Econometrics Journal (1998), volume **1**, pp. 1–25.

Cointegration Analysis in the Presence of Flexible Trends

VOLKERT SIERSMA
Technical University Eindhoven

PHILIP HANS FRANSES[†]
Econometric Institute
Erasmus University Rotterdam

RICHARD D. GILL
Mathematical Institute
University of Utrecht

[†] *Econometric Institute, Erasmus University,*
P.O.Box 1738, NL-3000 DR Rotterdam, The Netherlands.
E-mail: `franses@few.eur.nl`

Received: May 1998

Summary Intercept and deterministic trend functions are known to have a substantial effect on cointegration analysis, and notably on the asymptotic distributions of various test statistics. In this paper we propose a unifying approach to the analysis of cointegrated vector autoregressions by allowing for a wide class of trend functions. Next, estimates of these trends are incorporated in the asymptotic distributions of the test statistics. This approach allows incorporating elaborate drift functions in cointegration analysis, while avoiding the issue of the significance of the trend these functions give rise to. Simulation techniques can yield the appropriate critical values.

Keywords: *cointegration testing, trend functions, asymptotic distribution*

1. INTRODUCTION

Cointegration analysis for trending economic time series amounts to investigating the presence of common stochastic trends. A useful approach to test for cointegration, which is based on a vector autoregressive model (VAR) specification, is summarized in Johansen (1995). To estimate the cointegrating relations and the common trends, which both can be of economic interest, likelihood-based test statistics can be constructed.

When a VAR model includes for example an unrestricted intercept term, the common trends in the vector system contain a deterministic linear trend component. This linear trend component can be removed by imposing the appropriate restriction on the vector of intercepts. When the model includes an unrestricted deterministic linear trend term, the common trends display quadratic trend behaviour, see, e.g. Johansen (1994).

As the common trends reflect the driving forces behind the economic time series, the inclusion of deterministic regressors should somehow match with the empirically observed patterns in the data. At present there are no formal methods to obtain insights into the most appropriate way how one should account for deterministics in possibly cointegrated VAR models. Currently, one therefore analyses several options, e.g. restricted versus unrestricted parameters of the trend term, in order to investigate the robustness of the estimated cointegrating relations. Needless to say that with more than one option, one may face widely varying estimation results which are difficult to compare and to evaluate.

In this paper we propose a cointegration analysis which overcomes the drawbacks of choosing between a limited number of alternative trend specifications which each lead to a different asymptotic theory for the relevant test statistics. The main idea of our proposal is that a wide class of trend functions is allowed for in the VAR model. The relevant estimates of these trends are then incorporated in the asymptotic distribution of the test statistic. In the extremal cases, our approach equals the Johansen approach. A useful side-effect of our approach is that more general functions than only linear and quadratic trends can be used. For example, trend functions such as t^{-1} or $t^{-\frac{1}{2}}$ can be considered. Putting these in the cointegrating equations, our method allows one to investigate the stationarity of, for example, $z(t) = y(t) - x(t) - \phi t^{-\frac{1}{2}}$. This model is useful in case one examines convergence between for example country-specific macroeconomic aggregates.

In section 2 relevant models are reviewed with a focus on the consequences of the inclusion of drift terms. An alternative asymptotic approach is chosen to result in an alternative testing procedure for cointegration where the influence of the deterministic terms is dealt with more carefully. In section 3 the derivation is outlined of the maximum likelihood estimates of the parameters in an error correction mechanism model and in section 4 an outline of the proof is given of the main theorem on the asymptotic distributions of the trace test in various asymptotic regimes for the drift term.

2. COINTEGRATION MODELS WITH TRENDS

This section deals with the influence of the deterministic model terms on the asymptotic distribution of the tests for cointegration. After the influence is examined, an asymptotic setting is developed in which a wide class of deterministic terms can be incorporated in the analysis. Finally, an alternative testing procedure is constructed from the theoretical results.

2.1. The Model

Consider the p-dimensional vector autoregressive process X_t defined by

$$X_t = \Pi_1 X_{t-1} + \ldots + \Pi_k X_{t-k} + \Phi D_t + \epsilon_t, \quad t = 1, \ldots, T \tag{1}$$

for fixed starting values $X_{-k+1}, \ldots, X_0$ and i.i.d. errors ϵ_t with covariance matrix Ω. The deterministic term D_t can include all kinds of deterministic functions of time, restricted only by integrability. Following Engle and Granger (1987), This VAR process can be rewritten in error correction mechanism (ECM) form:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \epsilon_t, \tag{2}$$

where $\Pi = \sum_{i=1}^{k} \Pi_i - I$ and $\Gamma_i = -\sum_{j=i+1}^{k} \Pi_j$, and where Δ is the first difference operator. If there is at least one root $L = 1$ in the characteristic polynomial $|\Pi(L)| = |I - \Pi_1 L - \ldots - \Pi_k L^k|$, the process has a unit root, which is denoted by I(1). When the process is I(1), it can be shown that the matrix Π is singular, i.e. $\Pi = \alpha\beta'$ for some $p \times r$ matrices α and β.

There is a reverse representation result when we consider processes which are I(1) that gives the non-differences series in terms of parameters of the process in ECM form. This representation is useful to show the impact of model assumptions on the non-differenced process. The original result is due to Engle and Granger (1987) and a proof is found in Johansen (1995).

Denote by $\beta_\perp$ a $p \times (p-r)$ matrix with the property that $\beta'\beta_\perp = 0$. The space spanned by the columns of $\beta_\perp$ is now orthogonal to the space spanned by β, i.e. we have a decomposition of the p-dimensional space into the directions determined by the columns of β and $\beta_\perp$.

Theorem 1. Granger's representation theorem *If for the process (1)* $|\Pi(L)| = 0$ *implies that* $|L| > 1$ *or* $L = 1$, *and rank*$(\Pi) = r < p$ *in (2), then there are* $p \times r$ *matrices* β *and* α *such that*

$$\Pi = \alpha\beta'. \tag{3}$$

If, additionally, the matrix $\alpha_\perp'\Gamma\beta_\perp$ *is of full rank, then* ΔX_t *and* $\beta' X_t$ *are trend stationary and we can represent the solution of (2) by*

$$X_t = C\sum_{i=1}^{t}(\epsilon_i + \Phi D_i) + C(L)(\epsilon_t + \Phi D_t) + P_{\beta_\perp} X_0, \tag{4}$$

where $C = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'$, $\Gamma = I - \sum_{i=1}^{k}\Gamma_i$ *and* $C(L) = \sum_{j=1}^{\infty} C_j L^j$ *with exponentially fast decreasing coefficients* C_j *and* $P_{\beta_\perp} = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'$, *the projection on the space spanned by the columns of* $\beta_\perp$. *The process* X_t *is a cointegrated I(1) process with cointegrating vectors* β.

In the representation (4), X_t can be split in a trend stationary part, $C(L)(\epsilon_t + \Phi D_t) + P_{\beta_\perp} X_0$ and a non-stationary part, $C\sum_{i=1}^{t}(\epsilon_i + \Phi D_i)$. From this it is seen that $\beta' X_t - E(\beta' X_t)$ is stationary, since $\beta' C = 0$. The columns of β are the r cointegrating relations for X_t, whereas in the directions $\beta_\perp$ the non-stationary random walk components will dominate. The cointegration assumption $\Pi = \alpha\beta'$ states that there are *at most* r cointegrating relations. This is exploited in the testing procedures described below.

In the extreme case that rank(Π)=0, there are no such α and β that $\Pi = \alpha\beta'$ and both $\alpha_\perp$ and $\beta_\perp$ are square matrices. Then C is of full rank and the series is I(1) in all directions. In this case $\beta = 0$ and we have no cointegration. In particular the series $\beta_\perp' X_t$ is I(1) and without cointegration.

In the other extreme case rank(Π)=p and we can take $\alpha = \Pi$ and $\beta = I_p$ and conclude that X_t is stationary in all directions and the choice of $\beta = I_p$ shows us that all the individual series are I(0) and so the series itself is I(0).

In the following the error correction representation (2) with $\Pi = \alpha\beta'$ is used as the general model for I(1) series with r cointegrating relations, denoted as H_r. The advantage of the ECM form over the conventional VAR-model (1) is that in the former the long- and short-term dynamics are isolated in $\alpha\beta'$ and $\Gamma_1, \ldots, \Gamma_{k-1}$, respectively. The Granger's representation of the ECM form is used whenever properties of the non-differenced process are to be examined.

2.2. The Impact of the Drift Term

The role of the deterministic term is found to be crucial in inference and estimation procedures involving the above models. For different forms of the deterministic terms the asymptotic distributions of test statistics will also differ. It is important to know what the properties of the process are for various forms of the deterministic part ΦD_t before focusing on the asymptotics. Since the part ΦD_t enters in our analysis in a model for the first differenced series, this part is called the drift of the process. The deterministic part of the non-differenced series is called the trend of the process. We now proceed with a brief treatment of some forms of drifts in cointegrated processes often used for practical purposes.

A cointegrated process without drift amounts to a constant trend and stationary cointegrating relations. In case of no drift the Granger representation has the form

$$X_t = C \sum_{i=1}^{t} \epsilon_i + C(L)\epsilon_t + P_{\beta_\perp} X_0.$$

In the process X_t a constant $P_{\beta_\perp} X_0$ is observed. The cointegrating relations however are centered around zero, this is because $\beta' X_t = \beta' C(L)\epsilon_t$ is a zero mean, stationary process. This process, denoted by $H_0(r)$, can be used to model a series when the individual variables are known to be proportional to each other in the long run.

A process with a constant drift term can be used to model a series with a linear trend. With this constant drift, i.e. $\Phi D_t = \mu$, the process can be written as

$$X_t = C \sum_{i=1}^{t} \epsilon_i + C\mu t + C(L)(\epsilon_t + \mu) + P_{\beta_\perp} X_0.$$

In this case the process itself has a linear trend but the cointegrating relations only allow for a constant since $\beta' X_t = \beta' C(L)(\epsilon_t + \mu)$. In this process, denoted by $H_1(r)$, the equilibrium relations are proportional with a constant added.

A process with a general drift term ΦD_t implies a trend of $\Phi C \sum_{i=1}^{t} D_i + C\Phi(L)(D_t) + P_{\beta_\perp} X_0$, i.e. a term proportional to the primitive function of D_t, the drift itself and a constant. Thus a polynomial drift of degree k for example, implies a trend which is a polynomial of order $k+1$.

Foe the process $H_1(r)$ with a constant drift term, a restriction on the drift parameters changes its properties. Such processes exhibit linear trends and a constant term in the cointegrating relations. From the Granger representation it is seen that the linear part of the trend enters the process through the coefficient $C\mu$, i.e. through the combination $\alpha'_\perp \mu$. As a nested process of the full constant drift process a restricted process arises when $\mu = \alpha\rho$, i.e. $\alpha'_\perp \mu = 0$. Now $C\mu = C\alpha\rho = 0$ and the linear term cancels. Still this process allows for a constant in the r cointegrating relations, but the non-differenced process itself loses the deterministic linear trend term and only a constant remains.

In general we can have the model restriction that $\alpha'_{\perp}\Phi_i = 0$ for some columns of Φ. Those parts of the deterministic drift will not contribute to the trend of the process as primitives of the drift, i.e. to the part $C\Phi\sum_{i=1}^{t} D_i$ of the trend.

To distinguish between the restricted process and the full process is important. Even though it is a nested case of a process with a constant drift, its asymptotics are different as shown in Johansen (1995) for some specific cases. The choice between the use of the full model or the restricted model may be based on economic insight, when modelling economic time series. Frequently, however, especially when the drift term gets more complicated, we may need tests to see which is the most appropriate. Given the wide range of possibilities, these tests may not be easy to use and interpret. Therefore, it is the purpose of our paper to suggest a cointegration method that does not require such tests.

We see that implicitly one has to choose between the several ways of dealing with the trend, before focusing on the amount of cointegrating relations. Given that it is well known that mistakenly in- or excluding deterministic terms biases estimates of the cointegration rank, this aspect can be inconvenient. In the next subsection a more general way of dealing with the trend term in the testing procedure is presented where we do not have to choose between various ways to handle trend behaviour.

2.3. The Asymptotic Regime for the Drift Term

To deal with more general trends in a more general way, a new asymptotic framework is needed . To do this, two new approaches are introduced.

Asymptotic Regimes. In economic modelling only part of the underlying processes is observed. Economic theory may indicate a time trend, but this might not be so clear from the data at hand. Or, different economic views may contradict one another on the deterministic part of the process. In general, trends are many times not clearly visible in the data and economic theory is sometimes not clear about them either. This uncertainty about the trend is focused on in our inference.

A process with a constant drift, i.e. $\Phi D_t = \mu$, is used to explain our approach. This process has Granger representation

$$X_t = C\sum_{i=1}^{t}\epsilon_i + C\mu t + C(L)(\epsilon_t + \mu) + P_{\beta_\perp}X_0,\ t = 1,\ldots,T. \tag{5}$$

A process generated by this model exhibits a dominating linear time trend $C\mu t$ in the long run, since it is the only factor without a finite limit if premultiplied by $T^{-\frac{1}{2}}$. To focus on the uncertainty of the trend, the constant μ is made dependent on T, denoted by μ_T. Then the representation (5) becomes

$$X_t = C\sum_{i=1}^{t}\epsilon_i + C\mu_T t + C(L)(\epsilon_t + \mu_T) + P_{\beta_\perp}X_0. \tag{6}$$

Three distinct cases of asymptotic behaviour for the constant drift are distinguished. Firstly, we have the dominating case of $\mu_T = \mu$ as in (5). This case corresponds with a near certainty that the linear trend is supported by the data when T is large. Asymptotically, the p-value of rejecting a linear trend moves to zero. Secondly, we have a balanced case of $\mu_T = \mu T^{-\frac{1}{2}}$. Now the deterministic part will also converge to a finite limit when we premultiply (6) by $T^{-\frac{1}{2}}$. The linear trend is now blurred by the stochastic process. In terms of chances there is a 10% to 90% chance that the linear term is accepted statistically. In this case the p-value of rejecting a linear trend is asymptotically distributed somewhere between a uniform distribution on $[0,1]$ and a distribution with all mass at zero. Finally we have the case where the linear part vanishes as the rest converges, i.e. $\mu_T = \mu T^{-1}$. This case corresponds with an insignificant probability that the linear term is present, i.e. economic theory may suggest a linear term, but there is no evidence of a linear trend in the data. The p-value will be uniformly distributed in the limit.

The notion of the three asymptotic regimes devloped to a general model for the drift gives rise to the following three different cases, all having different asymptotics.

Definition 1. *Define* $\Phi_{1,T} = \Phi_1$, $\Phi_{2,T} = \Phi_2 T^{-\frac{1}{2}}$ *and* $\Phi_{3,T} = \Phi_3 T^{-1}$ *and define* $X_t^{(i)}$ *for* $i = 1,2,3$ *as the solution of*

$$\Delta X_t^{(i)} = \alpha\beta' X_{t-1}^{(i)} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-1}^{(i)} + \Phi_{i,T} D_t + \epsilon_t, \tag{7}$$

such that by Granger's representation theorem, $X_t^{(i)}$ *is given as*

$$X_t^{(i)} = C \sum_{i=1}^{t} \epsilon_i + C\Phi_{i,T} \sum_{i=1}^{t} D_i + C(L)(\epsilon_t + \Phi_{i,T} D_t) + P_{\beta_\perp} X_0. \tag{8}$$

The meaning of these cases is the same as in the above model with linear trend. In the dominating case of $i = 1$ the significancy of the trend is guaranteed. In the balanced case of $i = 2$ the deterministic part is partly hidden by the stochastic part. In the vanishing case of $i = 3$ the trend is not discernable from the stochastics of the process.

Note that all three cases have the same estimators and models, and that only different assumptions are made about the asymptotic behaviour of the trends. The above asymptotic assumptions on the drift parameters are for the purpose of determining the asymptotic certainty of the linear part of the trend only. Likelihood ratio tests, such as described in Johansen (1995), may be quite clear about whether or not a trend term should be included, but (extensions of) these tests are what we want to avoid in the following.

Asymptotic Approach. From (8) it is seen that in general the deterministic part of the series is composed of the drift and lags of itself, a constant term, and the drift summed over the elapsed time. The main difference between trend and drift is thus the extra sum of drift terms seen in the trend. Since this sum is in general not a closed expression, its convergence is not straightforeward.

We aim to impose an asymptotic approach that deals with all possible drift terms at the same time. In most previous literature on cointegration, asymptotics were done simply by moving the time to infinity. Here however, we rewrite the drift term as $D_t = d(\frac{t}{T})$. By doing this, the dataset becomes larger when T increases, but time does not move to infinity. It is as if the data was collected at a higher frequency.

The following convergence result is an important support for this alternative asymptotic approach.

Lemma 2. *For $u \in [0,1]$ we have that if $T \to \infty$*

$$T^{-1} \sum_{i=1}^{[Tu]} d(\tfrac{i}{T}) \to \int_0^u d(y)dy \tag{9}$$

This lemma shows that the trend in the series should be composed of a term proportional to a primitive function of the drift function $d(u)$, and the drift itself plus a constant term. This result holds for the whole class of integrable functions d and thus states that in our asymptotic approach all sums of reasonable drift functions converge if divided by T.

Again this alternative definition of the drift term is an asymptotics assumption; similar to the assumptions on the parameters $\Phi_{i,T}$, which are also asymptotics assumptions. If the model was used to forecast a series, the drift should be taken D_t since we are actually moving further in time, and T fixed in $\Phi_{i,T}$ to make Φ constant in time.

The asymptotical framework that is constructed in the above serves two purposes. The first is to establish a balanced case between the two extreme cases where the asymptotics of the trend are clear. It is already stated that assumptions about the trend are sometimes not wanted and the freedom there is in the balanced case is exploited in the testing procedure below. The second is to be able to include a general class of functions in the analysis rather than two or three most used ones. This is achieved by the unorthodox approach to asymptotics that does not extend into infinite time and so bounds the trend limits to an $\mathcal{O}(T^{-\frac{1}{2}})$ function.

2.4. The Asymptotic Distribution of the Rank Test

The above approaches have their influence on the distribution of the statistic of the rank test. The three asymptotic regimes we established in the previous section are expected to give different asymptotic distributions for this statistic. This is seen clearly in the theorem below. Not only does it discern between the asymptotic regimes, it is also stated in a general way to account for deterministic functions of a wide class.

The so called *trace* test is considered, which is the test for H_r in the general model H_p for a certain model $D_t = d(\frac{t}{T})$ for the drift. The corresponding test statistic is given by

$$-2 \log Q(H_r|H_p) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i), \tag{10}$$

where $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ are the ordered eigenvalues of the eigenvalue problem (28) below. The asymptotic distribution of the trace test statistic for the three cases of asymptotic behaviour of the drift term is given by the following theorem. The proof of this theorem is found in Section 4.

Theorem 3. The asymptotic distribution of the rank test statistic *The limit distribution of the likelihood ratio test statistic for the hypothesis* $\Pi = \alpha\beta'$, *where* α *and* β *are* $p \times r$ *matrices, is given by*

$$tr\left\{\int_0^1 (dB)F' \left[\int_0^1 FF'du\right]^{-1} \int_0^1 F(dB)'\right\} \tag{11}$$

where B *is a* $p-r$ *dimensional Brownian motion and* F *depends on the model for and the asymptotic behaviour of the deterministic term. If the deterministic term* $\Phi D_t = \Phi_{i,T} d(\frac{t}{T})$ *for the three cases of asymptotic behaviour* $i = 1, 2, 3$ *from definition 1 and if* $\alpha_\perp \Phi_i \neq 0$ *then in case* $i = 1$ *of a dominating trend, we define* $d_b(u)$ *as the stacked entries* $d_j(u)$ *of* $d(u)$ *for which* $\int_0^u d_j(y)dy \notin span(d(u))$ *with* m *the dimension of* $d_b(u)$ *and we have*

$$F(u) = \begin{pmatrix} B(u) - A_{11}d(u) \\ \int_0^u d_b(y)dy - A_{12}d(u) \end{pmatrix} \tag{12}$$

where A_{11} *and* A_{12} *are determined by correcting the* $p-r-m$ *dimensional* $B(u)$ *and* $\int_0^u d_b(y)dy$, *respectively, for the drift term* $d(u)$. *In the case* $i = 2$ *of balanced trends we have*

$$F(u) = B(u) + \alpha'_\perp \Phi_2 \int_0^u d(y)dy - A_2 d(u) \tag{13}$$

where A_2 *is determined by correcting* $B(u) + \alpha'_\perp \Phi_2 \int_0^u d(y)dy$ *for the deterministic drift* $d(u)$. *Finally, in the case* $i = 3$ *of vanishing trends*

$$F(u) = B(u) - A_3 d(u) \tag{14}$$

where A_3 *is determined by correcting* $B(u)$ *for the deterministic* $d(u)$ *process.*

From this theorem it is seen that the asymptotic distribution depends on the amount of cointegrating relations and the chosen model for the drift term. Only in the case $i = 2$ of balanced trends it depends also on the factor $\alpha'_\perp \Phi_2$. This is the general theorem that gives the distributions for the testing procedure described below.

2.5. Testing for Cointegration with Flexible Trends

The three aysmptotic regimes have a clear connection. From the above the three different types of asymptotic behaviour of the deterministic part are seen to result in different asymptotic distributions of the likelihood ratio test for the cointegrating rank. Until now these three cases are investigated separately. When the interaction between the asymptotic behaviours is examined, it is to be expected that the asymptotic distributions of the rank test will converge to a second if the asymptotic behaviour becomes the other.

Denote the three asymptotic trace test distributions for a certain model by tr_1, tr_2 and tr_3 corresponding to the cases $i = 1, 2, 3$, and focus on the ECM-model (7) in case $i = 2$ of balanced trends:

$$\Delta X_t = \alpha'\beta X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-k} + \Phi_2 T^{-\frac{1}{2}} D_t + \epsilon_t \tag{15}$$

The next result shows the relation between the balanced case and the two extremal cases.

Theorem 4. Convergence of the asymptotic regimes *For the model (15) with fixed deterministic part we have that*

$$\Phi_2 \to 0 \Rightarrow tr_2 \xrightarrow{w} tr_3 \tag{16}$$

and

$$\Phi_2 \to \infty \Rightarrow tr_2 \xrightarrow{w} tr_1 \tag{17}$$

where $\xrightarrow{w}$ *denotes convergence in distribution.*

The proof of this theorem is given in section 4. Note that $\Phi_2 = 0$ implies $\alpha'_\perp \Phi_2 = 0$ which is exactly the condition necessary to let the deterministic part not contribute to the trend as a primitive function as described in Section 2.2, which, in case of a constant drift term, leads to the model $H_1^\star(r)$.

An alternative testing procedure can be derived from the above convergence theorem. We already stated that the usual way to do cointegration analysis is first to choose between models with or without some columns of $\alpha'_\perp \Phi = 0$, for example choose $H_1^\star(r)$ instead of $H_1(r)$, and then test for the number of stable relations. The above theorem presents an alternative. Instead of choosing first, our testing is done in case $i = 2$ of balanced trends and the influence of the various parts of the drift becomes clear by examining the factor $\alpha'_\perp \Phi_2$. This yields the appropriate function since it takes into account the appropriate amount of influence for the different parts of the drift. In this way a drift or trend term can be included in our models which is maybe needed according to economic theory, but whose presence is not immediately clear from the data. This testing procedure is especially suited in case one is not interested in the trends of the series itself but merely in the trends in the cointegrating relations, since one cannot be sure of the statistical or even economical significance of the trend in the series itself.

The special asymptotic behaviour of the parameters of the drift term, which is assumed in the above testing procedure, does not blur the interpretation of the resulting model. If the model is used for analysis of economic data, the full model is used.

The main drawback is here that the asymptotic distributions are dependent on the parameters Φ_2 and $\alpha_\perp$ and thus cannot be tabulated. The asymptotic distributions can however be simulated relatively fast as will become clear from the linear drift example in the next subsection.

The above result also states that the asymptotic distribution of the trace test is continuous in the parameters of the model. This is the main prerequisite for the bootstrap technique to hold. With the bootstrap, the test statistic is calculated and the model is estimated. Then the estimated residuals are reordered randomly and with the estimated model a new series is produced, a bootstrap series. From many of these bootstrap series, the test statistics are calculated which form a simulation of the actual distribution of the test statistic, under an estimated model. For more on the bootstrap we refer to Efron and Tibshirani (1993).

2.6. A Model with a Linear Drift

We illustrate our findings with a process where the deterministic drift is a linear function, i.e. $\Phi D_t = \mu_0 + \mu_1 t$. This process is also investigated in Johansen (1994) and Perron and Campbell (1993).

The unrestricted process with a linear drift with r cointegrating relations is denoted as $H_2(r)$, conform statndard notation for a process with a linear drift. From its Granger representation it is seen that such a process exhibits a quadratic trend in the non-differenced series and a linear trend in the cointegrating relations. Alternatively, the restricted process $H_2^\star(r)$ is defined as the same model with the additional restriction that $\alpha'_\perp \mu_1 = 0$. This process still has the linear trend in the cointegrating relations but the quadratic trend from the unrestricted process has disappeared. Johansen (1995) gives a likelihood ratio test for testing between these two alternatives. With our approach this test is avoided. If, for example, the divergence of some macroeconomic factors of two countries is investigated, the trends in those series are not our main interest. Of more interest is the possible existence of a cointegrating relation, which would show that the factors move apart linearly in time, rather than stay at a fixed distance.

To specify the asymptotic distributions, the functional F from Theorem 3 in the three asymptotic regime cases for the model with a linear drift is given below. Denote $\bar{B}_i = \int_0^1 B_i(y)dy$, the mean of the ith entry of a p-dimensional standard Brownian motion, such that $B_i(u)$ corrected for a linear drift becomes

$$B_{i|(1,u)}(u) = B_i(u) - \left(4\bar{B}_i - 6\int_0^1 yB_i(y)dy\right) - \left(12\int_0^1 yB_i(y)dy - 6\bar{B}_i\right)u.$$

In the dominating trend case (case $i = 1$), the quadratic part of the trend needs seperate treatment from the rest. The asymptotic distribution of the trace test statistic testing $H_2(r)$ in the general VAR model is now given by (11) with for F the definition

$$F_i(u) = B_{i|(1,u)}(u), \quad i = 1, \ldots, p-r-1$$
$$F_{p-r}(u) = u^2 - u - \frac{1}{6},$$

Observe that this is the same result as in Johansen (1994). Next, in case of the balanced trend asymptotics (case $i = 2$), the functional F is

$$F_i(u) = B_{i|(1,u)}(u) + \alpha'_\perp \Phi_2 (u^2 - u - \frac{1}{6}), \quad i = 1, \ldots, p-r.$$

Here, both $\alpha_\perp$ and Φ_2 need to be estimated before this limit distribution can be used for testing. This gives no problems since maximum likelihood estimators can be constructed for both as is demonstrated in section 3. Finally, The case of vanishing trends (case $i = 3$) gives for F

$$F_i(u) = B_{i|(1,u)}(u), \quad i = 1, \ldots, p-r.$$

As an illustration of the testing procedure from 2.5 several trace test distributions for testing H_{p-1} in H_p are simulated. It can be shown that in case $i = 1$ the trace test distribution is a $\chi^2(1)$ distribution. The limit distribution in case $i = 2$ of balanced trends is simulated for various values of $\alpha'_\perp \Phi_2$, which is one-dimensional in this case. The Brownian motion is approximated by a random walk with $T = 400$ entries and the cumulative distribution is based on 6000 simulation experiments. The computing time for these simulation of the six distributions was approximately 30 minutes on a computer with a Pentium 60MHz processor. The simulated distribution functions are

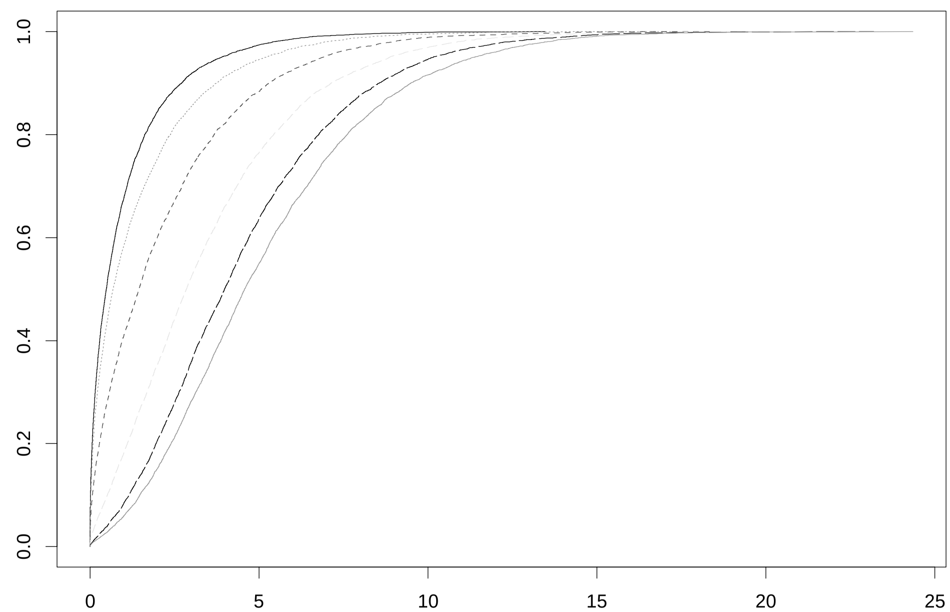

Figure 1. Asymptotic distribution functions for the trace test of H_{p-1} in H_p for various asymptotic behaviours of the linear trend. From top to bottom: the dominating case, four balanced cases with $\alpha'_\perp \Phi_2 = 2,\ 1,\ 0.5,\ 0.25$ respectively, and the vanishing case.

shown in Figure 1. From this figure the distributions are seen to move upward to the dominating distribution when Φ_2 is raised and to converge downward to the distribution of the vanishing case when Φ_2 approaches zero, conform the theorem. There are considerable differences between the graphed distribution functions, especially around the 95% confidence bound, which indicates that the influence of the linear trend is clearly important in the testing procedure.

The testing methodology suggested above implies that the testing is done in the balanced case of $i = 2$ and the distribution simulated. Automatically the correct influence of the trend part is reflected in the asymptotic distribution. This simulated distribution will converge to the distribution for the dominating case if the linear part is pre-eminent in the dataset. The same holds for the vanishing case if there is no statistical evidence of a linear trend in the data.

Here, the dominating distribution equals the asymptotic distribution for testing $H_2(p-1)$ in $H_2(p)$ and the distribution for the vanishing case is the asymptotic distribution for testing $H_2^\star(p-1)$ in $H_2^\star(p)$. Both distributions are tabulated in Osterwald-Lenum (1992) and a comparison is in Table 1.

Table 1. A comparison between the asymptotic distributions for the trace test simulated above and those simulated in Osterwald-Lenum (1992).

	50%	80%	90%	95%	97.5%	99%
dominating trend	0.49	1.62	2.71	3.89	5.04	6.51
Osterwald-Lenum	0.44	1.66	2.69	3.76	4.96	6.65
vanishing trend	4.58	7.61	9.49	11.35	12.95	14.84
Osterwald-Lenum	5.55	8.65	10.49	12.25	14.21	16.26

2.7. Concluding Remarks

In this article the method of cointegration testing developed by Johansen (1988) is generalized in a way that trends are dealt with more carefully. An unorthodox asymptotic approach amounts to a theoretical setting in which the asymptotic distribution of the trace test can be analysed for the large class of all left continuous drift functions. To focus on the statistical uncertainty of the trend, three cases of asymptotic behaviour for the trend are distinguished. For these cases a general theorem which states the asymptotic distributions in each case for the trace test with general drift terms is presented.

The key result however is a convergence theorem that links the three asymptotic trend regimes. The dominating and vanishing cases are (in the special cases of a linear and quadratic trend) identified with known cases in literature. The interesting case is a balanced case, where the asymptotic certainty of the trend is incorporated in the asymptotic distribution itself. The convergence theorem now accounts for continuity in the drift parameters of the asymptotic distribution of the trace test, i.e. when there is a trend in the process, the balanced distribution will converge to the dominating distribution, etc.

This theoretical excercise amounts to an alternative testing procedure for the trace test. By using the balanced asymptotic distribution, the cointegration analist does not have to choose on forehand between the two cases of trend or no trend, as is now the case. Sometimes one is not interested in the trend in the process and by using this flexible trend distribution one abandons hypotheses concerning trends.

A short simulation analysis of the asymptotic distributions for various asymptotic regimes for a model with linear drift shows that different specifications of the trend can lead to widely varying confidence bounds for the trace test. Therefore it is important to deal with the trend hypotheses more carefully as our proposed method does.

This new testing procedure needs some further remarks. The general nature and the inclusion of estimated parameters in the flexible trend asympotic distribution makes it impossible to tabulate confidence bounds, not even in special cases. Therefore the users have to simulate the distribution themselves for each analysis. This can be a time consuming and complex task. Further research has to be done to make fast computation of confidence bounds possible and to make this procedure accessible to applied analists. Furthermore, the convergence theorem gives an important prerequisite for the bootstrap method to hold. Further investigation of the bootstrap in cointegration can benefit from the above results.

3. STATISTICAL ANALYSIS

This section gives a survey of the statistical analysis of the ECM-form models for I(1) series and the derivation of the likelihood-ratio tests for the number of cointegrating relations. The analysis given below leans heavily upon the one in Johansen (1995), chapter 6. Little effort is made to give the complete derivations of the estimators. It is mainly used as a vehicle for the analysis in the next section and to point out the techniques used.

To be able to give maximum likelihood estimators for the parameters in the error correction model (2), we assume that the errors ϵ_t are independent normally distributed with zero mean and covariance matrix Ω. Following the notation in Johansen (1995), we define $Z_{0t} = \Delta X_t$, $Z_{1t} = X_{t-1}$, $Z_{2t} = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}, D'_t)'$ and Ψ being the matrix of parameters corresponding to Z_{2t}, such that the model now has the form

$$Z_{0t} = \alpha\beta' Z_{1t} + \Psi Z_{2t} + \epsilon_t, \quad t = 1, \ldots, T \tag{18}$$

Estimation in the above formula is not straightforward since the parameter matrix of Z_{1t} is assumed of reduced rank, say $r < p$. The log likelihood function, apart from a constant, now is

$$\begin{aligned} -2\log L(\Psi, \alpha, \beta, \Omega) &= T\log|\Omega| + \\ &+ \sum_{t=1}^{T}(Z_{0t} - \alpha\beta' Z_{1t} - \Psi Z_{2t})'\Omega^{-1}(Z_{0t} - \alpha\beta' Z_{1t} - \Psi Z_{2t}) \end{aligned} \tag{19}$$

For given α and β the estimator for Ψ is easily found by examining the first order conditions. We have

$$\hat{\Psi}(\alpha, \beta) = M_{02}M_{22}^{-1} - \alpha\beta' M_{12}M_{22}^{-1} \tag{20}$$

where the M_{ij} for $i, j = 0, 1, 2$ are the product moment matrices $T^{-1}\sum_{t=1}^{T} Z_{it}Z'_{jt}$. Denote by R_{0t} and R_{1t} the residuals of the regression of ΔX_t respectively X_{t-1} on the deterministic terms and the lagged differences Z_{2t}. These are found to be

$$R_{0t} = Z_{0t} - M_{02}M_{22}^{-1}Z_{2t} \tag{21}$$

and

$$R_{1t} = Z_{1t} - M_{12}M_{22}^{-1}Z_{2t} \tag{22}$$

By inserting the estimator (20) for Ψ and the above residuals in (19), the concentrated or profile likelihood function becomes

$$\begin{aligned} -2\log L(\alpha, \beta, \Omega) &= T\log|\Omega| + \\ &+ \sum_{t=1}^{T}(R_{0t} - \alpha\beta' R_{1t})'\Omega^{-1}(R_{0t} - \alpha\beta' R_{1t}) \end{aligned} \tag{23}$$

This is the same likelihood as we would obtain when investigating the regression of R_{0t} on R_{1t}, only now with a parameter matrix $\alpha\beta'$ of incomplete rank. The technique now used to obtain maximum likelihood estimators for the remaining parameters is known as reduced rank regression of R_{0t} on R_{1t}. This is extensively treated in Anderson (1951). Consider the moments

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt} \quad i,j = 0,1 \tag{24}$$

The estimators of α and Ω are obtained by an ordinary regression of R_{0t} on $\beta' R_{1t}$. Since both series are stationary, the usual theory can be applied. We have the maximum likelihood estimators

$$\hat{\alpha}(\beta) = S_{01}\beta(\beta' S_{11}\beta)^{-1} \tag{25}$$
$$\hat{\Omega}(\beta) = S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10} \tag{26}$$

For the estimator for β we investigate again the likelihood function concentrated on β. Since the last term in (23) maximizes for Ω^{-1} a matrix of zeros, we have apart from a constant factor that

$$\begin{aligned} L^{-2/T} &= |\hat{\Omega}(\beta)| = \\ &= |S_{00}||\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})|/|\beta' S_{11}\beta| \end{aligned} \tag{27}$$

Maximizing the likelihood function is now done by maximizing the above expression. From a classical result on eigenvalues and eigenvectors this is done by solving the eigenvalue problem

$$|\rho S_{11} - (S_{11} - S_{10}S_{00}^{-1}S_{01})| = 0$$

which for $\lambda = 1 - \rho$ changes into

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0 \tag{28}$$

Hereby we find p eigenvalues λ_i and their corresponding eigenvectors v_i. The β that minimizes (27) now is a $p \times r$ matrix where the r eigenvectors corresponding to the r largest eigenvalues enter as columns, i.e. $\hat{\beta} = (v_{(1)}, \ldots, v_{(r)})$. Note that the estimator is normalized such that $\hat{\beta}' S_{11}\hat{\beta} = I$, the identity matrix. The eigenvectors diagonalize $S_{01}S_{00}^{-1}S_{01} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_r)$ such that for the maximized likelihood function we find, by inserting $\hat{\beta}$ in (27), that

$$L_{\max}^{-2/T} = |S_{00}| \prod_{i=1}^{r} (1 - \hat{\lambda}_i) \tag{29}$$

Note that the estimator for β is normalized by $\beta' S_{11}\beta = I$ to make the estimator identified. Linear combinations of the cointegrating vectors will make again a perfectly good cointegrating vector. It is better to say that the *cointegrating space* is estimated, a basis given by $\hat{\beta}$.

From the maximum likelihood estimation it is easy to construct likelihood ratio test statistics for testing several model assumptions. The most important test we shall consider is the likelihood ratio test on the number of cointegrating relations. From (29) we find that the likelihood ratio test statistic for testing the hypothesis of r or less cointegrating relations H_r in the general hypothesis of p or less cointegrating relations H_p in a certain model is

$$-2\log Q(H_r|H_p) = -T\sum_{i=r+1}^{p}\log(1-\hat{\lambda}_i) \qquad (30)$$

This statistic is called the *trace test statistic*. However, this statistic is not asymptotically chi-square distributed, since the variables are not stationary. The asymptotic distribution is of the multivariate Dickey-Fuller type and is given in theorem 3 in section 2. From this theorem it is seen that this distribution depends on $p-r$, the model for the deterministic part and the factor $\alpha'_{\perp}\Phi$. The asymptotic distribution of the trace test statistic and its dependence on the deterministic term is derived in the next section.

Testing for the amount of cointegrating relations is done as follows. We start with a test for the H_0 in H_p. If we cannot reject H_0, the amount of cointegrating relations is estimated by $\hat{r}=0$. If H_0 is rejected, we test H_1 in H_p and again, if we cannot reject H_1, $\hat{r}$ is taken 1. If we can, we test H_2 in H_p, etc. This way we have an estimator for r that is known to converge to the true value in the sense of Johansen (1995). Another likelihood ratio test that is often performed is the test of H_r in H_{r+1} for $r=0,\ldots,p-1$. The test statistic belonging to this test is called the *maximum eigenvalue test statisic* and is given by

$$-2\log Q(H_r|H_{r+1}) = -T\log(1-\hat{\lambda}_{r+1}) \qquad (31)$$

An estimator for the amount of cointegrating relations by this test is given by the same procedure as with the trace test. Its asymptotic distribution is only a slight variation of the asymptotic distribution of the trace test. We focus on the trace test, however, since the derivations are the same with both tests.

4. ASYMPTOTIC DISTRIBUTION OF THE TRACE TEST

In section 2 the asymptotic distributions of the rank test for the various asymptotic regimes for the drift term was given without proof. In this section an outline of the proof of theorem 3 and 4 based on the analysis of Johansen (1995) is given, but now with a different asymptotic approach. The parts of the proof is described at large with a focus on the differences between our approach and the one described in (Johansen 1995). For a closer examination of the more basic components one is referred to Johansen (1995).

4.1. Asymptotic behaviour of X_t

Time series can be viewed as functions in $D[0,1]^p$, the space of p-dimensional functions om the unit interval that are right continuous and have left limits. we write the observation time points t as the integer part of Tu for $t = 1, \ldots, T$ and $u \in [0,1]$, i.e. $t = [Tu]$. Viewed as functions of u, time series processes are elements of $D[0,1]^p$. Applied to the VAR-model for series with cointegration, the above time axis adjustment gives

$$X_{[Tu]} = C \sum_{i=1}^{[Tu]} \epsilon_i + C\Phi \sum_{i=1}^{[Tu]} D_i + C(L)\epsilon_{[Tu]} + C(L)\Phi D_{[Tu]} + P_{\beta_\perp} X_0 \qquad (32)$$

where $\epsilon_t, t = 1, \ldots, T$ are assumed independent and identically (not necessarily normal) distributed with zero mean and variance Ω[1]. The two stochastic parts here are

$$X_T^{\text{stat}}(u) = C(L)\epsilon_{[Tu]} \quad \text{and}$$
$$X_T^{\text{sum}}(u) = C \sum_{i=1}^{[Tu]} \epsilon_i$$

that are respectively stationary and $I(1)$ and we want to find weak limits as $T \to \infty$ for both. The following will use some results concerning weak convergence of $D[0,1]$ functions. For a rigid treatment of those results one is referred to van der Vaart and Wellner (1996).

For $X_T^{\text{stat}}(u)$ the function $C(L)$ is a polynomial in L with exponentially fast decreasing coefficients. For such processes $T^{-\frac{1}{2}} X_T^{\text{stat}}(u)$ converges in probability, and hence weakly, to zero. The other stochastic part $X_T^{\text{sum}}(u)$ has a Brownian motion on $[0,1]$ as its weak limit. This is proved by a result concerning the weak convergence of sums of i.i.d. variables, known as Donsker's invariance principle. This principle states that for a sequence ϵ_t of p-dimensional i.i.d. variables with mean zero and variance Ω holds that for $u \in [0,1]$ when $T \to \infty$

$$T^{-\frac{1}{2}} \sum_{i=1}^{[Tu]} \epsilon_i \ \xrightarrow{w}\ W(u) \qquad (33)$$

for a Brownian motion $W(u)$ with variance Ω. From investigating both parts, The stochastic part of the series premultiplied by $T^{-\frac{1}{2}}$ is found to converge in distribution.

Consider now the asymptotics of the series itself. In the directions of the cointegrating relations the series is stationary around the deterministic drift. The asymptotic behaviour in the non-stationary directions $\beta_\perp$ is not so straightforeward. In the case of a dominating trend where $\Phi_{1,T} = \Phi_1$, the classic case, write $\tau = C\Phi_1$. Now τ are all the directions where the sum of the drift is present and these need to be isolated in the dominating case. In general τ is not of full rank m, the amount of deterministic factors, but of a lower rank $m^\star$. Choose thus $\tau = \tau^\star \tau_0'$ with $\tau^\star : p \times m^\star$ and $\tau_0 : m \times m^\star$ matrices and define $d^\star(u) = \tau_0' d(u)$. Let in the following τ be $\tau^\star$, the directions where the integrated drift is present, and $d(u)$ be $d^\star(u)$, how the various drifts enter in each direction. It may seem hypothetical to consider τ's that are of incomplete rank, but they will appear quite often. Take for example a 2-dimensional series modelled by an error correction model with two deterministic factors ($m = 2$). If there is one cointegrating relation, we have since $\text{span}(\tau) \subset \text{span}(\beta_\perp)$ that τ, being a $p \times 2$ matrix, has rank 1.

[1]By $C(L)\epsilon_{[Tu]}$ is actually meant $(C(L)\epsilon_\cdot)_{[T\cdot]}$, i.e. first transforming the series ϵ by the polynomial in the lag operator $C(L)$ and then transforming $t = [Tu]$; in the same way for $C(L)\Phi D_{[Tu]}$.

Now choose γ, a $p\times(p-r-m^\star)$ matrix, orthogonal to both τ and β such that (β,τ,γ) span all of R^p. Define also the normalisation $\bar{c}=c(c'c)^{-1}$ for any c. For the process X_t we have the following convergence result:

Theorem 5. *Let the process* $X_t^{(i)}$ *and* $\Phi_{i,T}$ *for* $i=1,2,3$ *be defined as in definition 1. And define* $B_{1,T}=(\bar{\gamma},T^{-\frac{1}{2}}\bar{\tau})$ *and* $B_{i,T}=\beta_\perp$ *for* $i=2,3$. *Then as* $T\to\infty$ *and* $u\in[0,1]$

$$T^{-\frac{1}{2}}B_{1,T}'X_{[Tu]}^{(1)} \overset{w}{\to} \begin{pmatrix} \bar{\gamma}'CW(u) \\ \int_0^u d(y)dy \end{pmatrix} \tag{34}$$

$$T^{-\frac{1}{2}}B_{2,T}'X_{[Tu]}^{(2)} \overset{w}{\to} \beta_\perp C(W(u)+\Phi_2\int_0^u d(y)dy) \tag{35}$$

$$T^{-\frac{1}{2}}B_{3,T}'X_{[Tu]}^{(3)} \overset{w}{\to} \beta_\perp CW(u) \tag{36}$$

Proof For $i=1,2,3$ we have from (32) that

$$X_{[Tu]}^{(i)}=C\sum_{i=1}^{[Tu]}\epsilon_i+C\Phi_{i,T}\sum_{i=1}^{[Tu]}d(\tfrac{i}{T})+C(L)(\epsilon_{[Tu]}+\Phi_{i,T}d(\tfrac{[Tu]}{T}))+P_{\beta_\perp}X_0$$

The trend-stationary part $C(L)(\epsilon_{[Tu]}+\Phi_{i,T}d(\frac{[Tu]}{T}))+P_{\beta_\perp}X_0$ vanishes in the limit in all three cases by stationarity of $C(L)\epsilon_{[Tu]}$ and the fact that $d(u)$ is finite on $u\in[0,1]$. Let $S_T(u)=T^{-\frac{1}{2}}\sum_{i=1}^{[Tu]}\epsilon_i$ and $I_T(u)=T^{-1}\sum_{i=1}^{[Tu]}d(\frac{i}{T})$. By plugging in definition 1 we find the following.

$$T^{-\frac{1}{2}}\bar{\gamma}'(C\sum_{i=1}^{[Tu]}\epsilon_i+C\Phi_1\sum_{i=1}^{[Tu]}d(\tfrac{i}{T}))=\bar{\gamma}'CS_T(u) \tag{37}$$

$$T^{-1}\bar{\tau}'(C\sum_{i=1}^{[Tu]}\epsilon_i+C\Phi_1\sum_{i=1}^{[Tu]}d(\tfrac{i}{T}))=\bar{\tau}'CT^{-\frac{1}{2}}S_T(u)+I_T(u) \tag{38}$$

$$T^{-\frac{1}{2}}\beta_\perp'(C\sum_{i=1}^{[Tu]}\epsilon_i+C\Phi_2\sum_{i=1}^{[Tu]}d(\tfrac{i}{T}))=\beta_\perp'C(S_T(u)+\Phi_2 I_T(u)) \tag{39}$$

$$T^{-\frac{1}{2}}\beta_\perp'(C\sum_{i=1}^{[Tu]}\epsilon_i+C\Phi_3\sum_{i=1}^{[Tu]}d(\tfrac{i}{T}))=\beta_\perp'C(S_T(u)+\Phi_3 T^{-\frac{1}{2}}I_T(u)) \tag{40}$$

where we use that $\bar{\gamma}'C\Phi_1=\bar{\gamma}'\tau=0$ and that $\bar{\tau}'\tau=I$, the identity matrix. Applying Donsker's invariance principle to $S_T(u)$ and lemma 2 to $I_T(u)$ finishes the proof. □

From this theorem, the effects of the different trend assumptions are clearly visible. In the first case we need a division of the space spanned by the columns of $\beta_\perp$ and a higher order of T to make the trend-part converge. In both the other cases we do not need a division, but the influence of a trend disappears in the third case of vanishing trends.

4.2. Asymptotic behaviour of R_{1t}

The asymptotics of the residual processes (21) and (22) are the key part of the analysis, since it is here that the deterministic part enters the asymptotic distribution. The two processes R_{0t} and R_{1t} are the residuals obtained by regressing ΔX_t respectively X_{t-1} on the deterministic drift D_t and the lagged differences ΔX_{t-1}, $\ldots, \Delta X_{t-k+1}$, see Section 3. The asymptotics of R_{0t} are clear, since the series is stationary. The weak convergence of R_{1t}, however, is important for the understanding of the several asymptotic distributions of the rank tests.

The residuals of a regression of a random variable Y_t on another Z_t for $t = 1, \ldots, T$ are given by plugging in the least squares estimator

$$Y_t - \Big((\sum_{t=0}^{T} Z_t Z_t')^{-1} \sum_{t=0}^{T} Z_t Y_t' \Big)' Z_t$$

or

$$Y_t - S_{YZ} S_{ZZ}^{-1} Z_t$$

where S_{YZ} and S_{ZZ} denote the two product moment matrices. These residuals are called Y_t *corrected for* Z_t and denoted $Y_{t|Z}$. As in section 2 define $D_t = d(\frac{t}{T})$ and assume that $Y_{[Tu]} \xrightarrow{w} Y(u)$ for some $Y(u) \in D[0,1]$. $d(\frac{[Tu]}{T}) \to d(u)$, when d is left continuous. Since this is also enough to make d integrable, we consider only left continuous functions d. The product moment matrices behave like

$$T^{-1} \sum_{t=0}^{T} d(\tfrac{t}{T}) d'(\tfrac{t}{T}) = \int_0^1 d(\tfrac{[Tu]}{T}) d'(\tfrac{[Tu]}{T}) du \;\; \xrightarrow{w} \;\; \int_0^1 d(u) d'(u) du$$

and

$$T^{-1} \sum_{t=0}^{T} Y_t d'(\tfrac{t}{T}) = \int_0^1 Y_{[Tu]} d'(\tfrac{[Tu]}{T}) du \;\; \xrightarrow{w} \;\; \int_0^1 Y(u) d'(u) du$$

so that by application of the continuous mapping theorem, the weak limit in $D[0,1]$ of Y_t corrected for the deterministic D_t is found to be

$$Y(u) - \int_0^1 Y(u) d'(u) du (\int_0^1 d(y) d'(y) dy)^{-1} d(u).$$

For clarity in the formulas to follow we summarize the above by

Lemma 6. *Define the functionals $A_T(Y)$ and $A(Y)$ for $Y \in D[0,1]$ by*

$$A_T(Y) = \sum_{t=0}^{T} Y(t) d'(\tfrac{t}{T}) (\sum_{t=0}^{T} d(\tfrac{t}{T}) d'(\tfrac{t}{T}))^{-1} \tag{41}$$

$$A(Y) = \int_0^1 Y(y) d'(y) dy (\int_0^1 d(y) d'(y) dy)^{-1} \tag{42}$$

such that $A_T(Y)$ is the least squares estimator of A in a linear model $Y_t = Ad(\frac{t}{T}) + \nu_t$. Now $A_T(Y) \overset{w}{\Rightarrow} A(Y)$ and for a stochastic variable Y_t with $Y_{[Tu]} \overset{w}{\Rightarrow} Y(u)$ for some $Y(u) \in D[0,1]$ the residuals $Y_t - A_T(Y)d(\frac{t}{T})$ of a regression of Y_t on the deterministic drift $d(\frac{t}{T})$ have the property that for $T \to \infty$ and $u \in [0,1]$

$$Y_{[Tu]} - A_T(Y_{[T\cdot]})d(\tfrac{[Tu]}{T}) \overset{w}{\Rightarrow} Y(u) - A(Y)d(u) \tag{43}$$

The process R_{1t} is defined as the residual of a regression of X_{t-1} on the deterministic drift and the lagged differences. Since we have three different cases of asymptotics, the process $R_{1t}^{(i)}$ is defined for $i = 1,2,3$ as the residuals of a regression of $X_{t-1}^{(i)}$ on the drift and the lagged differences, i.e. $R_{1t}^{(i)}$ plays the role of R_{1t} in case the parameters concerning the drift are given by $\Phi_{i,T}$. We have

Theorem 7. *Let $R_{1t}^{(i)}$ for $i = 1,2,3$ be the residuals defined above. Let the functional $A(X)$ be defined as in (42). Then as $T \to \infty$ and $u \in [0,1]$*

$$T^{-\frac{1}{2}}B_{1,T}'R_{1[Tu]}^{(1)} \overset{w}{\Rightarrow} \begin{pmatrix} \bar{\gamma}'C\Big(W(u) - A(W(\cdot))d(u)\Big) \\ \int_0^u d(y)dy - A(\int_0^{\cdot} d(y)dy)d(u) \end{pmatrix} = G_1(u) \tag{44}$$

$$\begin{aligned} T^{-\frac{1}{2}}B_{2,T}'R_{1[Tu]}^{(2)} &\overset{w}{\Rightarrow} \beta_\perp C\Big(W(u) + \Phi_2\int_0^u d(y)dy - \\ &\quad -A(W(\cdot))d(u) - \Phi_2 A(\int_0^{\cdot} d(y)dy)d(u)\Big) \qquad (45) \\ &= G_2(u) \end{aligned}$$

$$T^{-\frac{1}{2}}B_{3,T}'R_{1[Tu]}^{(3)} \overset{w}{\Rightarrow} \beta_\perp C\Big(W(u) - A(W(\cdot))d(u)\Big) = G_3(u) \tag{46}$$

Proof Examine the case where there is only one lagged difference term present. First all cases are treated similarly. $R_{1[Tu]}$ is the residual series of a regression of X_{t-1} on D_t and ΔX_{t-1}. Note that this is the same as first correcting both X_{t-1} and ΔX_{t-1} for the drift D_t and then correcting for each other. For the series R_{1t} we now have that

$$R_{1t} = (X_t - A_T(X)D_t) - S_{X\Delta|D}S_{\Delta\Delta|D}^{-1}(\Delta X_t - A_T(\Delta X)D_t) \tag{47}$$

where

$$S_{X\Delta|D} = T^{-1}\sum_{t=0}^{T}(X_t - A_T(X)D_t)(\Delta X_t - A_T(\Delta X)D_t)'$$

and

$$S_{\Delta\Delta|D} = T^{-1}\sum_{t=0}^{T}(\Delta X_t - A_T(\Delta X)D_t)(\Delta X_t - A_T(\Delta X)D_t)'$$

Note from the model equations that the process $\Delta X_t - A_T(\Delta X)D_t$ is a stationary process and $(X_t - A_T(X)D_t)$ is I(1). By the Law of Large numbers for ergodic processes we have that $S_{\Delta\Delta|D}$ converges to a mean, i.e. $S_{\Delta\Delta|D}^{-1} = \mathcal{O}_p(1)$. By a result on convergence to stochastic integrals it can be found that $S_{X\Delta|D}$ converges to a certain stochastic integral and hence is also $\mathcal{O}_p(1)$. This result can be found as result (B.20) in Johansen (1995) and uses the techniques of Chan and Wei (1988). Since the stationary series $(\Delta X_t - A_T(\Delta X)D_t)$ premultiplied by $T^{-\frac{1}{2}}$ converges to zero in probability, the whole last term in (47) vanishes when examining the weak limit of $T^{-\frac{1}{2}}R_{1[Tu]}$. Inclusion of more lagged differences gives the same results.

Returning to the three different cases, we see that

$$T^{-\frac{1}{2}}B'_{i,T}A_T(X^{(i)}_{[T\cdot]}) = A_T(T^{-\frac{1}{2}}B'_{i,T}X^{(i)}_{[T\cdot]}).$$

The required results follow by theorem 5 combined with lemma 6. □

It is important to note that the above theorem, however true, is not altogether useful in all cases. In the case $i = 1$ of a dominating trend examine the case where for one of the components, say $d_i(u)$, in $d(u)$ holds that $\int_0^u d_i(y)dy = d_j(u)$ for some other component $d_j(u)$ in $d(u)$. From (44) we see that a zero entry appears in the vector process $G_1(u)$, which is highly unwanted in the coming proofs. Here it is wrong to ignore the fact that the process $X_t^{(1)}$ itself might exhibit a trend component $\int_0^u d_i(u)$. By another choice of $B_{1,T}$ possible zeros can be removed from the limit process. There is no problem of this kind in the cases $i = 2$ and 3 of balanced and vanishing trends.

In general, define $d_a(u)$ as the stacked components $d_i(u)$ in $d(u)$ for which hold that $\int_0^u d_i(y)dy \in \text{span}(d_j(u),\ j = 1,\ldots,m)$ and define $d_b(u)$ as the stacked components for which this does not hold. In this way we can write $d(u) = (d_a(u), d_b(u))$ and correspondingly $\Phi_1 = (\Phi_{1a}, \Phi_{1b})$. Now the following more useful result for the dominating trend case $i = 1$ holds.

Extension to theorem 7 *Let $R_{1t}^{(1)}$ be defined as theorem 7 and $A(X)$ as in (42). Define $\tau = C\Phi_{1b}$ and follow the same procedure as described above to theorem 5 when τ is of incomplete rank. Choose γ orthogonal to τ and define $B^\star_{1,T} = (\bar{\gamma}, T^{-\frac{1}{2}}\bar{\tau})$. Then as $T \to \infty$ and $u \in [0,1]$*

$$T^{-\frac{1}{2}}B^{\star\prime}_{1,T}R^{(1)}_{1[Tu]} \quad \xrightarrow{w} \quad \begin{pmatrix} \bar{\gamma}'C\,(W(u) - A(W(\cdot))d(u)) \\ \int_0^u d_b(y)dy - A(\int_0^{\cdot} d(y)dy)d(u) \end{pmatrix} = G_1^\star(u) \tag{48}$$

In the following proofs this convergence result is used for the dominating case $i = 1$.

4.3. The Asymptotic behaviour of the Moments

From the asymptotics of the process itself, the asymptotics of the moments in the eigenvalue problem (28) can be derived. This investigation is done much the same way as in Johansen (1995), but now in a slightly more general context.

In order to facilitate understanding the theorems, let from here the asymptotic trend behaviour be fixed, i.e. R_{1t} is $R_{1t}^{(i)}$ for one of the three trend asymptotics, B_T is $B^\star_{1,T}$ or $B_{i,T}$ for $i = 2, 3$, etc. Also, recall the definition of the moments S_{ij} that appears in the eigenvalue problem

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt},\ i,j = 0,1. \tag{49}$$

For these the following results hold.

Theorem 8. *Define*

$$E\left(\begin{pmatrix} \Delta X_{t|D} \\ \beta' X_{t-1|D} \end{pmatrix}^{\otimes 2} \Big| \Delta X_{t-1|D}, \ldots, \Delta X_{t-k+1|D}\right) = \begin{pmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{pmatrix} \quad (50)$$

where $c^{\otimes 2} = cc'$ for all vectors c. For these conditional expectations we have the following exact relations

$$\Sigma_{00} = \alpha \Sigma_{\beta 0} + \Omega \quad (51)$$
$$\Sigma_{0\beta} = \alpha \Sigma_{\beta\beta} \quad (52)$$

and hence the technical lemma 10.1 from Johansen (1995) holds. For $T \to \infty$ we have

$$S_{00} \xrightarrow{P} \Sigma_{00} \quad (53)$$
$$\beta' S_{11} \beta \xrightarrow{P} \Sigma_{\beta\beta} \quad (54)$$
$$\beta' S_{10} \xrightarrow{P} \Sigma_{\beta 0} \quad (55)$$

Proof The two results (51) and (52) follow directly from the model equation (2) by calculating the appropriate conditional expectations and the fact that $\Delta X_{t|D}$ and $\beta' X_{t-1|D}$ are stationary with zero mean. Correct first for the deterministic part and then for $Z_{2t|D}$, the vector of stacked lagged differences corrected for deterministics, to obtain the formula

$$S_{00} = S_{\Delta\Delta|D} - S_{\Delta Z|D} S_{ZZ|D}^{-1} S_{Z\Delta|D}$$

with

$$S_{\Delta\Delta|D} = T^{-1} \sum_{t=1}^{T} \Delta X_{t|D} \Delta X'_{t|D},$$
$$S_{\Delta Z|D} = T^{-1} \sum_{t=1}^{T} \Delta X_{t|D} Z'_{2t|D}$$

and

$$S_{ZZ|D} = T^{-1} \sum_{t=1}^{T} Z_{2t|D} Z'_{2t|D}.$$

Since both $\Delta X_{t|D}$ and $Z_{2t|D}$ are stationary and ergodic, these product moments converge in probability, according to the law of large numbers to their population values. Now

$$\begin{aligned} S_{00} &\xrightarrow{P} E(\Delta X_{t|D} \Delta X'_{t|D}) - E(\Delta X_{t|D} Z'_{2t|D}) E(Z_{2t|D} Z'_{2t|D})^{-1} E(Z_{2t|D} \Delta X'_{t|D}) \\ &= E(\Delta X_{t|D} \Delta X'_{t|D} | Z_{2t|D}) = \Sigma_{00} \end{aligned}$$

which proves result (53). The other two are proved similarily. □

In the above theorem the conditional second moment is used where in Johansen (1995) the conditional variance is used. This is due to our more general approach whereas in Johansen (1995) all proofs are for a process with a constant drift.

The three following results are central in the analysis of the aysmptotic distribution of the rank test.

Theorem 9. *When $T \to \infty$, we have the three results*

$$T^{-1}B_T'S_{11}B_T \overset{w}{\to} \int_0^1 G(u)G(u)'du \tag{56}$$

$$B_T'S_{1\epsilon} = B_T'(S_{10} - S_{11}\beta\alpha') \overset{w}{\to} \int_0^1 G(u)(dW(u))' \tag{57}$$

$$B_T'S_{11}\beta = \mathcal{O}_p(1) \tag{58}$$

PROOF From definition (49) of S_{11} and theorem 7 it is easily seen that

$$T^{-1}B_T'S_{11}B_T = T^{-1}\sum_{t=1}^{T}(T^{-\frac{1}{2}}B_T'R_{1t})(T^{-\frac{1}{2}}B_T'R_{1t})' \overset{w}{\to} \int_0^1 GG'du$$

The second result cannot be proved by the continuous mapping theorem, since the functional $F(x,y) \to \int_0^1 x(u)(dy(u))'$ is not continuous in general. The proof here involves again the results on convergence to stochastic integrals (Johansen 1995). Since $R_{0t} - \alpha\beta'R_{1t} = \epsilon_t$, we have that

$$B_T'S_{1\epsilon} = T^{-1}\sum_{t=1}^{T} B_T R_{1t}\epsilon_t$$

The proof is done by applying results (B.20) and (B.23) from Johansen (1995) to respectively the stochastic part and the deterministic part in R_{1t}. The desired result is found by summation of the two parts.

For the last result note that

$$B_T'S_{11}\beta = T^{-1}\sum_{t=1}^{T}(B_T'R_{1t})(\beta'R_{1t})$$

which, since $B_T'R_{1t}$ is I(1) and $\beta'R_{1t}$ is I(0), converges again to a stochastic integral which gives the result. □

4.4. The Asymptotic Distribution of the Rank Test

The main ingredients for the proof of theorem 3 in Section 2.4 are stated in the previous theorems 8 and 9. The structure of the proof in Johansen (1995) is kept in the following outline of the proof of theorem 3.

PROOF For $i = 1, 2, 3$ in the general case of $\Phi D_t = \Phi_{i,T} d(\frac{t}{T})$, the likelihood ratio test statistic of testing H_r in H_p is given by

$$-2 \log Q(H_r|H_p) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i), \tag{59}$$

where the eigenvalues $\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p$ are the smallest solutions to the eigenvalue problem (28)

Define $C_T = (\beta, T^{-\frac{1}{2}} B_T)$ with B_T as in theorem 5. Since C_T is square and of full rank for all T, the solutions of (28) are the same as the solutions of $|C_T' S(\lambda) C_T| = 0$. By using previous results we find that

$$\begin{aligned} |C_T' S(\lambda) C_T| \;\; &\xrightarrow{w} \;\; \left| \begin{pmatrix} \lambda \Sigma_{\beta\beta} & 0 \\ 0 & \lambda \int_0^1 GG' du \end{pmatrix} - \begin{pmatrix} \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} & 0 \\ 0 & 0 \end{pmatrix} \right| = \\ &= \;\; |\lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta}| \left| \lambda \int_0^1 GG' du \right| \end{aligned}$$

which shows that the eigenvalueproblem has $p - r$ zero roots and r positive roots.

With similar arguments, also $|(\beta, B_T)' S(\lambda)(\beta, B_T)| = 0$ has the same roots as (28). We have that

$$\begin{aligned} |(\beta, B_T)' S(\lambda)(\beta, B_T)| &= \left| \begin{pmatrix} \beta' S(\lambda)\beta & \beta' S(\lambda) B_T \\ B_T' S(\lambda)\beta & B_T' S(\lambda) B_T \end{pmatrix} \right| \\ &= |\beta' S(\lambda)\beta| \, |B_T' \left(S(\lambda)\right. \\ &\quad \left. - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1} \beta' S(\lambda)\right) B_T| \end{aligned} \tag{60}$$

Now let $T \to \infty$ and $\lambda \to 0$ such that $\rho = T\lambda$ stays fixed. From earlier results we see that the first term in (60) tends to

$$-\Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} + o_p(1) \tag{61}$$

and hence this part has no roots for ρ. Another consequence of the convergence results is that in the limit $B_T' S(\lambda)\beta$ converges to

$$-B_T' S_{10} \Sigma_{00}^{-1} \Sigma_{0\beta} + o_p(1) \tag{62}$$

For the second term in (60), we arrive at

$$\rho T^{-1} B_T' S_{11} B_T - B_T' S_{10} N S_{01} B_T$$

where N equals $\alpha_\perp (\alpha_\perp' \Omega \alpha_\perp)^{-1} \alpha_\perp'$ by application of lemma 10.1 from Johansen (1995). The $p - r$ smallest roots of (60) normalized by T converge to the roots of

$$\left| \rho \int_0^1 GG' du - \int_0^1 G(dW)' \alpha_\perp (\alpha_\perp' \Omega \alpha_\perp)^{-1} \alpha_\perp' \int_0^1 (dW) G' \right| = 0 \tag{63}$$

by theorem 9. The roots of this expression are invariant under linear transformations of the processes G and $\alpha_\perp' W$. Now it can be shown that (63) can be expressed as

$$\left| \rho \int_0^1 FF' du - \int_0^1 F(dB)' \int_0^1 (dB) F' \right| = 0 \tag{64}$$

where B is a standard Brownian motion in $p-r$ dimensions and F is defined as in the theorem. How this is done in case $i=1$ is shown in Johansen (1995). For the other two cases define $B = (\alpha_\perp' \Omega \alpha_\perp)^{-1/2} \alpha_\perp' W$ and transform G by $(\alpha_\perp' \Gamma \beta_\perp)(\beta_\perp' \beta_\perp)^{-1}$. From this transformation it is clear that the Φ_2 from (45) becomes $\alpha_\perp' \Phi_2$ in the limit distribution (13). Thus, the smallest $p-r$ roots normalized by T converge to the roots of (64). From (59) we find the desired result

$$\begin{aligned} -2\log Q(H(r)|H(p)) &= T \sum_{i=r+1}^{p} \hat{\lambda}_i + o_p(1) \overset{w}{\to} \sum_{i=1}^{p-r} \rho_i \\ &= tr\left\{ \int_0^1 (dB)F' \left[\int_0^1 FF'du \right]^{-1} \int_0^1 F(dB)' \right\} \end{aligned}$$

□

4.5. The Convergence of the Asymptotic Regimes

A proof of the convergence result, theorem 4 from Section 2.5, concludes this technical section.

PROOF From theorem 3, the proof of (16) easily follows. To prove (17) we split again $d(u) = (d_a(u), d_b(u))$ for $d_a(u)$ the stacked $d_i(u)$ in $d(u)$ for which holds that $\int_0^u d_i(y)dy \in \text{span}(d_j(u),\ j=1,\ldots,m)$ and $d_b(u)$ the rest. Split $\alpha_\perp' \Phi_2 = (\Phi_{2a}^{\alpha_\perp}, \Phi_{2b}^{\alpha_\perp})$, correspondingly. Then, define $\tau = \Phi_{2b}^{\alpha_\perp}$ choose γ orthogonal to τ such that (τ, γ) is a square nonsingular matrix. Note that we can write

$$F(u) = B(u) + \Phi_{2b}^{\alpha_\perp} \int_0^u d_b(y)dy - A_2 d(u)$$

since the part $\int_0^u d_a(y)dy$ vanishes by correcting for it. Define $C_1 = (\bar{\gamma}, \bar{\tau})$ where the normalization is defined as $\bar{c} = c(c'c)^{-1}$ for any c. Since C_1 is a nonsingular transformation of R^{p-r}, the trace test distribution does not change if we premultiply F by C_1. Since $\bar{\tau}'\Phi_{2b}^{\alpha_\perp} = I$, the identity, we have

$$C_1'F(u) = \begin{pmatrix} \bar{\gamma}'B(u) - A_{21}d(u) \\ \bar{\tau}'B(u) + \int_0^u d_b(y)dy - A_{22}d(u) \end{pmatrix}$$

Now, since by definition $\bar{\tau} \to 0$ for $\Phi_2 \to \infty$, we find that

$$\Phi_2 \to \infty \Rightarrow C_1'F(u) \overset{w}{\to} \begin{pmatrix} \bar{\gamma}'B(u) - A_{21}d(u) \\ \int_0^u d_b(y)dy - A_{22}d(u) \end{pmatrix}$$

To remove the factor $\bar{\gamma}$ in the upper part of the right hand side, $C_1'F(u)$ is again premultiplied by a matrix C_2 defined

$$C_2 = \begin{pmatrix} (\bar{\gamma}'\bar{\gamma})^{-1/2} & 0 \\ 0 & I \end{pmatrix}$$

which is nonsingular. The process $C_2C_1'F(u)$ converges weakly to the required distribution. □

REFERENCES

Anderson, T. W. (1951). Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *Annals of Mathematical Statistics 22* pp. 327-351.

Chan, N. H. and C. Z. Wei (1988). Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *The Annals of Statistics 16(1)* pp. 367-401.

Efron, B. and R. J. Tibshirani (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica 55* pp. 251-276.

Johansen S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control 12* pp. 231-254.

Johansen S. (1994). The Role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables. *Econometric Reviews 13(2)* pp. 205-229.

Johansen S. (1995). *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*. New York: Oxford University Press.

Osterwald-Lenum, M. (1992). A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and Statistics 54(3)* pp. 461-472.

Perron, P. and J. Y. Campbell (1993) A note on Johansen's Cointegration Procedure when Trends are Present. *Empirical Economics 18* pp. 777-789.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer Verlag.