^{1+\eta}} \mathbf{E}\{(\varepsilon'_t \varepsilon_t)^{1+\eta} | \mathcal{F}_{t-1}\}.$$

In combination we obtain $\mathbf{E}\{\varepsilon'_t \varepsilon_t 1_{(\varepsilon'_t \varepsilon_t > a)} | \mathcal{F}_{t-1}\} \leq a^{-\eta} \mathbf{E}(|\varepsilon_t|^{2+2\eta} | \mathcal{F}_{t-1})$, which vanishes as $a \rightarrow \infty$ uniformly in t , since the $\mathbf{E}(|\varepsilon_t|^{2+2\eta} | \mathcal{F}_{t-1})$ is uniformly bounded by assumption. ■

Remark C.4 We give an example of a martingale difference sequence $(\varepsilon_t, \mathcal{F}_t)$ satisfying $\sup_{t \in \mathbb{N}} \mathbf{E}|\varepsilon_t|^4 < \infty$ but violating Assumption 3.1(iii.a), i.e. $\sup_{t \in \mathbb{N}} \mathbf{E}\{|\varepsilon_t|^2 1_{(|\varepsilon_t|^2 > a)} | \mathcal{F}_{t-1}\} \rightarrow 0$ in probability, which is from Anderson and Kunitomo (1992). Consider the probability space $\{(0, 1], \mathcal{F}, \mathbf{P}\}$, where \mathcal{F} is the Borel field on $(0, 1]$ and \mathbf{P} is the uniform distribution. Consider also a dyadic sequence with indices $n = 1, 2, \dots$ and $k_n = 1, \dots, 2^n$, so that $t = \sum_{j=1}^{n-1} 2^j + k_n$, and define

$$\varepsilon_t(\omega) = \varepsilon_{nk_n}(\omega) = \begin{cases} n & \text{if } 2k - 1 < 2^n \omega \leq 2k, \\ -n & \text{if } 2k - 2 < 2^n \omega \leq 2k + 1, \\ 0 & \text{otherwise.} \end{cases}$$

We note that $\mathbf{E}\varepsilon_t = 0$ while $\mathbf{E}\varepsilon_t^4 = n^4/2^n$ is uniformly bounded in n . The natural filtration of ε_t is then given by the σ -fields $\mathcal{F}_0 = \sigma\{[0, 1]\}$, $\mathcal{F}_1 = \sigma\{(0, \frac{1}{4}], (\frac{1}{4}, \frac{1}{2}], \mathcal{F}_0\}$, $\mathcal{F}_2 = \sigma\{(\frac{1}{2}, \frac{3}{4}], (\frac{3}{4}, 1], \mathcal{F}_1\}$, $\mathcal{F}_3 = \sigma\{(0, \frac{1}{8}], (\frac{1}{8}, \frac{1}{4}], \mathcal{F}_2\}$ and so on. We find that $\mathbf{E}(\varepsilon_t | \mathcal{F}_{t-1}) = 0$ while $[\sup_{t \in \mathbb{N}} \mathbf{E}\{|\varepsilon_t|^2 1_{(|\varepsilon_t|^2 > a)} | \mathcal{F}_{t-1}\}](\omega) = \infty$ for all a and all ω , so that the random variables $\sup_{t \in \mathbb{N}} \mathbf{E}\{|\varepsilon_t|^2 1_{(|\varepsilon_t|^2 > a)} | \mathcal{F}_{t-1}\}$ cannot vanish in probability.

Remark C.5 A related cointegration model allowing for heteroscedasticity was considered by Cavaliere, Rahbek and Taylor (2010). It was assumed that $T^{-1} \sum_{t=1}^T \mathbf{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) \rightarrow \Omega$ in probability (Assumption 3.1(ii)) and $\sup_{t \in \mathbb{N}} \mathbf{E}|\varepsilon_t|^4 < \infty$, and the results of Hannan and Heyde (1972) were used for asymptotic arguments. These assumptions appear to be insufficient in that Hannan and Heyde (1972) require that Assumption 3.1(ii) should hold almost surely. In addition, it is necessary to ensure the invertibility of Σ_u , see (30), which is guaranteed under Assumption 3.1, as demonstrated in the proof of Lemma C.3 above.

C.2 Several lemmas for the partial systems

The asymptotic properties of the product moment matrices, S_{ij} for $i, j = 0, 1$ defined in (19), are investigated so as to adapt Lemmas 10.1 and 10.3 of Johansen (1995) to the present model. We do this by combining various ideas and techniques from HJNR, JMN and Kurita and Nielsen (2009). These papers assume normal innovations, which we have generalised as Assumptions 3.2 and 3.1 in our study. This means that we have to be careful when defining the limits of product moments of various non-integrated components. This issue is addressed in the following lemma:

Lemma C.6 Suppose that Assumptions 2.1 and C.1 are satisfied. Let

$$\mathbf{V}_t = \begin{pmatrix} \varepsilon_t \\ \Delta X_t \\ \beta' X_{t-1} + \gamma_j t \end{pmatrix}, \quad Q_t = \begin{pmatrix} \Delta X_{t-1} \\ \vdots \\ \Delta X_{t-k+1} \end{pmatrix} \quad \text{and} \quad \bar{Q}_t = \begin{pmatrix} Q_t \\ 1 \end{pmatrix}.$$

Let $v_j = T_j/T$ be relative break points for $j = 0, \dots, q$ and define the sample product moment matrix of \mathbf{V}_t corrected for Q_t and a constant as

$$M_{\mathbf{V}\mathbf{V}, \mathbf{Q}, 1} = \sum_{j=1}^q \frac{1}{T} \sum_{t=T_{j-1}+k}^{T_j} \left\{ \mathbf{V}_t - \sum_{s=T_{j-1}+k}^{T_j} \mathbf{V}_t \bar{Q}_s' \left(\sum_{s=T_{j-1}+k}^{T_j} \bar{Q}_s^{\otimes 2} \right)^{-1} \bar{Q}_t \right\}^{\otimes 2}.$$

Then, as $T \rightarrow \infty$ with fixed relative break points v_j we get that $M_{\mathbf{V}\mathbf{V}.Q,1}$ converges in probability to a positive definite matrix with the structure

$$\begin{pmatrix} \Omega & \Omega & 0 \\ \Omega & \Sigma_{xx} & \Sigma_{x\beta} \\ 0 & \Sigma_{\beta x} & \Sigma_{\beta\beta} \end{pmatrix}, \quad (33)$$

where $\Sigma_{x\beta} = \alpha\Sigma_{\beta\beta}$ and $\Sigma_{xx} = \alpha\Sigma_{\beta x} + \Omega$ hold.

Proof of Lemma C.6. We start with the homogenous equation (12). For $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$, this equation can always be solved as

$$\beta' \tilde{\mathbf{X}}_t = \sum_{s=1}^{t-T_{j-1}-k} (I_{\mathbf{r}} + \beta' \alpha)^{t-T_{j-1}-k-s} \beta' \boldsymbol{\nu}_{\varepsilon T_{j-1}+k+s} + (I_{\mathbf{r}} + \beta' \alpha)^{t-T_{j-1}-k} \beta' \tilde{\mathbf{X}}_{T_{j-1}+k},$$

and, in the first sub-sample period or $j = 1$, the initial value $\beta' \tilde{\mathbf{X}}_{T_0+k} = \beta' \tilde{\mathbf{X}}_k$ can be treated as fixed, so that the process $\beta' \tilde{\mathbf{X}}_t$ for $T_0 + k < t \leq T_1$ becomes uniformly bounded in probability by noting that it equals \mathbf{U}_t in Assumption C.1. Similarly, by iterating over all the other start-up values, the process $\beta' \tilde{\mathbf{X}}_t$ for $T_{j-1} + k < t \leq T_j$ and $j = 2, \dots, q$ is also uniformly bounded in probability. Since the number of breaks is finite, $\beta' \tilde{\mathbf{X}}_t$ is uniformly bounded in probability jointly for $1 \leq j \leq q$.

Next, the Granger-Johansen representation (14) implies that, for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$,

$$\Delta X_t = C\varepsilon_t + \Delta U_t + \tau_{\ell,j} \quad \text{and} \quad \beta' X_t + \gamma_j t = \beta' U_t + \beta' \tau_{c,j},$$

where

$$U_t = (I - C\Psi)\bar{\beta}' \tilde{\mathbf{X}}_t - C \sum_{i=1}^{k-1} \Gamma_i \sum_{\ell=0}^{i-1} \Delta \tilde{\mathbf{X}}_{t-\ell}.$$

Since $\beta' \tilde{\mathbf{X}}_t$ is uniformly bounded in probability, it follows that U_t is also uniformly bounded in probability. Note that U_t is identical throughout all sub-sample periods. The intercepts $\tau_{\ell,j}$ and $\beta' \tau_{c,j}$ are eliminated from ΔX_t and $\beta' X_t + \gamma_j t$ respectively, when demeaning them within each sub-sample period. Consequently, we can apply the Law of Large Numbers (30) in Assumption C.1 to

$$\frac{1}{T_j - T_{j-1} - k} \sum_{t=T_{j-1}+k}^{T_j} \left\{ \begin{pmatrix} \mathbf{V}_t \\ Q_t \end{pmatrix} - \frac{1}{T_j - T_{j-1} - k} \sum_{s=T_{j-1}+k}^{T_j} \begin{pmatrix} \mathbf{V}_s \\ Q_s \end{pmatrix} \right\}^{\otimes 2}, \quad (34)$$

which, for $1 \leq j \leq q$, converges in probability to a positive definite matrix denoted as

$$\begin{pmatrix} N_{\varepsilon\varepsilon}^{(j)} & N_{\varepsilon x}^{(j)} & N_{\varepsilon\beta}^{(j)} & N_{\varepsilon q}^{(j)} \\ N_{x\varepsilon}^{(j)} & N_{xx}^{(j)} & N_{x\beta}^{(j)} & N_{xq}^{(j)} \\ N_{\beta\varepsilon}^{(j)} & N_{\beta x}^{(j)} & N_{\beta\beta}^{(j)} & N_{\beta q}^{(j)} \\ N_{q\varepsilon}^{(j)} & N_{qx}^{(j)} & N_{q\beta}^{(j)} & N_{qq}^{(j)} \end{pmatrix},$$

for $N_{\varepsilon\varepsilon}^{(j)} = \Omega$ by Assumption C.1. Since both X_{t-1} and Q_t consist of the past values of ε_t , it follows from (30) that $N_{\varepsilon\beta}^{(j)} = 0$ and $N_{\varepsilon q}^{(j)} = 0$ hold. Note that the model equation is

$$\varepsilon_t = \Delta X_t - \alpha(\beta' X_{t-1} + \gamma_j t) - \mathbf{\Gamma} Q_t + \mu_j, \quad (35)$$

for $\mathbf{\Gamma} = (\Gamma_1, \dots, \Gamma_{k-1})$. We can derive three properties from this equation. First, we post-multiply (35) by ε'_t and then exploit $N_{\varepsilon\beta}^{(j)} = 0$ and $N_{\varepsilon q}^{(j)} = 0$ to find

$$\Omega = N_{\varepsilon\varepsilon}^{(j)} = N_{x\varepsilon}^{(j)} - 0 - 0 = N_{x\varepsilon}^{(j)},$$

which is the first property. The next one is

$$(N_{x\beta}^{(j)}, N_{xq}^{(j)}) \begin{pmatrix} N_{\beta\beta}^{(j)} & N_{\beta q}^{(j)} \\ N_{q\beta}^{(j)} & N_{qq}^{(j)} \end{pmatrix}^{-1} = (\alpha, \mathbf{\Gamma}), \quad (36)$$

where the left hand side is the limit of the sample regression coefficient for ΔX_t regressed on $\beta' X_{t-1} + \gamma_j t$, Q_t and an intercept. This property is demonstrated by substituting $\varepsilon_t + \alpha(\beta' X_{t-1} + \gamma_j t) + \mathbf{\Gamma} Q_t - \mu_j$ from (35) into ΔX_t ; we then arrive at the limit result

$$(N_{x\beta}^{(j)}, N_{xq}^{(j)}) = (N_{\varepsilon\beta}^{(j)}, N_{\varepsilon q}^{(j)}) + (\alpha, \mathbf{\Gamma}) \begin{pmatrix} N_{\beta\beta}^{(j)} & N_{\beta q}^{(j)} \\ N_{q\beta}^{(j)} & N_{qq}^{(j)} \end{pmatrix},$$

from which (36) follows by noting that $N_{\varepsilon\beta}^{(j)} = 0$ and $N_{\varepsilon q}^{(j)} = 0$. The third property is

$$\Omega = N_{xx}^{(j)} - (\alpha, \mathbf{\Gamma}) \begin{pmatrix} N_{\beta\beta}^{(j)} & N_{\beta q}^{(j)} \\ N_{q\beta}^{(j)} & N_{qq}^{(j)} \end{pmatrix} \begin{pmatrix} \alpha' \\ \mathbf{\Gamma}' \end{pmatrix}. \quad (37)$$

The left hand side of (37) is the limit of the sample product moment of ε_t regressed on $\beta' X_{t-1} + \gamma_j t$, Q_t and an intercept, due to $N_{\varepsilon\beta}^{(j)} = 0$ and $N_{\varepsilon q}^{(j)} = 0$. The right hand side of (37) is the limit of the sample product moment of ΔX_t regressed on $\beta' X_{t-1} + \gamma_j t$, Q_t and an intercept, where we have exploited the identity (36).

Now, we return to (34) and partial out Q_t to obtain

$$\begin{aligned} & \frac{1}{T_j - T_{j-1} - k} \sum_{t=T_{j-1}+k}^{T_j} \left\{ \mathbf{v}_t - \sum_{s=T_{j-1}+k}^{T_j} \mathbf{v}_s \bar{Q}'_s \left(\sum_{s=T_{j-1}+k}^{T_j} \bar{Q}_s^{\otimes 2} \right)^{-1} \bar{Q}_t \right\}^{\otimes 2} \\ & \xrightarrow{\text{P}} \begin{pmatrix} N_{\varepsilon\varepsilon}^{(j)} & N_{\varepsilon x}^{(j)} & N_{\varepsilon\beta}^{(j)} \\ N_{x\varepsilon}^{(j)} & N_{xx}^{(j)} & N_{x\beta}^{(j)} \\ N_{\beta\varepsilon}^{(j)} & N_{\beta x}^{(j)} & N_{\beta\beta}^{(j)} \end{pmatrix} - \begin{pmatrix} N_{\varepsilon q}^{(j)} \\ N_{xq}^{(j)} \\ N_{\beta q}^{(j)} \end{pmatrix} (N_{qq}^{(j)})^{-1} (N_{q\varepsilon}^{(j)}, N_{qx}^{(j)}, N_{q\beta}^{(j)}), \quad (38) \end{aligned}$$

for which it should be noted that $N_{\varepsilon\varepsilon}^{(j)} = N_{\varepsilon x}^{(j)} = \Omega$ while $N_{\varepsilon\beta}^{(j)} = 0$ and $N_{\varepsilon q}^{(j)} = 0$. Thus, (38) is reduced to

$$\begin{pmatrix} \Omega & \Omega & 0 \\ \Omega & \Sigma_{xx}^{(j)} & \Sigma_{x\beta}^{(j)} \\ 0 & \Sigma_{\beta x}^{(j)} & \Sigma_{\beta\beta}^{(j)} \end{pmatrix},$$

where

$$\begin{pmatrix} \Sigma_{xx}^{(j)} & \Sigma_{x\beta}^{(j)} \\ \Sigma_{\beta x}^{(j)} & \Sigma_{\beta\beta}^{(j)} \end{pmatrix} = \begin{pmatrix} N_{xx}^{(j)} & N_{x\beta}^{(j)} \\ N_{\beta x}^{(j)} & N_{\beta\beta}^{(j)} \end{pmatrix} - \begin{pmatrix} N_{xq}^{(j)} \\ N_{\beta q}^{(j)} \end{pmatrix} (N_{qq}^{(j)})^{-1} (N_{qx}^{(j)}, N_{q\beta}^{(j)}).$$

Furthermore, noting $\sum_{j=1}^q \Delta v_j = 1$ and $\sum_{j=1}^q (\Delta v_j) \Omega = \Omega$, we define

$$\Sigma_{ik} = \sum_{j=1}^q (\Delta v_j) \Sigma_{ik}^{(j)} \quad \text{and} \quad N_{lm} = \sum_{j=1}^q \Delta v_j N_{lm}^{(j)}, \quad (39)$$

for $i, k = x, \beta$ and $l, m = q, x, \beta$. The use of Slutsky's theorem then leads to (33).

It is left to show that $\Sigma_{x\beta} = \alpha \Sigma_{\beta\beta}$ and $\Sigma_{xx} = \alpha \Sigma_{\beta x} + \Omega$. For the first expression, we apply the identities in (39) to (36) so as to obtain

$$(N_{x\beta}, N_{xq}) = (\alpha, \mathbf{\Gamma}) \begin{pmatrix} N_{\beta\beta} & N_{\beta q} \\ N_{q\beta} & N_{qq} \end{pmatrix}. \quad (40)$$

Taking partitioned inversion in (40) results in $\alpha = N_{x\beta \cdot q} N_{\beta\beta \cdot q}^{-1} = \Sigma_{x\beta} \Sigma_{\beta\beta}^{-1}$. For the second expression, we apply the identities in (39) to (37) to find

$$\Omega = N_{xx} - (\alpha, \mathbf{\Gamma}) \begin{pmatrix} N_{\beta\beta} & N_{\beta q} \\ N_{q\beta} & N_{qq} \end{pmatrix} \begin{pmatrix} \alpha' \\ \mathbf{\Gamma}' \end{pmatrix}. \quad (41)$$

Inserting (40) into (41) and taking its partitioned inversion, we arrive at

$$\Omega = N_{xx} - (N_{x\beta}, N_{xq}) \begin{pmatrix} N_{\beta\beta} & N_{\beta q} \\ N_{q\beta} & N_{qq} \end{pmatrix}^{-1} \begin{pmatrix} N_{\beta x} \\ N_{q x} \end{pmatrix} = N_{xx \cdot q} - N_{x\beta \cdot q} N_{\beta\beta \cdot q}^{-1} N_{\beta x \cdot q} = \Sigma_{xx} - \alpha \Sigma_{\beta x},$$

by noting from (40) that $N_{x\beta \cdot q} N_{\beta\beta \cdot q}^{-1} = \alpha$ holds. \blacksquare

Recalling the decomposition $\Delta X_t = (\Delta Y_t', \Delta Z_t')'$, we find the following equivalence in the lower-right submatrix of (33) in Lemma C.6:

$$\begin{pmatrix} \Sigma_{xx} & \Sigma_{x\beta} \\ \Sigma_{\beta x} & \Sigma_{\beta\beta} \end{pmatrix} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yz} & \Sigma_{y\beta} \\ \Sigma_{zy} & \Sigma_{zz} & \Sigma_{z\beta} \\ \Sigma_{\beta y} & \Sigma_{\beta z} & \Sigma_{\beta\beta} \end{pmatrix} = \Sigma.$$

Under the normality assumption for ε_t as in HJNR, we could form the conditional variance of the two elements ΔY_t and $\beta' X_{t-1} + t\gamma_j$ given the element ΔZ_t . Moving away from normality under Assumptions 3.2 and 3.1, we need to consider instead the limit of a product moment matrix consisting of *linear combinations* of these elements, defined in the following manner:

$$\begin{pmatrix} \Sigma_{yy \cdot z} & \Sigma_{y\beta \cdot z} \\ \Sigma_{\beta y \cdot z} & \Sigma_{\beta\beta \cdot z} \end{pmatrix} = A' \Sigma A = \begin{pmatrix} \Sigma_{yy} & \Sigma_{y\beta} \\ \Sigma_{\beta y} & \Sigma_{\beta\beta} \end{pmatrix} - \begin{pmatrix} \Sigma_{yz} \\ \Sigma_{\beta z} \end{pmatrix} \Sigma_{zz}^{-1} (\Sigma_{zy}, \Sigma_{z\beta}), \quad (42)$$

which appears in (43) in Lemma C.9, and where

$$A' = \begin{pmatrix} I & -\Sigma_{yz} \Sigma_{zz}^{-1} & 0 \\ 0 & -\Sigma_{\beta z} \Sigma_{zz}^{-1} & I \end{pmatrix}.$$

Let us recall the weak exogeneity condition $\alpha_z = 0$, which implies

$$\alpha = \begin{pmatrix} \alpha_y \\ 0 \end{pmatrix} \quad \text{and} \quad \alpha_{\perp} = \begin{pmatrix} \alpha_{y\perp} & 0 \\ 0 & I_{p-m} \end{pmatrix}.$$

Finally, recall from (3) that the limit variance of innovations in the partial equation equals $\Omega_{yy \cdot z} = \Omega_{yy} - \Omega_{yz} \Omega_{zz}^{-1} \Omega_{zy}$ under Assumption C.1. Within each sub-sample period the setup here is identical to that of HJNR. We therefore obtain the following equation, which adapts equation 10.6 in Johansen (1995, Lemma 10.1).

Lemma C.7 (HJNR, Lemma 4) *Suppose that Assumptions 2.1 and C.1 are satisfied under $\alpha_z = 0$. It then follows that*

$$\alpha_{y\perp}(\alpha'_{y\perp}\Omega_{yy\cdot z}\alpha_{y\perp})^{-1}\alpha'_{y\perp} = \Sigma_{yy\cdot z}^{-1} - \Sigma_{yy\cdot z}^{-1}\Sigma_{y\beta\cdot z}(\Sigma_{\beta y\cdot z}\Sigma_{yy\cdot z}^{-1}\Sigma_{y\beta\cdot z})^{-1}\Sigma_{\beta y\cdot z}\Sigma_{yy\cdot z}^{-1}.$$

We now explore the limit of the common trends within each sub-sample period. Define

$$B_T^{*'} = \begin{pmatrix} \alpha'_{\perp}\Gamma & 0 \\ 0 & T^{-1/2} \end{pmatrix} \begin{pmatrix} I_p & -\tau_{\ell,j} \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad X_t^* = \begin{pmatrix} X_{t-1} \\ t \end{pmatrix},$$

and the next lemma is a combination of Lemma A.1 in JMN and Lemma 5 in HJNR.

Lemma C.8 *Suppose that Assumptions 2.1 and C.1 are satisfied. Consider the $(p-r+1)$ -dimensional process $T^{-1/2}B_T^{*'}X_{\text{int}(Tu)}^*$ on $D[0,1]$ endowed with the Skorokhod metric with common distortion across the dimensions. Let W_u be a $(p-r)$ -dimensional Brownian motion with variance Ω for $0 \leq u \leq 1$. For $v_{j-1} \leq u < v_j$ and $1 \leq j \leq q$, the process $X_{\text{int}(Tu)}^*$ satisfies*

$$T^{-1/2}B_T^{*'}(X_{\text{int}(Tu)}^* - X_{\text{int}(Tv_{j-1})}^*) \xrightarrow{D} \left\{ \begin{array}{c} \alpha'_{\perp}(W_u - W_{v_{j-1}}) \\ u - v_{j-1} \end{array} \right\}.$$

The convergence holds jointly for $1 \leq j \leq q$ and $0 \leq u \leq 1$.

Proof of Lemma C.8. The Granger-Johansen representation (14) implies that, for $1 \leq j \leq q$ and $T_{j-1} + k < t \leq T_j$,

$$T^{-1/2}\alpha'_{\perp}\Gamma(X_t - \tau_{\ell,j}t) \approx T^{-1/2}\alpha'_{\perp}\Gamma(C \sum_{s=T_{j-1}+k+1}^t \varepsilon_s + U_t + \tau_{c,j}).$$

Since $\alpha'_{\perp}\Gamma C = \alpha'_{\perp}$ has full row rank and U_t is bounded in probability as shown in the proof of Lemma C.6, the random walk component $\alpha'_{\perp}\sum_{s=T_{j-1}+k+1}^t \varepsilon_s$ dominates $\alpha'_{\perp}\Gamma U_t$. The initial value $\tau_{c,j}$ could be large when $j > 1$, but it is eliminated when taking differences $X_{\text{int}(Tu)}^* - X_{\text{int}(Tv_{j-1})}^*$. Thus, the first element of $T^{-1/2}B_T^{*'}(X_{\text{int}(Tu)}^* - X_{\text{int}(Tv_{j-1})}^*)$ converges to $\alpha'_{\perp}(W_u - W_{v_{j-1}})$ by the Functional Central Limit Theorem (31) in Assumption C.1. The second element also converges as desired since $T^{-1}\text{int}(Tu)$ converges to u . As the number of breaks is finite, the convergence holds jointly for $1 \leq j \leq q$. \blacksquare

Decompose $W_u = (W'_{1u}, W'_{2u})'$, in which the dimensions of W_{1u} and W_{2u} are m and $p-m$, respectively. Let us recall the notation $X_{t-1}^{\circ} = (X'_{t-1}, tE'_t)'$, which was used in reduced rank regression in Section 3.1. The next lemma establishes the asymptotic theory for the product moment matrices S_{ij} for $i, j = 0, 1$ defined in (19).

Lemma C.9 *Suppose that Assumptions 2.1 and C.1 are satisfied under $\alpha_z = 0$. Define $\beta^{\circ} = (\beta', \gamma)'$, $\tau_{\ell} = (\tau_{\ell,1}, \dots, \tau_{\ell,q}) \in \mathbf{R}^{p \times q}$ and*

$$B_T^{\circ'} = \begin{pmatrix} \alpha'_{\perp}\Gamma & 0 \\ 0 & T^{-1/2}I_q \end{pmatrix} \begin{pmatrix} I_p & -\tau_{\ell} \\ 0 & I_q \end{pmatrix}.$$

It then follows that

$$\begin{pmatrix} S_{00} & S_{01}\beta^\circ \\ \beta^{\circ\prime}S_{10} & \beta^{\circ\prime}S_{11}\beta^\circ \end{pmatrix} \xrightarrow{\text{P}} \begin{pmatrix} \Sigma_{yy\cdot z} & \Sigma_{y\beta\cdot z} \\ \Sigma_{\beta y\cdot z} & \Sigma_{\beta\beta\cdot z} \end{pmatrix}, \quad (43)$$

$$T^{-1}B_T^{\circ\prime}S_{11}B_T^\circ \xrightarrow{\text{D}} \int_0^1 F_u F_u' du, \quad (44)$$

$$B_T^{\circ\prime} (S_{10} - S_{11}\beta^\circ \alpha_y') \xrightarrow{\text{D}} \int_0^1 F_u d(W_{1u} - \omega W_{2u}), \quad (45)$$

$$B_T^{\circ\prime}S_{11}\beta^\circ = \text{O}_P(1), \quad (46)$$

where

$$F_u = \begin{pmatrix} \alpha_\perp' W_u \\ u e_u \end{pmatrix} - \int_0^1 \begin{pmatrix} \alpha_\perp' W_s \\ s e_s \end{pmatrix} e_s' ds \left(\int_0^1 e_s e_s' ds \right)^{-1} e_u.$$

Proof of Lemma C.9. Recall the decomposition $\Delta X_t = (\Delta Y_t', \Delta Z_t)'$.

For (43), we start with the left hand side of (38) and further partial out ΔZ_t from the process, to which we then apply the Law of Large Numbers (30) in Assumption C.1. Follow the proof of Lemma C.6 afterwards, supplemented with the definition of the limit expression (42), in order to verify (43).

For (44) we combine Lemma C.8, the continuous mapping theorem and Johansen (1995, Lemma 10.3).

For (45) we note $B_T^{\circ\prime} (S_{10} - S_{11}\beta^\circ \alpha_y') = B_T^{\circ\prime} S_{1\varepsilon}$. The Law of Large Numbers (30) in Assumption C.1 implies $B_T^{\circ\prime} S_{1\varepsilon} = B_T^{\circ\prime} (T - k)^{-1} \sum_{t=k+1}^T \mathcal{Z}_{1,t-1} \varepsilon'_{y\cdot z,t} + \text{o}_P(1)$, in which $\mathcal{Z}_{1,t}$ is the demeaned version of X_t° as defined in (18). By (29) and the Granger-Johansen representation in Theorem 2.2, we can replace $B_T^{\circ\prime} \mathcal{Z}_{1,t}$ with the demeaned version of $(\sum_{s=k+1}^t \varepsilon_s' \alpha_\perp, tE_t')$. The stochastic integral (32) in Assumption C.1 then gives (45).

For (46) we follow the strategy used for (45); see also Johansen (1995, Lemma 10.3). ■

C.3 Proofs of the theorems in Section 3

Proof of Theorem 3.4. Follow the proof of Theorem 11.1 in Johansen (1995) by using Lemmas C.7 and C.9 given above instead of his Lemmas 10.1 and 10.3, and also utilise invariance properties with respect to non-singular linear transformations as in the proof of Theorem 1 in HJNR. ■

Proof of Theorem 3.5. The proof presented here is based on Doornik (1998, Sec. 9). The asymptotic distribution of the LR test statistic in the partial model in Theorem 3.4 is rewritten as

$$\begin{aligned} \text{tr} \left\{ \int_0^1 dB_u^{(m-r)} G_u' \left(\int_0^1 G_u G_u' du \right)^{-1} \int_0^1 G_u dB_u^{(m-r)'} \right\} \\ = \sum_{i=1}^{m-r} \int_0^1 dB_{i,u} G_u' \left(\int_0^1 G_u G_u' du \right)^{-1} \int_0^1 G_u dB_{i,u}' = \sum_{i=1}^{m-r} \mathbb{T}_i. \end{aligned}$$

The process \mathbb{T}_i for $i = 1, \dots, m - r$ is a function of $B_{i,u}$ and G_u , both of which are functions of the $(p - r)$ -dimensional standard Brownian motion B_u . Inspection of these

functions shows that they are invariant to the relabelling of the coordinates of B_u , so that $\mathbb{T}_1, \dots, \mathbb{T}_{m-r}$ are identically distributed and any pairs $\mathbb{T}_j, \mathbb{T}_k$ are also identically distributed. Hence,

$$\begin{aligned} \mathbb{E} \left(\sum_{i=1}^{m-r} \mathbb{T}_i \right) &= \sum_{i=1}^{m-r} \mathbb{E}(\mathbb{T}_i) = (m-r) \mathbb{E}(\mathbb{T}_1), \\ \text{Var} \left(\sum_{i=1}^{m-r} \mathbb{T}_i \right) &= \sum_{j=1}^{m-r} \sum_{k=1}^{m-r} \text{Var}(\mathbb{T}_j, \mathbb{T}_k) = \sum_{j=1}^{m-r} \text{Var}(\mathbb{T}_j) + \sum_{j \neq k}^{m-r} \text{Var}(\mathbb{T}_j, \mathbb{T}_k) \\ &= (m-r) \text{Var}(\mathbb{T}_1) + (m-r)(m-r-1) \text{Cov}(\mathbb{T}_1, \mathbb{T}_2). \end{aligned}$$

In order to relate the moments of the limit distributions of the LR test statistics in the partial and full models, we evaluate the above expressions for $m-r$ in general and for $m-r = p-r$. For the means of the limit distributions, we find

$$\mathbb{E} \left(\sum_{i=1}^{m-r} \mathbb{T}_i \right) = (m-r) \mathbb{E}(\mathbb{T}_1) \quad \text{and} \quad \mathbb{E} \left(\sum_{i=1}^{p-r} \mathbb{T}_i \right) = (p-r) \mathbb{E}(\mathbb{T}_1).$$

Solving both equations for $\mathbb{E}(\mathbb{T}_1)$ and equating the resulting expressions yield

$$\mathbb{E} \left(\sum_{i=1}^{m-r} \mathbb{T}_i \right) = \left(\frac{m-r}{p-r} \right) \mathbb{E} \left(\sum_{i=1}^{p-r} \mathbb{T}_i \right).$$

For their variances, we obtain a set of equations similarly, which are solved for $\text{Var}(\mathbb{T}_1)$ to find the desired expression. ■

References

- Anderson, T.W. and Kunitomo, N. (1992). Asymptotic distributions of regression and autoregression coefficients with martingale difference disturbances. *Journal of Multivariate Analysis* 40, 221–243.
- Bårdsen, G, Eitrheim, Ø, Jansen, E.S. and Nymoen, R. (2005). *The Econometrics of Macroeconomic Modelling*. Oxford University Press, Oxford.
- Berenguer-Rico, V. and Nielsen, B. (2017). Marked and weighted empirical processes of residuals with applications to robust regressions. Department of Economics Discussion Paper Series 841, University of Oxford.
- Berenguer-Rico, V. and Wilms, I. (2018). White heteroscedasticity testing after outlier removal. Department of Economics Discussion Paper Series 853, University of Oxford.
- Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons.
- Boswijk, H.P. (1992). *Cointegration, Identification and Exogeneity* (Tinbergen Institute Research Series 37). Thesis Publishers, Amsterdam.
- Boswijk, H.P. (1995). Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics* 69, 133–158.
- Boswijk, H.P. and Doornik, J. A. (2005). Distribution approximations for cointegration tests with stationary exogenous regressors. *Journal of Applied Econometrics* 20, 797–810.
- Boswijk, H.P., Cavaliere, G., Rahbek, A. and Taylor, A. R. (2016). Inference on co-integration parameters in heteroskedastic vector autoregressions. *Journal of Econometrics* 192, 64–85.

- Brown, B.M. (1971). Martingale central limit theorems. *Annals of Mathematical Statistics* 42, 59–66.
- Cavaliere, G., De Angelis, L. and Fanelli, L. (2018). Co-integration rank determination in partial systems using information criteria. *Oxford Bulletin of Economics and Statistics* 80, 65–89.
- Cavaliere, G., Rahbek, A. and Taylor, A. R. (2010). Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory* 26, 1719–1760.
- Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367–401.
- Doornik, J.A. (1998). Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* 12, 573–593.
- Doornik, J.A. (2003). Asymptotic tables for cointegration tests based on the Gamma-distribution approximation. Available in: https://www.doornik.com/research/coigamma_tables.pdf
- Doornik, J.A. (2013). *Ox: An Object-Oriented Matrix Language*. Timberlake Consultants LTD, London.
- Doornik, J.A. and Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* 70, 927–939.
- Doornik, J.A., Hendry, D. F., Nielsen, B. (1998). Inference in cointegrating models: UK M1 revisited. *Journal of Economic Surveys* 12, 533–572.
- Doornik, J.A. and Hendry, D.F. (2013). *Modelling Dynamic Systems - PcGive 14: Volume 2*. Timberlake Consultants LTD, London.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Engle, R.F., Hendry, D.F. and Richard, J.-F. (1983). Exogeneity. *Econometrica* 51, 277–304.
- Ericsson, N.R. and MacKinnon, J.G. (2002). Distributions of error correction tests for cointegration. *Econometrics Journal* 5, 285–318.
- Godfrey, L.G. (1978). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica* 46, 1303–1313.
- Hannan, E.J. and Heyde, C.C. (1972). On Limit Theorems for Quadratic Functions of Discrete Time Series. *Annals of Mathematical Statistics* 43, 2058–2066.
- Harbo, I., Johansen, S., Nielsen, B. and Rahbek, A. (1998). Asymptotic inference on cointegrating rank in partial systems. *Journal of Business and Economic Statistics* 16, 388–399.
- Hansen, B. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
- Hansen, P. R. (2005). Granger’s representation theorem: A closed-form expression for $I(1)$ processes. *Econometrics Journal* 8, 23–38.
- Harvey, A. and Thiele, S. (2017). Co-integration and control: California smoking and German re-unification. Mimeo, University of Cambridge.
- Hendry, D.F. and Massmann, M. (2007). Co-breaking: Recent advances and a synopsis of the literature. *Journal of Business and Economic Statistics* 25, 33–51.
- Inoue, A. (1999). Tests of cointegrating rank with a trend-break. *Journal of Econometrics* 90, 215–237.
- Jakubowski, A., Ménin, J. and Pages, G. (1989). Convergence en loi des suites d’intégrales stochastiques sur l’espace D^1 de Skorokhod. *Probability Theory and Related Fields* 81, 111–137.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

- Johansen, S. (1992a). Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics* 52, 389–402.
- Johansen, S. (1992b). Testing weak Exogeneity and the order of cointegration in the UK money demand data. *Journal of Policy Modeling* 14, 313–334
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- Johansen, S., Mosconi, R. and Nielsen, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal* 3, 216–249.
- Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press, Oxford.
- Kurita, T. (2011). Local power of likelihood-based tests for cointegrating rank: Comparative analysis of full and partial systems. *Journal of Time Series Analysis* 32, 672–679.
- Kurita, T. and Nielsen, B. (2009). Cointegrated vector autoregressive models with adjusted short-run dynamics. *Quantitative and Qualitative Analysis in Social Sciences* 3, 43–77.
- Kurita, T., Nielsen, H.B. and Rahbek, A. (2011). An $I(2)$ cointegration model with piecewise linear trends. *Econometrics Journal* 14, 131–155.
- Kurtz, T.G. and Protter, P.E. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, 1035–1070.
- Kurtz, T.G. and Protter, P.E. (1996). Weak convergence of stochastic integrals and differential equations. In: Talay D., Tubaro L. (eds) *Probabilistic Models for Nonlinear Partial Differential Equations, Lecture Notes in Mathematics* 1627. Springer, Berlin, Heidelberg.
- Lai, T.L. and C.Z. Wei (1982) Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* 10, 154–166.
- Lai, T.L. and Wei, C.Z. (1985). Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. P.R. Krishnaiah, ed., *Multivariate Analysis VI*, 375–393.
- MacKinnon, J.G., Haug, A.A. and Michelis, L. (1999). Numerical distributions of likelihood ratio tests of cointegration. *Journal of Applied Econometrics* 14, 563–577.
- Nielsen, B. (1997). Bartlett correction of the unit root test in autoregressive models. *Biometrika* 84, 500–504.
- Nielsen, B. (2005). Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. *Econometric Theory* 21, 534–561.
- Nielsen, B. (2006). Order determination in general vector autoregressions. In Ho, H-C., Ing, C-K. and Lai, T.L. (eds.) *Time Series and Related Topics: In Memory of Ching-Zong Wei, IMS Lecture Notes and Monograph* 52, 93–112.
- Paruolo, P. (1997). Asymptotic inference on the moving average impact matrix in cointegrated $I(1)$ VAR systems. *Econometric Theory* 13, 79–118.
- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–1401.
- Schreiber, V. (2015). Cointegration analysis of the exchange rate and bilateral trade between the United Kingdom and Germany. Thesis for M.Phil. in Economics, University of Oxford.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.